

## A emergência de prefixos na aquisição de português brasileiro: formas analisáveis e a relação com o *input*

### The emergence of prefixes in the acquisition of Brazilian Portuguese: analyzable forms and the relationship with the input

Julia Svazati Assine<sup>1</sup>, Indaiá de Santana Bassani<sup>2</sup>

Universidade Federal de São Paulo — Escola de Filosofia, Letras e Ciências Humanas, Brasil<sup>3</sup>

#### RESUMO

Este artigo descreve a emergência dos prefixos *a-*, *eN-* e *deS-* na produção de crianças falantes de português brasileiro com foco na produção de formas analisáveis a partir de: comparação da ocorrência de formas não-prefixadas e prefixadas com raiz comum; análise da frequência da forma na fala adulta (*input*); produção de formas inovadoras. O *corpus* em análise contém 160 sessões de gravação de produções espontâneas de três crianças entre três e cinco anos. Os resultados mostram que, dentre as formas prefixadas com composicionalidade semântica, 53% apresentam presença prévia da forma sem o prefixo; isto é, são possivelmente analisáveis; a emergência de formas prefixadas analisáveis aumenta com o avanço da idade (p.valor = .030) e o prefixo mais analisável e inovador é o *deS-*.

#### PALAVRAS-CHAVE:

Prefixos. Emergência. Aquisição de linguagem. Frequência. Morfologia.

#### ABSTRACT

This paper describes the emergence of prefixes *a-*, *eN-* and *deS-* in the production of Brazilian Portuguese speaking children focusing on the production of analyzable forms from: comparison of the occurrence of non-prefixed and prefixed forms with common roots; analysis of form frequency in adult speech (*input*); production of innovative forms. The *corpus* under analysis contains 160 recording sessions of spontaneous productions by three children between three and five years old. The results show that, among the forms prefixed with semantic compositionality, 53% have a previous presence of the form without the prefix; that is, they are possibly analyzable; the emergence of analyzable prefix forms increases with age (p. value = .030), and the most analytical and innovative prefix is *deS-*.

#### KEYWORDS:

Prefixes. Emergence. Language acquisition. Frequency. Morphology.

Recebido em: 15/05/2020

Aceito em: 14/08/2020

<sup>1</sup> E-mail: [julia-assine@hotmail.com](mailto:julia-assine@hotmail.com) | ORCID: 0000-0002-4394-7849.

<sup>2</sup> E-mail: [indaia.bassani@unifesp.br](mailto:indaia.bassani@unifesp.br) | ORCID: 0000-0002-5277-2008.

<sup>3</sup> Esta pesquisa foi realizada com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Código de Financiamento 001.

## 1. Introdução

Quando se trata de estudar a aquisição da linguagem no paradigma gerativista, historicamente, a morfologia parece despertar menos interesse em comparação aos estudos de outros componentes da gramática, como a sintaxe e a fonologia<sup>4</sup>. Entre os estudos em aquisição de morfologia, a investigação sobre a formação de palavras por prefixação também parece despertar menos interesse em comparação ao número de pesquisas que investigam os processos de morfologia flexional e da formação de palavras por sufixação (Ferrari Neto, 2012).

Este trabalho, que trata de um tópico ainda pouco explorado na literatura, a investigação da emergência de prefixos na aquisição de morfologia por crianças falantes do português brasileiro, se soma aos trabalhos de Lima (2006) e Barbosa e Gomes (2014) ao abordar a aquisição de afixos derivacionais do português brasileiro. Adiante, na seção 2, mostraremos que nosso trabalho corrobora os resultados desses estudos na medida em que apresentam frequência semelhante dos prefixos na produção infantil e sua relação com a produção adulta.

De caráter primordialmente descritivo, o objetivo geral deste artigo é relatar a emergência dos prefixos *a-*, *eN-* e *deS-* na produção de três crianças paulistas em fase de aquisição do português brasileiro como língua materna. Para o desenvolvimento deste trabalho, utilizamos um *corpus* constituído de 160 sessões<sup>5</sup> de gravação de produções espontâneas, que compreende a faixa etária dos três aos cinco anos de idade<sup>6</sup>. A seleção dos dados foi realizada por meio do *software* AntConc.

Além disso, também foram coletados, classificados e descritos os dados da fala de adultos em interação com as crianças, com o objetivo de atender a parte dos seguintes objetivos específicos:

---

<sup>4</sup> Vale ressaltar que esse cenário pode ser diferente em outros paradigmas teóricos. No paradigma conexionista, por exemplo, existem muitos trabalhos sobre a aquisição de morfologia flexional. Essa relevância dada à morfologia flexional dentro do paradigma conexionista pode ser vista, principalmente, em alguns trabalhos de Steven Pinker e colaboradores, tais como *Overregularization in Language Acquisition* (1992) e *Regular and Irregular Morphology and the Psychological Status of Rules of Grammar* (1994).

<sup>5</sup> As 160 sessões que formam esse *corpus* são parte do projeto *Análise dos sistemas fonético e fonológico* desenvolvido na Universidade de São Paulo e foram gentilmente fornecidas pela Profa. Dra. Raquel Santana Santos.

<sup>6</sup> O banco de dados estudado contém gravações de crianças de um até seis anos não completos. Como nos interessava buscar a produção de morfologia derivacional e, de acordo com Clark (2001, p. 374), a produção de afixos derivacionais aumenta de forma significativa entre os três e quatro anos de idade, optamos pelo recorte a partir dos três anos.

---

- a) Verificar, nos dados do *corpus*, se a criança produz, para cada forma composicional prefixada, a mesma forma sem o prefixo, como indício de que ela está de fato segmentando e analisando a palavra produzida;
- b) Comparar, no *corpus*, a emergência das formas composicionais prefixadas nas produções infantis com a frequência recebida pelo *input*, ou fala adulta;
- c) Observar que tipo de produção composicional prefixada, presente no adulto, falta na criança.
- d) Descrever as formações inovadoras.

Consideramos em nossa pesquisa os prefixos *a-*, *eN-* e *deS-*<sup>7</sup> presentes em formações morfologicamente transparentes e semanticamente composicionais<sup>8</sup>. A decisão sobre a escolha desses prefixos foi feita, inicialmente, pela sua produtividade na língua (Schwindt, 2004) e, em seguida, com base na representatividade revelada nos dados infantis. Após a triagem realizada nas produções das crianças, os prefixos *a-*, *eN-* e *deS-* também foram selecionados para a busca nas produções adultas.

No que se refere às características dos prefixos selecionados, observa-se que: os prefixos *a-* e *eN-* são quase exclusivamente verbais, formando verbos de mudança e resultado. Por exemplo, nos verbos *amaciar* e *engordar* observa-se a mudança do argumento interno para o estado denotado pela raiz (nos casos, *mac-* e *gord-*) e nos verbos *engarrafar* e *acampar* há mudança do argumento interno para o lugar denotado pela raiz (nos casos *garraf-* e *camp-*) (Bassani, 2013). O prefixo *deS-* não faz seleção categorial rígida (Medeiros, 2010; Silva e Miotto, 2009), formando verbos e adjetivos, mas apresenta regularmente as seguintes contribuições semânticas: negação, como em *desleal*, quando o prefixo se anexa a uma raiz que foi categorizada como adjetivo ou nome; reversão, como em *desfazer*, quando o prefixo se anexa a um verbo que pressupõe um estado, nesse caso, que pressupõe a existência de algo feito; e separação, como em *descascar*, quando o prefixo se anexa a uma raiz que denota uma entidade inerente. Estamos utilizando neste trabalho o conceito de raiz tal como proposto pela teoria da Morfologia

---

<sup>7</sup> As letras em maiúscula significam que podem existir alomorfias e diferenças gráficas. As possibilidades para o prefixo *eN-* são *em-* e *en-*; para o prefixo *deS-* são *de/z/* e *de/s/*. Além disso, em nossas buscas, consideramos o alomorfe *diS-* e encontramos apenas um dado com essa sequência: *disjuntar*; portanto, não julgamos relevante a marcação dessa alomorfia na forma de citação do prefixo.

<sup>8</sup> Para a descrição e análise da emergência de prefixos em formas morfologicamente transparentes e semanticamente não composicionais em comparação a formas composicionais, ver Assine (2020) e Assine e Bassani (a sair).

Distribuída (MD) (Halle e Marantz, 1993), que pode ser definida como “uma unidade mínima compartilhada por uma série de estruturas derivadas em perspectiva sincrônica” (Bassani, 2013, p. 22)<sup>9</sup>.

Sobre o termo forma composicional citado acima, esse refere-se à classificação de composicionalidade semântica da formação: uma palavra pode possuir um significado composicional ou não (Marantz, 2007). No primeiro caso, é possível identificar em uma palavra a contribuição semântica das partes que compõem o significado do todo, como em *desligar*, em que o prefixo *deS-* reverte o evento de ligar; no segundo caso, embora uma palavra tenha transparência morfológica, isto é, embora seja possível identificar e segmentar afixo e base, não é possível identificar nesses morfemas a composição do significado do todo. Alguns exemplos são: *assombrar*, *desesperar* e *encaixar*. Nessas formações, podemos perceber que não há denotação direta do significado da raiz no significado da palavra; *assombrar* não é *causar/fazer sombra*; *desesperar* não significa *não esperar* ou *deixar de esperar* e *encaixar* não significa *colocar em/fazer caixa*.

Os resultados gerais apontam que a emergência de formas analisáveis tende a aumentar na produção infantil com o avanço da idade. Dentre os prefixos estudados, o *deS-* é o mais analisável e está presente nas formas inovadoras. As formas inovadoras encontradas são *disjuntar* e *descartelar* entre quatro e cinco anos de idade. Ainda, sobre as formas analisáveis presentes na produção infantil, a maior parte delas estão entre os dados pouco frequentes no *input*; isso significa que alta frequência e analisabilidade não são aspectos dependentes. Além disso, com o avanço da idade, os dados pouco frequentes no *input* se tornam maioria na produção infantil. Dessa forma, tanto as formas analisáveis quanto as formas pouco frequentes no *input* aumentam com a passagem da idade na produção das crianças.

O presente artigo se organiza da seguinte forma: a seção 1 detalha a Metodologia de seleção e classificação dos dados; a seção 2 apresenta os Resultados sobre a emergência dos prefixos *a-*, *eN-* e *deS-*; a seção 3 traz a Síntese e Discussão dos resultados frente a algumas previsões descritivas observadas em Clark (2001); e, finalmente, seguem-se as Conclusões e

---

<sup>9</sup> A definição de raiz no modelo da Morfologia Distribuída é cercada de diversas questões empíricas e teóricas. Discute-se, sobretudo, quais propriedades a raiz possui ao ser inserida na derivação sintática: se é provida de conteúdo fonológico e semântico, se é capaz de definir a estrutura argumental da palavra formada ou se possuem apenas índices de marcação sintática (Harley, 2014; Bassani e Minussi, 2015; Minussi e Bassani, 2017). Para os propósitos deste trabalho, basta-nos assumir que a raiz é o material nuclear fonológico e semântico compartilhado por uma série de palavras sincronicamente relacionadas.

Referências.

## 2. Metodologia de seleção e classificação dos dados

Nesta seção, apresentamos o *corpus* utilizado e o passo a passo da busca pelos dados. Na formação do *corpus*, consideramos apenas formas com transparência morfológica, isto é, formas que permitem a identificação formal clara do afixo e da base que, na maior parte dos casos, se iguala à raiz da palavra. Essas podem ter significado composicional ou não; desse modo, todos os dados encontrados foram classificados levando em consideração a presença ou a ausência de composicionalidade semântica. Nesse trabalho, serão utilizados apenas os dados com composicionalidade semântica. Também nesta seção, explicamos os critérios para a seleção dos prefixos em estudo.

Para chegarmos à escolha dos prefixos em estudo, fizemos uma primeira triagem que, com base em Schwindt (2004), teve como critério de seleção a produtividade prefixal em português. Assim, os prefixos atestados como mais produtivos por Schwindt (2004) e primeiramente selecionados em nossa amostra foram os seguintes: *a*<sup>-10</sup>, *eN*-, *deS*-, *eS*-, *re*- e *i(N)*<sup>-1112</sup>. No quadro 1, abaixo, apresentamos os prefixos com seus respectivos exemplos.

---

<sup>10</sup> Corresponde ao prefixo *a*- verbal e de origem latina, com significado de aproximação e mudança de estado, e não ao prefixo *a*- de origem grega, com significado de negação e privação, como em *amoral* e *afônico*.

<sup>11</sup> Um parecerista anônimo aponta que os prefixos inicialmente selecionados integram processos morfológicos diversos, a saber: prefixação simples (*re*-, *i(N)*- e *deS*-) e derivação parassintética (*a*-, *eN*-, *eS*- e *deS*-). Contudo, como se observou o critério de frequência para a seleção dos dados infantis, neste trabalho, os prefixos selecionados foram *a*-, *eN*- e *deS*-, que integram sobretudo formações parassintéticas. Como o presente artigo enfoca a emergência das formações morfológicas de uma perspectiva descritiva e geral, para o tratamento da diferença entre as formações morfológicas com os prefixos *a*- e *eN*-, por um lado, e *deS*-, por outro, remetemos o leitor a Assine e Bassani (a sair), que apresenta uma proposta de análise que diferencia formações morfológicas internas (parassíntese) para *a*- e *eN*- e formações morfológicas externas (prefixo adverbial) para *deS*-.

<sup>12</sup> Corresponde ao prefixo *i(N)*- de origem latina, com significado de negação e privação, e não ao prefixo *i(N)*- (também de origem latina) com significado de movimento para dentro, já contemplado pelo prefixo *eN*-.

---

Quadro 1 – Prefixos produtivos (Schwindt, 2004) e exemplos

PREFIXOS	EXEMPLOS
<i>a-</i>	acalmar, amaciar, adentrar.
<i>eN-</i>	engavetar, envenenar, enterrar.
<i>deS-</i>	desfazer, desmontar, desleal.
<i>eS-</i>	esfriar, esfaquear, expatriar.
<i>re-</i>	refazer, recadastrar, ressuscitar.
<i>i(N)-</i>	insuportável, ilegal, infeliz.

Esses prefixos foram considerados em nossa busca por dados infantis e, a partir dessa busca, delimitamos o número de prefixos pela frequência de aparição nos dados. Dada a ocorrência encontrada, decidimos considerar em nosso estudo os três prefixos mais frequentes. Esses prefixos mais frequentes são *a-*, *eN-* e *deS-*. Dos 638 *tokens*<sup>13</sup> encontrados na fala infantil, considerando dados composicionais e não composicionais, 390 são formados pelo prefixo *a-* (61,1%), 108 pelo *eN-* (17%) e 120 pelo *deS-* (18,8%). Os 20 *tokens* restantes (3,1%) correspondem a prefixos com baixa ocorrência: *eS-*, *re-* e *i(N)-*.

O *corpus* utilizado contém 160 sessões de gravação provenientes de interações espontâneas de três crianças falantes do português paulista: Am., Ar. e Lz. Esse número de sessões se divide da seguinte forma entre as três crianças: sete são de Am. e correspondem ao intervalo de três anos, quatro meses e nove dias (3;04.09) a quatro anos e 17 dias (4;00.17); 43 são de Ar. e correspondem ao intervalo de três anos e 12 dias (3;00.12) a quatro anos, 11 meses e 12 dias (4;11.12); e 110 são de Lz. e correspondem ao intervalo de três anos e nove dias (3;00.09) a cinco anos, seis meses e sete dias (5;06.07). No geral, as sessões de gravação cobrem a faixa de três a cinco anos e meio de idade e cada uma delas corresponde, em média, a trinta minutos de gravação. A construção desse material se deu por meio de observações longitudinais realizadas durante os anos 2000.

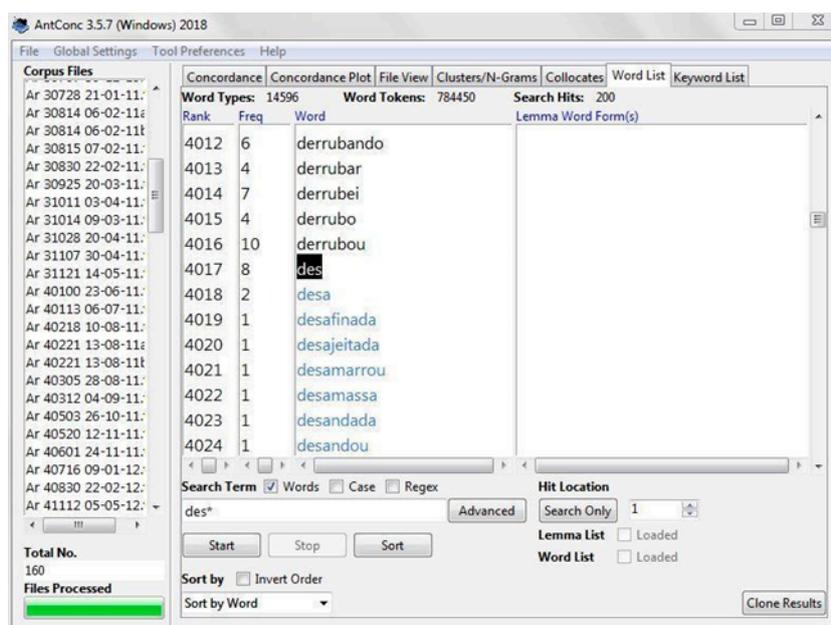
Em relação ao método aplicado para a seleção das palavras prefixadas, utilizamos o *software* AntConc<sup>14</sup>, desenvolvido para a análise de *corpus*. A busca pelos dados pode ser feita por meio de diferentes opções dentro do *software*. Após alguns testes com as diferentes opções, que

<sup>13</sup> Contagem de cada ocorrência de uma mesma palavra.

<sup>14</sup> O programa AntConc é gratuito e pode ser encontrado no seguinte link: <https://www.laurenceanthony.net/software/antconc/>.

são *Concordance*, *File View* e *Word List*, optamos pelo uso da ferramenta *Word List*. Essa ferramenta nos forneceu uma visão mais organizada dos dados e se destacou como o meio mais eficiente de busca. Na figura 1, abaixo, temos um exemplo da interface do *software* com a ferramenta *Word List* selecionada

Figura 1 – Interface do software AntConc



Fonte: AntConc (2018)

Conforme mencionado acima, um dos nossos objetivos é observar se existe, para cada forma prefixada, a ocorrência da mesma forma sem o prefixo, como possibilidade de que a criança analisa a estrutura da palavra. Por exemplo, queremos ver se para as formas *acalm*<sup>15</sup>, *enterr*- e *desmont*- existe a produção das respectivas formas *calm*-, *terr*- e *mont*-. Para isso, também utilizamos o *software* AntConc, mas com uma metodologia de busca diferente. Primeiramente, extraímos todas as partes em comum dos *types*<sup>16</sup> produzidos pelas crianças; uma forma que ocorreu prefixada, como *deslig*-, foi selecionada como forma comum entre os seguintes *types*: *desliga*, *desligado*, *desligar*, *desligou* e *desligue*. Depois disso, por meio da ferramenta *File View* (ferramenta que nos possibilita ter acesso a uma visão ampla do arquivo de transcrição), foi feita a busca pelas formas sem os prefixos, isto é, pelas possíveis formas alomórficas da raiz, em cada

<sup>15</sup> Aqui e no restante do artigo as formas em descrição e análise serão representadas sem a presença dos afixos temáticos e flexionais, dada sua irrelevância para este trabalho. Por exemplo, as diversas ocorrências flexionadas do verbo *desculpar* são representadas pela forma *desculp*-, em que se representam o prefixo e a raiz (ou base).

<sup>16</sup> Uma única ocorrência de palavra.

uma das crianças, separadamente. Por exemplo, em vez de abrirmos o total de transcrições das três crianças de uma só vez dentro do programa, abrirmos apenas o total de transcrições de uma determinada criança por vez e buscamos por cada uma das formas sem prefixo seguindo a ordem crescente das transcrições, que estão organizadas pela idade da criança. Dessa maneira, pudemos selecionar a primeira ocorrência de cada forma não prefixada em cada uma das crianças.

Como mostra a figura 2, a seguir, na classificação dessas formas (prefixadas e não prefixadas) consideramos as seguintes informações: idade da primeira produção da forma prefixada, idade da primeira produção da forma sem prefixo (quando há forma sem prefixo), se há ou não produção da forma sem prefixo, se as produções da forma com prefixo e da forma sem prefixo ocorrem na mesma criança e, por último, se há produção da forma sem prefixo antes da produção da forma com prefixo. Na figura, as células que aparecem em verde claro indicam a idade da primeira ocorrência da forma sem prefixo entre as três crianças, enquanto as células em verde escuro indicam a idade da primeira ocorrência da forma com prefixo. Essas variáveis possibilitam observar as crianças individualmente e a ordem de ocorrência das formas com e sem prefixo; sendo assim, elas podem nos indicar de modo mais minucioso se as crianças analisam ou não as palavras prefixadas. Caso apareça a forma simples antes da forma precedida por prefixo e essa tenha significado composicional, consideraremos a hipótese de possibilidade de análise por parte da criança<sup>17</sup>. Esse cenário pode ser observado, por exemplo, na linha 27 da figura 2, na qual a Lz. produziu a forma sem prefixo *col-* (presente no verbo *colar*) aos três anos, nove meses e 16 dias, e, somente depois, aos quatro anos, seis meses e 27 dias, produziu a forma com prefixo *descol-* (presente no verbo *descolar*).

---

<sup>17</sup> Um parecerista anônimo indica que a premissa de que a criança somente analisa a forma complexa quando há ocorrência da forma simples anteriormente é muito restritiva, pois se sabe que a produção infantil fica aquém da competência linguística e, ademais, o fato de o prefixo integrar outras formações analisáveis pode ser evidência de que há análise de formas prefixadas mesmo que não se ateste a formação sem prefixo de determinada palavra nos dados. Entretanto, adotamos neste trabalho uma posição mais conversadora e cuidadosa de hipótese analítica, pois nosso objetivo é atestar a emergência das formas prefixais e não a aquisição por completo das regras morfológicas de formação de palavras prefixadas. Assim, o trabalho contribui de modo meticuloso no que se refere à observação dos dados morfológicos e expressa como conclusão uma hipótese de que as crianças possivelmente analisam determinados dados prefixados cuja forma não prefixada fora atestada, e conclui que é necessário a realização de testes experimentais para que se possa confirmar essa hipótese.

---

Figura 2 – Classificação das raízes/formas sem prefixo

	C	F	G	I	J	L	M	O	P	Q	R
1	forma c/ prefixo	forma s/ prefixo	forma s/ prefixo Am.	forma c/ prefixo Am.	forma s/ prefixo Ar.	forma c/ prefixo Ar.	forma s/ prefixo Lz.	forma c/ prefixo Lz.	Presença forma s/ prefixo	Presença forma s/ e c/ prefixo na mesma criança	forma s/ antes de forma c/ prefixo
26	DESENROL-	enrol-			Ar 30728		Lz 30728	Lz 41023	SIM	SIM	SIM
27	DESCOL-	col-			Ar 30507		Lz 30916	Lz 40627	SIM	SIM	SIM
29	DESFAZ-	iz-/fiz-/faç-/fa	Am 30409		Ar 30012		Lz 30009	Lz 40319	SIM	SIM	SIM
31	DISJUNT-	junt-			Ar 30116		Lz 30212	Lz 40312	SIM	SIM	SIM
32	DESOBEDES-	obedeS-	Am 30716			Ar 31014			SIM	NÃO	SIM
35	DESLIG-	lig-			Ar 30012	Ar 40520	Lz 30407	Lz 30324	SIM	SIM	SIM E NÃO
37	DESMONT-	mont-			Ar 30211	Ar 30419	Lz 30120	Lz 30212	SIM	SIM	SIM
39	DESPEDAÇ-	pedaç-			Ar 30419		Lz 30205	Lz 30120	SIM	SIM	NÃO
40	DESCULP-	culp-				Ar 30012	Lz 30805	Lz 30226	SIM	SIM	NÃO
58	ENDURES-	dur-			Ar 30402		Lz 30107	Lz 31028	SIM	SIM	SIM
61	ENCH-	chei-	Am 31010		Ar 30100	Ar 30410	Lz 30317	Lz 30114	SIM	SIM	SIM E NÃO

Fonte: Assine (2020)

Antes de prosseguirmos, é importante destacar que para apresentação dos dados, a faixa etária dos três aos cinco anos e meio foi dividida em intervalos de seis meses para uma compreensão mais aprofundada dos resultados e, ainda, sobre a divisão de faixa etária, julgamos importante informar que o número de sessões de gravação e de palavras (no geral) em cada faixa é o seguinte:

Quadro 2 – Número de sessões e de palavras

IDADE	SESSÕES	TYPES	TOKENS
3;0 a 3;6	41	6820	198951
3;6 a 4;0	44	6610	212208
4;0 a 4;6	35	6160	166673
4;6 a 5;0	26	5140	130679
5;0 a 5;6	14	3560	75739

Finalmente, a classificação de frequência das formas nas produções infantis e adultas se deu conforme critério misto, pois, concomitantemente, consideramos o conhecimento que adquirimos em contato com os dados em relação a produção total das formas e fizemos um cálculo com base no total de *tokens*. Os tipos de frequência dos itens são um dos pilares dos padrões de *input* a que a criança é exposta (Ravid, 2019). Na próxima seção, apresentamos os principais resultados descritivos observados no *corpus*.

Quadro 3 – Faixas de frequência

	PRODUÇÃO ADULTA
frequente	Acima de 23 <i>tokens</i>
pouco frequente	De 2 a 23 <i>tokens</i>
nada frequente	Apenas 1 <i>token</i>

### 3. Resultados: a emergência dos prefixos *a-*, *eN-* e *deS-* na fala infantil

A seguir, apresentamos a emergência dos prefixos na produção infantil e a distribuição desses nas diferentes faixas etárias e nas diferentes raízes prefixadas. Trataremos, também, de observar as características dessas raízes em relação à possibilidade de serem analisáveis (ou seja, sua analisabilidade) e em relação à frequência que apresentam na produção adulta; isto é, queremos observar o quão frequentes são as raízes produzidas pelas crianças no *input*. No total, encontramos 57<sup>18</sup> raízes emergentes prefixadas que apresentam composicionalidade semântica; em outras palavras, emergiram 57 palavras prefixadas em que é possível observar a contribuição semântica do prefixo e da raiz para a forma final. Essas 57 formas emergentes estão distribuídas da seguinte maneira entre os estágios observados:

Tabela 1 – Distribuição das raízes prefixadas entre as faixas etárias

Prefixo	Idade					Total
	3 a 3;6	3;6 a 4	4 a 4;6	4;6 a 5	5 a 5;6	
<i>a-</i>	5 (27,8%)	9 (50%)	6 (60%)	3 (30%)	0 (0%)	23 (40,4%)
<i>eN-</i>	5 (27,8%)	5 (27,8%)	1 (10%)	2 (20%)	1 (100%)	14 (24,6%)
<i>deS-</i>	8 (44,4%)	4 (22,2%)	3 (30%)	5 (50%)	0 (0%)	20 (35%)
Total	18 (100%)	18 (100%)	10 (100%)	10 (100%)	1 (100%)	57 (100%)

<sup>18</sup> Estamos contando as formas que se repetem em crianças diferentes.

Levando em consideração os números apresentados no quadro 2 na seção anterior, ao olharmos para os totais de raízes prefixadas por faixa etária (tabela 1), devemos levar em conta que nos dois últimos intervalos (4;6 a 5 e 5 a 5;6 anos), os números de sessões de gravação e de *types* e *tokens* diferem de forma mais acentuada dos números presentes nos outros períodos. Assim sendo, o menor número de raízes na última faixa etária está relacionado a um menor número de dados em geral, e não a uma menor produção dos dados em estudo.

Para comparar os resultados gerais de nosso *corpus*, antes de seguir para o detalhamento por faixa etária, retomaremos dois trabalhos. Barbosa e Gomes (2014) fazem um levantamento de afixos derivacionais do português brasileiro na fala de crianças de dois a cinco anos de idade em três *corpora* diferentes. No primeiro *corpus* (PEUL/UFRJ, nove crianças), é encontrado apenas o prefixo *deS-*, que aparece num total de 11 palavras prefixadas. No segundo *corpus* (AQUIVAR, 34 crianças), em 12 ocorrências de palavras prefixadas, o prefixo mais frequente foi o *eN-*, com seis palavras (50%), seguido pelo *deS-*, com três (25%), *re-*, com duas palavras (16,7%) e, por último, o menos frequente, *eS-*, com uma palavra (8,3%). É importante ressaltar que as ocorrências dos dois *corpora* acima foram contadas em número de *tokens*. No terceiro *corpus* (ASPA-UFMG), sem detalhamento de tamanho, número de participantes e idade dos participantes, é feito um levantamento a partir dos prefixos encontrados nos *corpora* anteriores: *deS-*, *eN-*, *re-* e *eS-* e as ocorrências aparecem em número de *types*, a fim de avaliar produtividade dos quatro prefixos citados. O número de palavras prefixadas encontradas em *types* é de 1097 palavras. O prefixo *deS-* é o mais produtivo e precede 553 dados. Ocupando a segunda posição, está o *re-*, com 417 dados; em seguida, o *eN-*, com 107 dados e, por último, o *eS-*, com 20 dados. Considerando os resultados do *corpus* ASPA-UFMG, ao contrário do que encontramos, o prefixo *re-* parece ser muito frequente. Por outro lado, nos estranha a ausência do prefixo *a-*, que aparece em nossos dados como o prefixo com maior número de ocorrências na fala das crianças, tanto em *types* quanto em *tokens*.

No geral, os pontos de concordância entre os achados deste trabalho e os achados de Barbosa e Gomes (2014) são: a baixa frequência do prefixo *eS-* e a alta frequência dos prefixos *deS-* e *eN-* nos dados infantis. Gostaríamos de destacar que o número de *types* do *corpus* ASPA-UFMG, comparado ao número de *tokens* encontrados nos dois *corpora* anteriores, nos parece muito alto, e, como informamos acima, as autoras não especificam o tamanho do *corpus* ASPA-UFMG. Com base nisso, podemos dizer que há falta de informações importantes para a compreensão dos números apresentados.

Outro estudo que corrobora, em certa medida, com o nosso trabalho, é o de Lima (2006), que investiga a aquisição de morfemas derivacionais e compostos do português brasileiro por crianças de dois a sete anos de idade. Lima (2006) observa a frequência dos processos de formação de palavras e dos afixos. Além disso, é apresentada uma comparação entre a frequência dos afixos presentes na fala infantil e a frequência presente na fala dos adultos. Os dados da fala infantil são provenientes de 62 amostras dos bancos de dados AQUIFONO e DELICRI. Para a coleta dos dados da fala adulta, foram utilizados o *corpus* VARSUL e o Dicionário Aurélio Eletrônico. Sobre os processos de formação de palavras, a prefixação é o menos frequente na fala das crianças, resultando em apenas nove dados de um total de 788. Os prefixos que estão entre esses nove dados são *deS-* e *re-*. O prefixo *deS-* é o mais frequente e tem cinco ocorrências, todas correspondem a ocorrências da palavra *desligar*. As quatro ocorrências do prefixo *re-* correspondem a ocorrências da palavra *remexer* e todas se concentram na mesma faixa etária: três a quatro anos. Por outro lado, as cinco ocorrências do prefixo *deS-*, citadas acima, estão distribuídas em diferentes faixas etárias; três das cinco ocorrências pertencem à primeira faixa: dois a três anos. Isso pode ser uma indicação de que o prefixo *deS-* é adquirido antes do prefixo *re-*. As outras duas ocorrências restantes se dividem entre a terceira e a quarta faixa etária (quatro a cinco anos e cinco a seis anos). A última faixa etária, dos seis aos sete anos, não possui nenhum dado prefixado. Como conclui Lima (2006), nos seus dados, o aumento da produção de palavras formadas por prefixação só ocorre da primeira para a segunda faixa etária, isto é, dos dois até os quatro anos de idade. Em contrapartida, pelos resultados que apresentaremos adiante, parece que o aumento da produção de palavras prefixadas (em *types*) continua a partir dos quatro anos de idade. Em relação à comparação entre a frequência de *deS-* e *re-* na fala infantil adulta, a autora chega à conclusão de que os dados de ambos os *corpora* apresentam a mesma proporção que vemos nos dados infantis: o prefixo *deS-*, em todos os casos, precede mais palavras do que o prefixo *re-*. Nos nossos dados, conforme indicamos na seção 1, *re-* também aparece com menos frequência na fala das crianças, mas a discrepância entre as ocorrências dos dois prefixos é muito maior em comparação à discrepância mostrada no trabalho de Lima (2006).

Além disso, também percebemos, em nossos dados, correspondências entre as produções dos adultos e as das crianças. Por fim, assim como o estudo de Barbosa e Gomes (2014), comentado inicialmente, Lima (2006) não traz ocorrências de palavras prefixadas por *a-*, mas, mais do que isso, também não são encontradas ocorrências com o prefixo *eN-*, que é relevante tanto em nossos dados quanto nos dados de Barbosa e Gomes (2014).

No quadro 4, abaixo, enfocamos os dados emergentes da primeira faixa etária (3 a 3;6 anos).

Quadro 4 – Emergência das formas prefixadas – 3 a 3;6 anos

Idade	Prefixos			Criança
	<i>a-</i>	<i>eN-</i>	<i>deS-</i>	
3;00.12			<i>desculp-</i>	Ar
3;00.30	<i>acompanh-</i>			Ar
3;01.14		<i>ench-</i>		Lz
3;01.20			<u>desaparec-</u> , <i>despedaç-</i>	Lz
3;01.25	<i>assust-</i>			Ar
3;02.05		<i>engraç-</i>		Lz
3;02.11		<i>engraç-</i>		Ar
3;02.12			<u>desmont-</u>	Lz
3;02.19			<u>desarrum-</u>	Lz
3;02.26			<i>desculp-</i>	Lz
3;03.24			<i>deslig-</i>	Lz
3;04.02		<u>enterr-</u>		Ar
3;04.10	<i>aproxim-</i> , <u>aterriç-</u>	<u>ench-</u>		Ar
3;04.19			<u>desmont-</u>	Ar
3;05.14	<i>assopr-</i>			Ar

**Legenda:**

- Composicional e HÁ presença prévia da palavra sem o prefixo (na mesma criança).
- Composicional e NÃO HÁ presença prévia da palavra sem o prefixo (na mesma criança).

Começamos nossa descrição pela produção de Lz., que apresenta o maior número de dados. Nos dados que possuímos dessa criança, a emergência dos prefixos acontece da seguinte maneira: o primeiro prefixo a aparecer é o *eN-*, na forma *ench-* (do verbo *encher*); esse dado, por não apresentar presença prévia da sua forma simples e por estar entre os mais frequentes no *input*, com 74 ocorrências, é possivelmente adquirido pela criança como uma forma composicional não analisável. Em seguida, temos o prefixo *deS-*; esse prefixo surge nas formas *desaparec-* e *despedaç-*. A primeira é composicional com presença prévia da forma simples e a segunda é composicional sem presença (prévia) da forma simples. Ambas são pouco frequentes na fala adulta, com respectivamente cinco e três ocorrências. Em casos semelhantes a de *despedaç-*, em que o dado é composicional, não há produção (prévia) da forma simples e a frequência é mais baixa, podemos supor duas possibilidades: a que a criança realmente ainda não adquiriu forma não prefixada e, portanto, o dado não é analisado e pode ter sido apenas produzido como uma repetição da produção do adulto, ou que a criança já adquiriu e usou a forma não prefixada em

algum momento não gravado, apesar de não termos essa evidência em nossos dados. É importante destacar que há uma limitação prática nesse sentido: nossos dados, assim como toda amostra, não compreendem toda a fala da criança. Nesse caso, diante do que as evidências podem nos mostrar, iremos assumir que os dados sem ocorrência da produção prévia da forma não prefixada são dados possivelmente não analisáveis, independentemente da frequência que apresentam na produção adulta<sup>19</sup>.

Ao todo, nessa faixa etária, temos oito dados emergentes. O prefixo *deS-* é o mais frequente e está presente em seis desses dados. Vale notar que nenhum dado composicional prefixado por *a-* é produzido por essa criança na faixa etária em questão. Dentre os oito dados, cinco são do tipo sem presença da forma simples (*ench-*, *despedaç-*, *engraç-*, *desculp-* e *deslig-*) e três são do tipo com presença anterior da forma simples (*desaparec-*, *desmont-* e *desarrum-*). Tendo isso em vista, do total de dados emergentes, três (37,5%) possuem as características de um dado analisável pela criança e esses são precedidos apenas pelo *deS-*.

Por último, sobre a frequência na fala adulta das formas emergentes, temos cinco dados (62,5%) entre as formas frequentes (ver quadro 3). Eles são: *ench-* (74)<sup>20</sup>, *desculp-* (67), *deslig-* (53), *engraç-* (49) e *desmont-* (38). Como é possível observar, temos entre os dados frequentes quatro formas composicionais não analisáveis e uma forma composicional analisável, isto é, com ocorrência prévia da forma simples. Dos três dados que estão entre os pouco frequentes, dois são composicionais analisáveis: *desaparec-* (5) e *desarrum-* (3), e um é do tipo composicional não analisável: *despedaç-* (3).

Acerca da emergência dos prefixos nos dados produzidos por Ar., chegamos ao seguinte quadro: o primeiro prefixo a aparecer é o *deS-*, na forma *desculp-* (presente no verbo *desculpar* e seus derivados). Sobre a forma *desculp-*, ela nos parece ser uma interjeição de uso recorrente, que na produção adulta conta com 67 *tokens*. Além disso, é uma forma não analisável, pois não há na produção da criança a ocorrência prévia da forma sem o prefixo (*culp*). O segundo prefixo a surgir é o *a-*, em *acompanh-* e *assust-*; e o prefixo *eN-* é o último, na forma *engraç-*.

Desses dados emergentes, o único com pouca frequência nos adultos é o *acompanh-*; os outros estão entre os dados frequentes (*desculp-*, *assust-*, *engraç-*). Sobre a composicionalidade, nenhum desses dados são do tipo composicional com presença prévia da forma simples. O primeiro dado desse tipo surge aos três anos, quatro meses e dois dias, com o prefixo *eN-*

---

<sup>19</sup> Vide nota de rodapé 17.

<sup>20</sup> Número de *tokens*.

(*enterr-*).

Na produção de Ar., temos um total de dez formas que se manifestam ao longo de toda essa primeira faixa etária. Diferentemente do que vimos na produção de Lz., aqui o prefixo mais frequente é o *a-* e ele está presente em cinco dados (50%). Dos dez dados emergentes, seis são composicionais não analisáveis (*acompanh-*, *assust-*, *aproxim-*, *assopr-*, *desculp-* e *engraç-*) e quatro são analisáveis (*desmont-*, *ench-*, *aterriss-* e *enterr-*); isto é, 40% dos dados têm presença (prévia) da forma simples. Considerando apenas as formas composicionais analisáveis, o prefixo mais produtivo é o *eN-*.

Referente à frequência das formas no *input*, dos dez dados emergentes, cinco (50%) estão entre os frequentes; esses são: *ench-* (74), *desculp-* (67), *engraç-* (49), *desmont-* (38) e *assust-* (24). Os cinco dados restantes, que estão entre os pouco ou nada frequentes, são: *assopr-* (12), *acompanh-* (7), *enterr-* (4), *aproxim-* (3) e *aterriss-* (0). Dentre os cinco dados mais frequentes e dentre os cinco menos frequentes, três dados são composicionais não analisáveis.

No geral, quando olhamos exclusivamente as formas composicionais e com produção anterior da forma sem prefixo, constatamos que dos 18 dados produzidos nessa faixa etária, sete (39%) correspondem a esse tipo: *desaparec-*, *desmont-* (2) (presente em *desmontar*), *desarrum-* (presente em *desarrumar*), *enterr-*, *ench-* e *aterriss-*. Esses são os dados que melhor indicam que a criança pode estar analisando o prefixo; a maior parte deles é precedida pelo *deS-* (quatro dados).

Dessa maneira, nesse estágio, o prefixo *deS-* é o mais produtivo quando consideramos as formas analisáveis. Por fim, a respeito da relação entre a frequência presente na produção adulta e a emergência das formas nessa faixa etária, temos dez dados (55,6%) entre os frequentes (três são analisáveis) e oito dados (44,4%) entre os pouco ou nada frequentes (quatro são analisáveis).

A seguir, apresentamos os detalhes das formas que emergem na segunda faixa etária (3;6 a 4 anos).

Quadro 5 - Emergência das formas prefixadas – 3;6 a 4 anos

Idade	Prefixos			Criança
	<i>a-</i>	<i>eN-</i>	<i>deS-</i>	
3;07.28		<u>enrol-</u>	<u>desaparec-</u>	Ar
3;07.28		<i>enrol-</i>		Lz
3;08.15	<i>acredit-</i>			Ar
3;08.25	<u>afund-</u>			Lz

3;08.30	<i>acamp</i> <sup>21</sup> , <i>aproveit-</i>			<b>Ar</b>
3;09.16			<u>descans-</u>	Lz
3;09.30		<i>envenen-</i>		Lz
3;10.07	<i>assust-</i>			Lz
3;10.14			<i>desobedec-</i>	<b>Ar</b>
3;10.14		<i>envergonh-</i>		Lz
3;10.28		<u>endurec-</u>		Lz
3;11.16	<u>apont-</u> , <i>assopr-</i> , <i>afund-</i>			Am
3;11.21	<u>afund-</u>		<u>descans-</u>	<b>Ar</b>
<b>Legenda:</b>				
● Composicional e HÁ presença prévia da palavra sem o prefixo (na mesma criança).				
● Composicional e NÃO HÁ presença prévia da palavra sem o prefixo (na mesma criança).				

Tratando inicialmente da produção de Lz., temos a emergência de sete raízes prefixadas. Desse total de raízes, quatro (57,1%) são composicionais não analisáveis: *enrol-*, *envenen-*, *assust-* e *envergonh-*, e três (42,9%) apresentam presença (prévia) da raiz sem prefixo, isto é, são possivelmente analisáveis: *afund-*, *endurec-* e *descans-*. Podemos perceber que o prefixo *eN-* é o mais frequente e, proporcionalmente, o mais analisável é o *deS-*.

Sobre a frequência das formas, encontramos, entre os sete dados emergentes de Lz., duas (28,6%) formas classificadas como frequentes na produção adulta: *enrol-* (47) e *assust-* (24); e cinco (71,4%) formas classificadas como pouco frequentes: *descans-* (22), *afund-* (6), *endurec-* (5), *envenen-* (5) e *envergonh-* (3). Podemos notar, com isso, que os três dados analisáveis estão entre os pouco frequentes.

Com relação à produção de Ar., há um total de oito formas emergentes. Considerando todo o conjunto de dados, o prefixo *a-* é o mais frequente; no entanto, observando apenas os dados analisáveis, vemos que, proporcionalmente, o mais analisável é o *eN-* e em seguida o *deS-*. Dentre as oito raízes emergentes, quatro (50%) são composicionais não analisáveis: *acredit-*, *acamp-*, *aproveit-* e *desobedec-*; e quatro (50%) são analisáveis: *afund-*, *enrol-*, *desaparec-* e *descans-*.

Acerca da frequência, dois (25%) dos oito dados que aparecem na produção de Ar. nessa faixa etária são classificados como frequentes no *input*: *enrol-* (47) e *aproveit-* (36). Dentre esses, 1

<sup>21</sup> O dado *acamp-* apresenta composicionalidade semântica bem como a presença anterior da forma sem prefixo (*camp-*) na produção da criança; no entanto, não iremos considerá-lo como um dado analisável porque, pelo contexto, fica evidente que é uma repetição da produção adulta. Tanto na produção adulta quanto na infantil, há apenas uma ocorrência da forma *acamp-*. Dessa maneira, consideraremos que a forma em questão é do tipo composicional não analisável (formas em itálico/azul).

(50%) é composicional analisável. Os outros seis (75%) dados, classificados como pouco ou nada frequentes, são: *descans-* (22), *acredit-* (20), *afund-* (6), *desaparec-* (5), *desobedec-* (2) e *acamp-* (1). Dentre esses, três (50%) são composicionais analisáveis.

Por fim, acerca da produção de Am., encontramos três formas emergentes nessa faixa etária: *apont-*, *assopr-* e *afund-*. apenas *apont-* é analisável, correspondendo a 33,3% do total de dados. A respeito da frequência dos dados na produção adulta, os dados *assopr-* (12), *afund-* (6) e *apont-* (4) estão entre os pouco frequentes. Nessa faixa etária, observando a produção das três crianças, encontramos um novo cenário em comparação com a faixa etária anterior: na faixa de 3;6 a 4 anos a maior parte dos dados está entre os menos frequentes na fala adulta ao passo que na faixa de 3;0 a 3;6 anos, a maior parte dos dados está entre os mais frequentes.

Em uma observação geral, no intervalo ilustrado pelo quadro acima, também temos a emergência de 18 formas. O prefixo mais frequente nesse intervalo é o *a-*, mas considerando apenas as formas analisáveis, constatamos que, proporcionalmente, o mais produtivo é o *deS-*. Por último, acerca da frequência, vimos que 77,8% dos dados emergentes (14 dados) foram classificados como pouco ou nada frequentes na produção adulta e que, dentro desse conjunto, estão sete dos oito dados analisáveis.

Abaixo, no quadro 6, apresentamos um resumo da ordem de emergência de cada um dos prefixos em cada uma das crianças. O quadro inclui dados das duas primeiras faixas etárias, abrangendo o período de três a quatro anos, pois trata-se do primeiro ano estudado.

Quadro 6 – Emergência dos prefixos

Ordem	Lz.			Ar.			Am.		
	Prefixo	Tipo de forma	Freq. da forma no input	Prefixo	Tipo de forma	Freq. da forma no input	Prefixo	Tipo de forma	Freq. da forma no input
1º	<i>eN-</i>	<i>ench-</i>	F <sup>22</sup>	<i>deS-</i>	<i>desculp-</i>	F	<i>a-</i>	<i>apont-</i> <i>assopr-</i> <i>afund-</i>	PF
2º	<i>deS-</i>	<i>desaparec-</i> <i>despedaç-</i>	PF	<i>a-</i>	<i>acompanh-</i>	PF	–	–	–
3º	<i>a-</i>	<i>afund-</i>	PF	<i>eN-</i>	<i>engraç-</i>	F	–	–	–

<sup>22</sup> F = frequente e PF = pouco frequente.

Diante do quadro apresentado, podemos notar que as primeiras formas prefixadas são, em sua maioria, composicionais não analisáveis; além disso, não há a possibilidade de se chegar a uma sistematização confiável acerca da ordem de emergência dos prefixos, pois em cada uma das crianças a ordem é diferente.

A seguir, comentamos sobre as formas que surgem na terceira faixa etária (4 a 4;6 anos).

Quadro 7 – Emergência das formas prefixadas – 4 a 4;6 anos

Idade	Prefixos			Criança
	<i>a-</i>	<i>eN-</i>	<i>deS-</i>	
4;00.17	<i>acredit-</i>			Am
4;00.18	<i>assopr-</i>			Lz
4;01.13	<i>abaix-</i>			Ar
4;02.28	<i>apont-</i>			Lz
4;03.12	<i>atir-</i>			Ar
4;03.12			<i>disjunt-</i>	Lz
4;03.19			<i>desf(az, iz, aç, ar)-</i>	Lz
4;04.24	<i>ajunt-</i>			Lz
4;05.07		<i>entort-</i>		Lz
4;05.20			<i>deslig-</i>	Ar

**Legenda:**

- Composicional e HÁ presença prévia da palavra sem o prefixo (na mesma criança).
- Composicional e NÃO HÁ presença prévia da palavra sem o prefixo (na mesma criança).

Como é possível notar no quadro acima, nesta faixa etária há um total de dez dados emergentes. Grande parte desses dados, mais especificamente 60% deles, ou seja, seis dados, são produzidos somente pela Lz. Os quatro dados restantes estão distribuídos entre Ar., com três dados, e Am., com um dado.

Inicialmente, observando os seis dados produzidos pela Lz., que são: *assopr-*, *apont-*, *disjunt-* (presente em *disjunta*), *desf(az, iz, aç, ar)-* (presente em *desfazer*), *ajunt-* (presente em *ajuntar*) e *entort-* (presente em *entortar*), notamos que o prefixo mais frequente é o *a-*. Considerando apenas as formas analisáveis, os prefixos *a-* e *deS-* são os mais produtivos. Em relação às faixas etárias anteriores, aqui nota-se há uma importante diferença: a maior parte dos dados são analisáveis; dos seis dados composicionais, apenas *entort-* é do tipo não analisável (sem presença (prévia) da forma não prefixada).

Acerca da classificação dos dados emergentes diante da frequência presente na produção adulta, temos todos os seis dados entre os pouco ou nada frequentes: *assopr-* (12), *desf(az, iz, aç,*

ar)- (8), *apont-* (4), *entort-* (3), *disjunt* (1) e *ajunt-* (0). Sobre o dado *disjunt-* (usado no verbo *disjuntar* com significado de *separar*), ele é uma formação inovadora e a sua única ocorrência na produção adulta corresponde a uma repetição da produção inovadora da criança. Na situação em que esse dado foi produzido, a criança pede para a mãe não separar (*disjuntar*) as peças do quebra-cabeça<sup>23</sup>.

Sobre a produção de Ar., como citamos acima, há um total de três formas emergentes nessa faixa etária. Essas formas são: *abaix-* *atir-* e *deslig-*. Aqui, o prefixo mais frequente é o *a-* e, pela primeira vez, todos os dados são composicionais analisáveis. Referente à frequência das formas emergentes, *deslig-* (53) está entre as formas classificadas como frequentes na produção adulta, e *abaix-* (12) e *atir-* (3) são classificadas como pouco frequentes.

Sobre a produção de Am., nessa faixa etária encontramos apenas a emergência da forma *acredit-* (presente em *acreditar*). Esse dado, conforme já vimos anteriormente, é do tipo composicional sem a presença (prévia) da forma não prefixada. Ou seja, é um dado possivelmente não analisável. Sobre a frequência da forma no *input*, ela é classificada como pouco frequente, com 20 *tokens*.

Em uma breve análise conjunta dessa faixa etária, vemos que o prefixo *deS-* é, proporcionalmente, o mais analisável, e o prefixo *a-* é o mais frequente. Sobre a frequência das formas, 90% estão entre as formas classificadas como pouco ou nada frequentes na produção adulta. Por fim, sobre a analisabilidade dos dados, 80% são composicionais analisáveis. A seguir, apresentamos a quarta e penúltima faixa etária (4;6 a 5 anos).

Quadro 8 – Emergência das formas prefixadas – 4;6 a 5 anos

Idade	Prefixos			Criança
	<i>a-</i>	<i>eN-</i>	<i>deS-</i>	
4;06.01			<u>desapert-</u>	<b>Ar</b>
4;06.13	<i>atir-</i>			Lz
4;06.20			<i>desajeit-</i>	Lz
4;06.27			<u>descol-</u>	Lz
4;07.24	<i>adormec-</i> <sup>24</sup>			Lz

<sup>23</sup> “CHI: e não disjunta tá? MOT: tá eu não vou disjuntar não.”

<sup>24</sup> O dado *adormec-* apresenta composicionalidade semântica bem como a presença anterior da forma sem prefixo (*dorm-*) na produção da criança; no entanto, pelo contexto, podemos perceber claramente que se trata de uma repetição da produção adulta e não de uma forma que é analisada pela criança. Desse modo, a forma *adormec-* foi classificada como uma forma do tipo composicional não analisável (formas em itálico/azul).

4;08.14		<i>ensabo-</i>		Lz
4;09.19	<u><i>acalm-</i></u>			Lz
4;10.09		<u><i>enterr-</i></u>		Lz
4;10.23			<u><i>desenrol-</i></u>	Lz
4;11.00			<u><i>descartel-</i></u>	Lz
<b>Legenda:</b>				
● <u>Composicional e HÁ presença prévia da palavra sem o prefixo (na mesma criança).</u>				
● <u>Composicional e NÃO HÁ presença prévia da palavra sem o prefixo (na mesma criança).</u>				

Nesta faixa etária também temos um total de dez dados emergentes; nove (90%) desses dados são parte da produção de Lz. e um (10%) corresponde à produção de Ar. A princípio, observando a produção de formas emergentes pela Lz., dos nove dados, cinco (55,6%) são analisáveis: *descol-*, *acalm-*, *enterr-*, *desenrol-* e *descartel-*, e quatro (44,4%) são composicionais não analisáveis: *atir-*, *desajeit-*, *adormec-* e *ensabo-*. Sobre *descartel-*, esse dado é composicional, mas não há a forma sem prefixo (*cartel-*) na produção da criança; ou seja, esse dado deveria ser classificado como um dado composicional não analisável. No entanto, por se tratar de uma forma inovadora, criada pela criança, e pelo contexto, no qual a criança usa a palavra *descartela* para pedir à mãe que *destaque as cartelas* de um jogo, podemos notar que *descartel-* é analisável. Sendo assim, excepcionalmente sem atender ao critério de ocorrência da forma não prefixada, esse dado foi classificado como composicional analisável. Em uma visão geral dos dados e em uma visão que considera apenas as formas analisáveis, podemos afirmar que o prefixo *em-* é o mais frequente e o mais produtivo.

Referente à frequência das nove formas produzidas pela Lz., seis (66,7%) são classificadas como pouco frequentes no *input*: *desenrol-* (8), *acalm-* (5), *descol-* (4), *enterr-* (4), *atir-* (3) e *adormec-* (2). Os três dados restantes, que correspondem a 33,3% do total, estão entre as formas nada frequentes: *descartel-* (1), *desajeit-* (0) e *ensabo-* (0).

Sobre a produção de Ar., o dado emergente nesse estágio é *desapert-* (presente em *desaperta*<sup>25</sup>), que é do tipo composicional analisável. Sobre a frequência, *desapert-* está entre as formas classificadas como pouco frequentes no *input*, com três *tokens*.

Em uma sucinta observação conjunta dos dados, podemos afirmar que 60% são composicionais analisáveis; 100% estão entre os dados classificados como pouco ou nada frequentes no *input*; e, por último, nessa faixa etária há o surgimento de três formas que não

<sup>25</sup> Em princípio, *desapertar* poderia ser classificada como uma produção inovadora da criança, pois é formada em lugar de *afrouxar* ou *soltar*, mas essa ocorrência específica é uma repetição única da produção da fala adulta.

estão presentes na produção adulta do nosso *corpus*: *desajeit-*, *ensabo-* e *descartel-*<sup>26</sup>. A seguir, discorreremos sobre os dados da última faixa etária (5 a 5;6 anos).

Quadro 9 – Emergência das formas prefixadas – 5 a 5;6 anos

Idade	Prefixos			Criança
	<i>a-</i>	<i>em-</i>	<i>deS-</i>	
5;01.14		<u>ensolar-</u>		Lz
<b>Legenda:</b>				
● Composicional e HÁ presença prévia da palavra sem o prefixo (na mesma criança).				

Neste último estágio, temos uma forma emergente, pertencente à produção de Lz. Essa forma é *ensolar-* (presente na palavra *ensolarado*) e se trata de um dado composicional analisável. Sobre a frequência, *ensolar-*<sup>27</sup> é classificada como nada frequente, com apenas um *token*. Vale lembrar que o número reduzido de dados nas últimas faixas etárias se deve a um número reduzido de dados em geral e não a uma diminuição na produção das crianças, como já apontamos no início do capítulo.

#### 4. Síntese e discussão dos resultados

Um dos objetivos iniciais deste artigo é o de investigar se a criança, para cada forma composicional prefixada, produz a mesma forma sem o prefixo, como indício de que ela está de fato analisando a palavra. De acordo com o que mostramos durante a descrição dos dados, existem formas com composicionalidade semântica e essas podem aparecer previamente sem prefixo na produção das crianças (classificadas como composicionais analisáveis) ou podem não aparecer previamente sem prefixo na produção das crianças (classificadas como composicionais não analisáveis).

A análise descritiva univariada (tabela 2 abaixo) mostra que foram coletados 57 dados de palavras composicionais com prefixos (4 de Am., 22 de Ar. e 31 de Lz.), entre as idades de 3;0 a 5;6 anos. Para aplicação de análise estatística a partir da quantidade de dados encontrada, foi feito o

<sup>26</sup> Visto que a forma *descartel-* é criada pela criança, a única ocorrência que há na produção adulta corresponde a uma repetição da produção infantil.

<sup>27</sup> Pelo contexto, pudemos verificar que não se trata de repetição.

agrupamento das duas primeiras faixas etárias, resultando em uma primeira faixa etária agrupada de 3;0 a 4;0 anos, e das três demais, resultando na segunda faixa etária agrupada, de 4;1 a 5;6 anos. Nesta primeira faixa agrupada, há 36 dados e na segunda há 21 dados. Com relação ao tipo de formação, ou seja, palavras com prefixos analisáveis e não analisáveis, tem-se que 30 dados foram classificados como analisáveis e 27 como não analisáveis.

Tabela 2 – Análise Descritiva Univariada

Criança	Quant.	%	Idade	Quant.	%	Palavra com prefixo	Quant.
Am	4	7.0	3;0-4;0	36	63.2	Analisável	30
Ar	22	38.6	4;1-5;6	21	36.8	Não analisável	27
Lz	31	54.4	N=	57	100	N=	57
N=	57	100					

A fim de investigar a relação entre idade e produção de dados analisáveis, procedeu-se à análise descritiva bivariada, conforme tabela 3 abaixo:

Tabela 3 – Análise Descritiva Bivariada

Criança	Idade	Palavra com prefixo		Total
		Analisável	Não analisável	
Am	3;0-4;0	1	2	3
		33.3	66.7	100.0
	4;1-5;6	0	1	1
		0.0	100.0	100.0
Ar	3;0-4;0	8	10	18
		44.4	55.6	100.0
	4;1-5;6	4	0	4
		100.0	0.0	100.0
Lz	3;0-4;0	6	9	15
		40.0	60.0	100.0
	4;1-5;6	11	5	16
		68.8	31.2	100.0
Total		30	27	57
		52.6	47.4	100.0

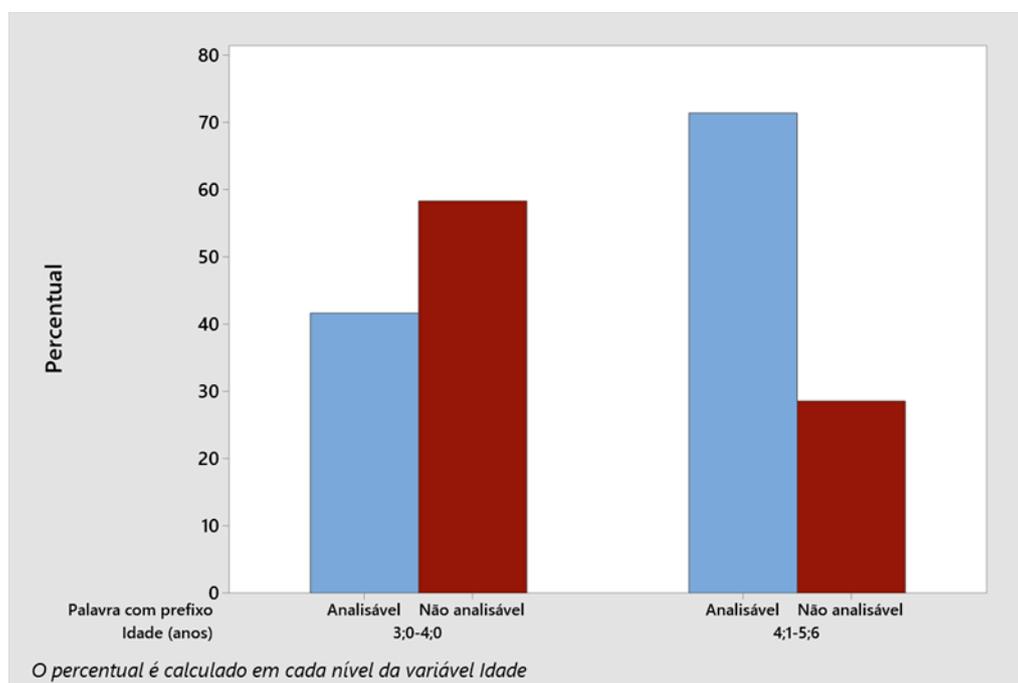
Quantidade

% da linha

Dos 57 dados com composicionalidade semântica produzidos pelas três crianças, 30, ou 52,6%, são composicionais analisáveis. Observando as diferentes faixas etárias, notamos que tanto na produção de Lz. quanto na produção de Ar., crianças com maior produção, a predominância na primeira faixa, de 3;0 a 4;0, é de formas composicionais não analisáveis; respectivamente, temos 60,0% e 55% e a quantidade de formas analisáveis é de 40% e 44,4% respectivamente. No entanto, na faixa seguinte, entre 4;1 a 5;6, o cenário muda: os dados analisáveis passam a somar 68,8% e 100% na produção de Lz. e Ar., respectivamente. O aumento da presença de dados analisáveis com a passagem da idade só não pode ser observado na produção de Am.; acreditamos que esse aumento não se evidencia na produção dessa criança pela quantidade limitada de dados, conforme vimos anteriormente.

É possível visualizar no gráfico 1, a seguir, a aparente relação entre as variáveis faixa etária e tipo de palavra com prefixo: quanto maior a idade, menor a presença de dados não analisáveis.

Gráfico 1 – Idade vs. tipo de palavra com prefixo



Para testar a hipótese de associação entre as variáveis idade e tipo de palavra com prefixo (analisável e não-analisável), aplicou-se o teste Qui-quadrado de Pearson (tabela 4)<sup>28</sup>. O resultado do teste mostrou que a relação entre essas variáveis é significativa:  $X^2(1, N = 57) = 4.7, p = .030$ .

<sup>28</sup> As análises foram desenvolvidas no Minitab® Statistical Software 19. [www.minitab.com](http://www.minitab.com).

Tabela 4 – Teste Qui-quadrado Variáveis Idade vs. Palavra com prefixo

<b>Estatística</b>	<b>Valor</b>	<b>GL</b>	<b>Valor P</b>
Pearson	4.712	1	0.030

Clark (2001) apresenta um conjunto de características relevantes sobre a aquisição da morfologia flexional, derivacional e composicional. Um dos primeiros pontos abordados pela autora diz respeito aos pré-requisitos para a aquisição da morfologia, ou seja, para a combinação de morfemas e formação de novas palavras. Segundo Clark (2001), a aquisição da morfologia envolve os seguintes passos: percepção das estruturas das palavras, identificação de bases e afixos e mapeamento de significados. A realização desses passos parece acontecer antes de a criança ser capaz de produzir os afixos, ou seja, a compreensão da morfologia está presente antes do aparecimento das primeiras produções da criança. Ao encontro dessa previsão, observamos uma escalada na produção das formas prefixais estudadas neste artigo: nas faixas iniciais, predominam as formas composicionais não analisáveis, em que os afixos que possivelmente são percebidos como parte da estrutura das palavras em algumas formações ainda não são largamente produzidos como partes de estruturas complexas do ponto de vista semântico e estrutural<sup>29</sup>. Nas faixas etárias seguintes, aumenta a produção de formas morfossemanticamente mais complexas devido a um prévio mapeamento dos significados prefixais. Tal fato fica evidente quando o prefixo *deS-* passa a ser produzido em formas inovadoras, em que é claro seu significado composicional analisável e não há ocorrência da mesma forma na fala adulta, ou seja, se descarta a produção infantil como mera repetição do *input*.

Sobre as formas prefixais, observando os dados das três crianças, o prefixo *deS-* é o que mais aparece em formas analisáveis enquanto o prefixo *a-* é o menos recorrente nessas formas, conforme mostra o quadro 10 a seguir. Observando a analisabilidade dos prefixos do modo individual, o *deS-* continua liderando como o mais analisável na produção de Lz. Por outro lado, na produção de Ar., *deS-* e *eN-* são igualmente mais analisáveis em comparação com o prefixo *a-*.

<sup>29</sup> Em conformidade com esses resultados, em Assine (2020) e Assine e Bassani (a sair), encontramos achados que mostram que a emergência de formas não composicionais precede a emergência de formas composicionais (analisáveis ou não). Além disso, as formas não composicionais que ocorrem nas primeiras faixas etárias com frequência são também muito frequentes na fala adulta.

Quadro 10 – Prefixos mais analisáveis por faixa etária

Faixa etária	Prefixo mais analisável (Lz.)	Prefixo mais analisável (Ar.)	Prefixo mais analisável (Am.)
1ª	deS-	eN-	-
2ª	deS-	eN-	-
3ª	a-/deS-	a-/deS-	-
4ª	deS-	deS-	-
5ª	eN-	-	-

Os dados inovadores e analisáveis que encontramos na produção infantil do *corpus*, *disjunt-* e *descartel-*, reforçam o caráter produtivo do prefixo *deS-*, pois, como podemos observar, eles são precedidos por esse prefixo. Essas formas aparecem nas frases do exemplo abaixo e suas ocorrências foram detalhadas anteriormente.

(1) Uso das formas *disjunt-* e *descartel-*:

a. “CHI: e não disjunta tá?”

MOT: tá eu não vou disjuntar não.” (Lz. – 4;03.12).

b. “CHI: descartela.

MOT: (es)tá.

MOT: descartela o que, Lz.?

MOT: destaca(r)?

CHI: é.” (Lz. – 4;11.00).

Outro objetivo que devemos retomar refere-se à comparação das formas emergentes na produção infantil com a frequência recebida pelo *input*, ou fala adulta. No quadro 11, abaixo, indicamos a distribuição dos tipos de frequência<sup>30</sup> das formas emergentes nas diferentes faixas etárias.

<sup>30</sup> No quadro, os tipos de frequência estão indicados pelas letras iniciais: F (frequente), PF (pouco frequente) e NF (nada frequente).

Quadro 11 – Tipo de frequência predominante nas formas emergentes

Faixa etária	Frequência dos dados											
	Lz.				Ar.				Am.			
	Total de formas	F %	PF %	NF %	Total de formas	F %	PF %	NF %	Total de formas	F %	PF %	NF %
1ª	8	63	37,5	0	10	50	40	10	0	-	-	-
2ª	7	29	71,4	0	8	25	62,5	13	3	-	100	-
3ª	6	0	83,3	17	3	33	66,7	0	1	-	100	-
4ª	9	0	66,7	33	1	0	100	0	0	-	-	-
5ª	1	-	-	100	0	-	-	-	0	-	-	-

Examinando cada uma das faixas etárias, percebemos uma maioria de dados *frequentes* na primeira faixa; na produção de Lz., 63% dos dados desse período são *frequentes* no *input* e na produção de Ar. esses dados *frequentes* na fala adulta correspondem a 50% do total da produção. Nas faixas etárias seguintes, a maioria dos dados são do tipo *pouco frequentes*. A aparição de dados *nada frequentes* na primeira faixa etária é mínima; com nenhum dado nas produções de Lz. e de Am. e com 10% dos dados na produção de Ar. Na produção de Lz., os dados *pouco frequentes* surgem na terceira faixa etária e na passagem dessa faixa para a próxima, há um aumento desses dados. Isto é, há a manifestação desses em maior quantidade com a progressão da idade. Por outro lado, na produção de Ar., os dados *nada frequentes* aparecem apenas nas duas primeiras faixas etárias, com, praticamente, a mesma fração; isto é, na produção de Ar., não há um aumento evidente da aparição desses dados com a passagem da idade. Sobre a produção de Am., conforme nos mostra o quadro acima, a emergência de dados *frequentes* e de dados *nada frequentes* é inexistente; todos os dados produzidos por essa criança estão entre os *pouco frequentes*.

Durante a descrição detalhada das faixas etárias, apresentada acima, vimos que na produção das crianças surgem algumas poucas formas que não estão presentes na produção adulta do nosso *corpus* ou que estão presentes com apenas um *token*, mas esse *token* corresponde a uma repetição que o adulto faz das formas criadas pelas crianças. Essas formas criadas pelas crianças e que possuem apenas um *token* na produção dos adultos são *descartel-* e *disjunt-*; as outras formas, sem nenhum *token*, são: *ajunt-*, *aterriss-*, *desajeit-* e *ensabo-*. Sobre a frequência, todas apresentam apenas uma ocorrência na fala infantil e, sobre a analisabilidade, *descartel-*, *disjunt-*, *ajunt-* e *aterriss-* são analisáveis. No total, são 57 formas emergentes

produzidas pelas três crianças; portanto, esses seis dados, específicos da produção infantil, correspondem a apenas 10,5% do conjunto. O contrário também ocorre. Existem formas prefixadas produzidas pelos adultos que não estão presentes entre as formas emergentes da produção infantil. Um dos objetivos deste trabalho, conforme indicamos anteriormente, é observar quais são as características dessas formas.

Na produção adulta, reunimos um total de 87 formas prefixadas com composicionalidade semântica. Desse total, 55,2%, ou 48 formas, não aparecem nos dados infantis. Referente à frequência das formas, esses dados estão entre os pouco frequentes e os nada frequentes. Entre os pouco frequentes, temos 18 dados, ou 37,5%; entre os nada frequentes, temos 30 dados, ou 62,5%. Ou seja, mais da metade dos dados que não são produzidos pelas crianças estão entre os classificados como *nada frequentes*; esses contêm apenas um *token* na produção adulta. Como mostra o quadro 11, essas formas *nada frequentes* são minoria nos intervalos observados da produção infantil. Há um aumento delas na produção de Lz., mas, ainda assim, não chegam a ser maioria. Nas produções das outras crianças, essas formas *nada frequentes* não aparecem, como é o caso na produção de Am.; ou aparecem pouco, como é o caso no conjunto de dados de Ar. Sobre os prefixos, 48% ou 23 dessas formas são precedidas pelo *deS-*, que, na produção infantil, é o prefixo que mais acompanha formas analisáveis. A outra parte das formas se divide entre os prefixos *a-* e *eN-*; o primeiro está em 13 formas e o segundo está em 12.

Tendo em vista o que apresentamos no parágrafo anterior, é possível afirmar que a maior parte das formas presentes na fala adulta que não são produzidas pelas crianças apresenta frequência muito baixa (uma ocorrência) e presença do prefixo *deS-*. Sabemos que a produtividade do *deS-* é um aspecto que existe em comum entre os dados infantis e os dados da produção adulta. Dessa maneira, o fator que pode implicar na ausência de certas formas na fala das crianças é a baixa frequência de uso pelos adultos, comentada aqui; outros possíveis fatores são a própria aquisição de vocabulário, que está sempre em curso, bem como as necessidades contextuais.

## 5. Conclusão

O objetivo geral deste estudo era descrever a emergência dos prefixos *a-*, *eN-* e *deS-* na produção de crianças falantes de português brasileiro com foco nas formas analisáveis. A partir das nossas observações, chegamos a alguns resultados que podem ser relevantes para o campo de

estudos em aquisição de morfologia.

Primeiramente, concluímos que a emergência de formas analisáveis tende a aumentar na produção infantil com o avanço da idade. Para testar a hipótese de associação entre as variáveis idade e tipo de palavra com prefixo (analisável e não-analisável), aplicou-se o teste Qui-quadrado de Pearson. O resultado do teste mostrou que a relação entre essas variáveis é significativa ( $p = .030$ ). Isto é, a produção de formas morfossemanticamente mais complexas aumenta a devido a um prévio mapeamento dos significados prefixais. Tal fato fica evidente quando o prefixo *deS-* passa a ser produzido em formas inovadoras, em que é claro seu significado composicional analisável e não há ocorrência da mesma forma na fala adulta, ou seja, se descarta a produção infantil como mera repetição do *input*. Essas formas inovadoras são *disjunt-* e *descartel-*, produzidas, respectivamente, na terceira e na quarta faixa etária. Conforme vimos, dentre os prefixos estudados, o *deS-* é o mais analisável. Ainda sobre as formas analisáveis presentes na produção infantil, podemos afirmar que a maior parte delas estão entre os dados *pouco frequentes no input*; isso significa que alta frequência e analisabilidade não são aspectos dependentes.

Sobre a frequência do *input*, vimos que na primeira faixa etária os dados emergentes predominantes são aqueles produzidos frequentemente pelos adultos; com o avanço da idade, o cenário é invertido e os dados que são *pouco frequentes no input* se tornam maioria na produção infantil. Sendo assim, podemos concluir que, além das formas analisáveis, formas *pouco frequentes no input* aumentam com a passagem da idade na produção das crianças. Sobre as formas *nada frequentes*, apesar de surgirem seis dados que não são produzidos pelos adultos (*desajeit-*, *ensabo-*, *disjunt-*, *ajunt-*, *aterriss-* e *descartel-*), a aparição dessas formas é mínima na produção infantil e corresponde a 10,5% do total da produção. A respeito dos dados da fala adulta que não foram produzidos pelas crianças no período estudado, vimos que a maior parte corresponde a formas *nada frequentes*, isto é, com apenas uma ocorrência, e o prefixo predominante é o *deS-*. Desse modo, concluímos que o fator que pode implicar na ausência de certas formas na fala das crianças é a baixa frequência de uso pelos adultos, comentada aqui; outros possíveis fatores são a própria aquisição de vocabulário, que está sempre em curso, bem como as necessidades contextuais.

Por fim, consideramos como possibilidade futura de pesquisa e como continuidade a este estudo a aplicação de um experimento, com base nas formas prefixadas encontradas, a fim de confirmarmos quais dados de fato são analisados pelas crianças.

## Referências

- ASSINE, J. S. *A emergência de prefixos na produção de crianças durante a aquisição de português brasileiro*. 2020. 111 f. Dissertação (Mestrado em Letras) - Escola de Filosofia, Letras e Ciências Humanas, Universidade Federal de São Paulo. Guarulhos, 2020.
- ASSINE, J. S.; BASSANI, I. S. Morfologia interna e externa na aquisição de prefixos no português brasileiro. *Revista do GELNE*, (a sair).
- BARBOSA, M. F. M.; GOMES, C. A. Aquisição de palavras complexas no português brasileiro: a emergência de morfologia derivacional na fala infantil. *Cadernos do CNLF (CiFEFil)*, v. 18, p. 203-218, 2014.
- BASSANI, I.S. *Uma abordagem localista para morfologia e estrutura argumental dos verbos complexos (parassintéticos) do português brasileiro*. 2013. 382f. Tese (Doutorado) – FFLCH, Universidade de São Paulo, São Paulo, 2013.
- BASSANI, I. S.; MINUSSI, R. D. Contra a seleção de argumentos pelas raízes: nominalizações e verbos complexos. *Revista Virtual de Estudos da Linguagem*, v. 13 , p. 139-173, 2015.
- CLARK, E. Morphology in Language Acquisition. In: Spencer; Andrew; Arnold M. Zwicky. (org.). *The Handbook of Morphology*. Blackwell Publishing, 2001.
- FERRARI NETO, J. Passos em direção a uma teoria da aquisição da morfologia. In: TAVEIRA DA CRUZ, R. (org.). *As interfaces da gramática*. 1. edição. Curitiba: Editora CRV, 2012, v. 1, p. 215-239.
- HALLE, M.; MARANTZ, A. Distributed Morphology and the pieces of inflection. In: HALE, K.; KEYSER, J. *The View from Building 20*. Cambridge: MIT Press, 1993, p. 111-176.
- HARLEY, H. On the identity of roots. *Theoretical Linguistics: An Open Peer Review Journal*, v.40, Issue 3-4, p.225-276, 2014.
- MARANTZ, A. Phases and words. Manuscrito. NYU, 2007.
- MARCUS, G. F. et al. Overregularization in Language Acquisition. *Monographs of the Society for Research in Child Development*, v. 57, n. 4, 1992. Disponível em <[www.jstor.org/stable/1166115](http://www.jstor.org/stable/1166115)>. Acesso em: 16 abr. 2020.
- MEDEIROS, A. B. Para uma abordagem sintático-semântica do prefixo des-. *Revista da ABRALIN*, v. 9, p. 95-121, 2010.
- MINUSSI, R. D.; BASSANI, I. S. Em favor do conteúdo semântico das raízes. *REVISTA LETRAS*, v. 96, p. 152-173, 2017.
- LIMA, P. A. N. *Morfemas derivacionais e compostos do português brasileiro na fala de crianças de dois a sete anos de idade*. 2006. 91f. Dissertação (Mestrado) – UFRGS, Porto Alegre, 2006.
- PINKER, S; PRINCE, A. Regular and irregular morphology and the psychological status of rules of grammar. In: S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.). *The reality of linguistic rules*.
-

Philadelphia: Benjamins, 1994.

RAVID, D. First-Language Acquisition of Morphology. *Oxford Research Encyclopedia of Linguistics*, 2019. Acesso em 14 de maio de 2020 em <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-603>.

SCHWINDT, L. C. Produtividade, transparência e estatuto prosódico de palavras derivadas por prefixação em português brasileiro e espanhol peninsular. *Organon* (UFRGS), Porto Alegre, v. 18, n. 36, p. 131-137, 2004.

SILVA, M. C; MIOTO, C. Considerações sobre a prefixação. *ReVEL*, v.7, n. 12, 2009.

---