

RELATO DE PESQUISA

# Pistas estatísticas e influências morfosintáticas na aquisição de perguntas com sintagma-QU em português brasileiro

Clariana Lara VIEIRA 

Universidade de São Paulo (USP)

Elaine Bicudo GROLLA 

Universidade de São Paulo (USP)



OPEN ACCESS

EDITADO POR

- Raquel Freitag (UFS)

AVALIADO POR

- Livia Oushiro (Unicamp)

- Ronaldo Manguiera Lima Jr. (UnB)

SOBRE OS AUTORES

- Clariana Lara Vieira

Conceptualização, Curadoria de Dados, Análise Formal, Investigação, Metodologia, Escrita – rascunho original.

- Elaine Bicudo Grolla

Conceptualização, Supervisão, Escrita – rascunho original.

DATAS

- Recebido: 03/06/2025

- Aceito: 04/06/2025

- Publicado: 03/04/2026

COMO CITAR

VIEIRA, C. L.; GROLLA, E. B.. (2026).

Pistas estatísticas e influências morfosintáticas na aquisição de perguntas com sintagma-QU em português brasileiro. *Revista da Abralín*, v. 24, n. 1, p. 1-33, 2025.

RESUMO

Em português brasileiro (PB), o movimento-QU em perguntas de constituinte é supostamente opcional. Se esse for o caso, é de se esperar que a estratégia em que o elemento QU permanece *in situ* seja mais frequente no início do desenvolvimento linguístico infantil, já que, de acordo com a análise adotada aqui, ele demandaria menos operações derivacionais. Sikansi (1999) e Grolla (2000, 2009) apontam, todavia, que a criança raramente produz a construção, apesar de ela ser frequente na fala adulta (Lopes-Rossi, 1996; Oushiro, 2010). Em vista dessa assimetria, o presente estudo visa a examinar a produção do QU-*in situ* por adultos e crianças a partir da coleta de dados naturalísticos. O *corpus* analisado inclui cerca de 154 horas de gravação, contendo perguntas-QU produzidas por 4 crianças em processo de aquisição de PB como primeira língua, com idades entre 1;02.28 e 5;06.07, e 23 adultos com os quais elas interagiam (*corpus* Santana-Santos). Os resultados sugerem que a estrutura não é tão produtiva na fala dirigida à criança como na fala entre adultos e indicam que a criança não se guia apenas pelas pistas estatísticas deixadas pelo *in-put*: embora a estratégia QU-*in situ* compreenda menos de 2% de sua produção, sua distribuição reflete também restrições de natureza gramatical e pragmática.

RESUMO EM OUTRA LÍNGUA

In Brazilian Portuguese (BrP), WH-fronting in questions is assumed to be optional. If this assumption holds true, we would expect the WH-*in situ* strategy to be more frequent at early stages of children's linguistic development, since, according to the

analysis adopted here, it would require fewer derivational steps. Sikansi (1999) and Grolla (2000, 2009), however, point out that children rarely produce this construction, despite it being frequent in adult speech (Lopes-Rossi, 1996; Oushiro, 2010). Given this asymmetry, the present study aims to examine the production of *WH-in situ* questions by adults and children using naturalistic data. The analyzed *corpus* includes approximately 154 hours of recordings, containing *WH*-questions produced by four children acquiring BrP as their first language, aged between 1;02.28 and 5;06.07, as well as 23 adults interacting with them (Santana-Santos *corpus*). The results suggest that the *WH-in situ* structure is not as productive in child-directed speech as it is in adult-to-adult interactions, and indicate that children are not guided solely by statistical cues provided by the input: although the *WH-in situ* strategy comprises less than 2% of their production, its distribution also reflects grammatical and pragmatic constraints.

## PALAVRAS-CHAVE

Aquisição de linguagem. *QU-in situ*. Dados espontâneos. Português brasileiro. Pistas estatísticas.

## PALAVRAS-CHAVE EM OUTRO IDIOMA

Language acquisition. *WH-in situ*. Spontaneous data. Brazilian Portuguese. Statistical cues.

## RESUMO PARA NÃO ESPECIALISTAS

No português brasileiro (PB), perguntas com elemento interrogativo (o que, como, onde etc) podem ser feitas de duas maneiras: movendo a palavra interrogativa para o início (“O que o gato comeu?”) ou deixando-a na posição original da sentença (“O gato comeu o quê?”). A segunda opção, chamada *QU-in situ*, seria supostamente mais econômica, por não exigir movimento, e poderia ser preferida pelas crianças que estão adquirindo linguagem. No entanto, pesquisas mostram que, embora adultos usem bastante essa estratégia, crianças raramente a utilizam. Este estudo analisou 154 horas de gravações com quatro crianças brasileiras, entre 1 e 5 anos, e 23 adultos interagindo com elas, para investigar a questão. Os resultados sugerem que o *QU-in situ* não é tão frequente na fala adulta e que as crianças parecem seguir os adultos usando pouquíssimo (menos de 2%) dessa forma mais simples. Isso sugere que as crianças optam pela estratégia que mais ouvem, mesmo sendo uma construção mais complexa linguisticamente.

## Introdução

Em português brasileiro (doravante PB), as perguntas-QU são realizadas com a permanência do constituinte-QU *in situ* (1a) ou com seu movimento para a periferia esquerda (1b), opção que denominaremos QU-*ex situ*<sup>1</sup> neste artigo:

- (1) a. A Gabriela viu o quê?  
 b. O que a Gabriela viu \_\_\_?

Semanticamente, tanto o QU-*ex situ* como o *in situ* são possíveis e aceitos como pedidos genuínos de informação e a posição sintática do elemento-QU, seja qual for, não parece acarretar implicações para a denotação das interrogativas. Ambas as estruturas, segundo Krifka (2011, p. 1744, tradução nossa), “criam uma proposição aberta deixando partes de sua descrição não especificadas”, obrigando o interlocutor a preencher as lacunas deixadas pelo constituinte-QU. Sintaticamente, contudo, essa opcionalidade suscita questões em uma abordagem minimalista (Chomsky, 1995, 1999), onde o movimento de sintagmas e núcleos ocorre de maneira restrita, apenas quando necessário<sup>2</sup>. Seria interessante explicar, então, como deve ser o tratamento de línguas que permitem tanto o QU-*ex situ* como o *in situ*: nesses casos<sup>3</sup>, como o do PB, estamos realmente diante de uma opcionalidade? Argumentaremos aqui que as duas estratégias não são totalmente opcionais. Fatores morfossintáticos e pragmáticos entram em jogo, determinando, em muitos casos, a escolha por uma ou outra estratégia.

Além disso, do ponto de vista do desenvolvimento linguístico, a literatura tem indicado que, em línguas de movimento-QU obrigatório e em línguas de movimento-QU proibido, as crianças geralmente não apresentam dificuldades em discernir, desde os estágios iniciais da aquisição, as permissões de sua gramática alvo. As crianças adquirindo inglês (Guasti, 2014) ou chinês (Chang, 1992), por exemplo, aparentam detectar precocemente se sua língua exige ou proíbe, respectivamente, o movimento do elemento-QU. Todavia, em línguas de movimento-QU opcional, essa definição depende de um *input* “variável”, no qual a criança é exposta tanto a perguntas QU-*ex situ*

<sup>1</sup> A literatura recente tem utilizado as denominações QU-*ex situ* para estruturas como “O que; o menino viu \_\_\_?”, em que o constituinte-QU aparece deslocado à esquerda, contrapondo-se à estrutura QU-*in situ*, “O menino viu o quê?”, sem movimento aparente da palavra-QU. Aqui adotaremos essa terminologia para facilitar a discussão, tendo em vista que estruturas *ex situ* podem assumir diversas formas em PB, como ilustrado abaixo:

- (i) O que; o menino viu ti? (QU-movido simples)  
 (ii) O que; que o menino viu ti? (QU-que)  
 (iii) O que; é que o menino viu ti? (QU-é-que)

Acompanhamos Mioto (1994), Sikansi (1999) e Grolla (2009) na análise do QU-movido simples (a), QU-que (b) e QU-é-que (c):

- a. QU-movido simples: [CP O que; [IP o João comprou ti]]  
 b. QU-que: [CP O que; [C que ] [IP o João comprou ti]]  
 c. QU-é-que: [CP2 O que; C [IP pro é [CP1 ti [C que][IP o João comprou ti]]

(Grolla, 2009, p. 86-87)

<sup>2</sup> Com base na teoria de Princípios e Parâmetros (Chomsky, 1981), os sintagmas interrogativos são licenciados via movimento para a periferia esquerda. Esse movimento, grosso modo, poderia ocorrer de forma aberta, gerando o QU-*ex situ*, ou coberta, i.e., em Forma Lógica (LF), de modo que seus efeitos não sejam lidos pela interface fonética, gerando o QU-*in situ* (Huang, 1982).

<sup>3</sup> Outras línguas que apresentam tanto QU-*in situ* como QU-*ex situ* incluem: francês (Zuckerman, 2001); espanhol (Jiménez, 1997; Uribe-Etxebarria, 2002; Reglero, 2004; Biezma, 2018, a.o.); dialetos do italiano (Munaro; Poletto; Pollock, 2001; Poletto; Pollock, 2015; Bonan, 2021, a.o.); e grego (Vlachos; Chiou, 2021, entre outros).

como QU-*in situ*. No PB adulto, estudos empíricos têm apontado uma produtividade de ambas as estratégias (Lopes-Rossi, 1996; Oushiro, 2012). Já na fala infantil, observamos um quadro curiosamente diferente: em estudos disponíveis na literatura, o QU-*in situ* foi raramente (ou nunca) produzido em dados espontâneos (Sikansi, 1999; Grolla, 2000) e as escassas ocorrências emergiram tardiamente, se comparadas a outras construções A-barra bem mais sofisticadas (Grolla, 2009; Silveira, 2011). Portanto, cabe investigar por que as crianças tendem a evitar a construção supostamente mais econômica, que não envolve movimento aparente, distanciando-se do padrão observado em interações entre adultos.

Para dar conta das assimetrias encontradas no PB, neste trabalho, defenderemos que o QU-*ex situ* e o QU-*in situ* não são alternativas em competição, estando associadas a diferenças cruciais, mesmo que sutis, que as distanciam e têm impacto para a aquisição de linguagem. Ainda mais marcante para o processo de aquisição parece ser a diferença entre a fala entre adultos e a fala dirigida à criança (*child-directed speech*, CDS) no tocante à produção de perguntas-QU. Essas diferenças deixam claro que não é possível comparar estudos analisando a fala infantil àqueles que se baseiam em conversas entre adultos (Lopes-Rossi, 1996 e Oushiro, 2012), pois um *corpus* dessa natureza não equivaleria a dados de um *input* em potencial. Por isso, esperamos encontrar uma medida mais adequada para comparação, observando as produções infantis e produções de adultos em interações com crianças.

Este artigo se organiza da seguinte forma: na seção 1, discutiremos os aspectos morfossintáticos que regem a distribuição do QU-*in situ* no PB; na seção 2, exploraremos seus aspectos pragmáticos; na seção 3, abordaremos estudos pertinentes sobre a aquisição de QU-*in situ*, considerando aqueles baseados em dados espontâneos e experimentais; na seção 4, descreveremos os dados da pesquisa e a metodologia utilizada na coleta; na seção 5, apresentaremos os resultados obtidos, examinando a influência dos fatores morfossintáticos e pragmáticos. Seguiremos, então, a uma discussão dos achados, que, em linhas gerais, sugerem que o comportamento infantil é marcado pela interação entre a frequência do *input* e uma sensibilidade precoce aos aspectos morfossintáticos da língua; por fim, concluiremos o trabalho.

## 1 QU-*ex situ* vs. QU-*in situ*: aspectos morfossintáticos

Em sua análise das perguntas-QU para o PB, Hornstein, Nunes & Grohmann (2005) enumeram os contextos que permitem ou não o movimento-QU: (i) o deslocamento do elemento-QU é obrigatório com complementizador interrogativo aberto (2), em interrogativas indiretas (3) e com elementos-QU inerentemente *non-D-linked*, como a expressão “que diabo” (4); em contrapartida, (ii) o deslocamento é proibido se ilhas sintáticas intervêm (5):

- (2) a. Como que você consertou o carro?  
b. \*Que você consertou o carro como?
  
- (3) a. Eu perguntei como (que) você consertou o carro.  
b. \*Eu perguntei (que) você consertou o carro como.

- (4) a. Que diabo você bebeu?  
b. \*Você bebeu que diabo?<sup>4</sup>
- (5) a. \*Que livro você conversou com o autor [que escreveu]?  
b. Você conversou com o autor [que escreveu que livro]?

Exemplos retirados de Hornstein, Nunes & Grohmann (2005, p. 42-43)

Figueiredo Silva & Grolla (2015) também destacam a obrigatoriedade do QU-*in situ* nos seguintes casos: (i) sentenças com sujeito nulo dêitico ou anafórico (6); e (ii) estruturas nas quais não há uma projeção funcional à qual o elemento-QU possa ser movido (7):

- (6) A: Eu falei...Eu falei...  
B: Falou o quê?  
B': \*O que falou?

- (7) a. O dodói de quem?  
b. \*De quem o dodói?

Exemplos retirados de Figueiredo Silva & Grolla (2015, p. 264)

Além desses contextos, notamos também que o QU-*in situ* é a única estratégia possível quando há apenas um fragmento da sentença, geralmente copiada da estrutura anterior (8):

- (8) a. Criança: eu vou falar. (corpus Leonardo, 4;05.11)  
Mãe: falar o quê?
- b. Criança: mamãe, toma! (corpus Maria, 1;10.24)  
Mãe: toma o quê?

Ademais, as sentenças em (6), (7) e (8) parecem possuir (e aqui emprestamos um termo usado por Ambar (2000)) um “sabor de eco”, pois a inclusão de um argumento nulo na sentença imediatamente anterior gera uma estrutura similar a uma pergunta-eco. Em outras palavras, esses casos são intrinsecamente dependentes de um antecedente imediatamente anterior e pedem um preenchimento ou esclarecimento do argumento nulo da sentença (para ampliar a discussão sobre essas construções, ver Ambar (2000), Chang (1997) e Vieira (2023)).

---

<sup>4</sup> Apesar de a construção com elementos non-D-linked *in situ*, como “você bebeu que diabo?”, ser considerada inaceitável no dialeto paulista, Lessa-de-Oliveira (2003) aponta sua aceitabilidade no dialeto baiano de Vitória da Conquista. Portanto, existe uma variação dialetal na aceitação dessa expressão em posição *in situ* na língua.

Considerando os casos acima, sintaticamente, há aqueles em que o movimento-QU é obrigatório (2, 3 e 4), aqueles em que é proibido (5, 6, 7 e 8) e aqueles em que as duas estratégias parecem opcionais (1). No entanto, conforme veremos na próxima seção, mesmo em contextos sintáticos que permitiriam as duas formas interrogativas, como (1), algumas restrições pragmáticas se impõem, distanciando mais uma vez os dois tipos de pergunta.

Neste artigo, assumiremos, junto com DeRoma (2011) e Lacerda (2020), que os constituintes-QU são formados pelo sintagma-QU e um operador Q, não adjacentes em PF. O sintagma interrogativo – considerado uma instância de foco (Rizzi, 1997) – não sofre movimento coberto para uma projeção de foco, mas é ligado por um operador coberto, via *Unselective binding*<sup>5</sup>. Isso permite que o sintagma-QU seja licenciado por seu operador, mesmo dentro de ilhas (5), como ilustrado nos esquemas abaixo (com diferentes níveis de encaixamento) retirados de Lacerda (2020:156):

- (9) a. [<sub>CP</sub> OP-FOC<sub>i</sub> [<sub>TP</sub> ... FOCUS<sub>i</sub>]]  
 b. [<sub>CP</sub> OP-FOC<sub>i</sub> [<sub>TP</sub> ... [<sub>CP</sub> ... FOCUS<sub>i</sub>]]]  
 c. [<sub>CP</sub> OP-FOC<sub>i</sub> [<sub>TP</sub> ... [<sub>Island</sub> ... FOCUS<sub>i</sub>]]]

O resultado da associação entre o constituinte-QU e o foco reflete diretamente no *status* sintático e pragmático do constituinte-QU. Sintaticamente, ele estará associado à projeção FocP na periferia esquerda. Dado que não há movimento sintático envolvido, isso implica em uma construção mais econômica que sua versão *ex situ*. Conforme veremos na próxima seção, isso terá consequências para a pragmática das perguntas com QU *in situ*.

## 2 QU-*ex situ* vs. QU-*in situ*: aspectos pragmáticos

Diversos trabalhos discutiram, em línguas de movimento-QU opcional, a percepção de que o QU-*in situ* e o *ex situ* diferem quanto ao grau de pressuposição exigido no aporte contextual (Chang, 1997; Boeckx, 1999; Hamlaoui, 2011; Baunaz; Bonan, 2023; Pires; Taylor, 2007; Oushiro, 2010; DeRoma, 2011; Vieira; Grolla, 2020; entre outros). Por “pressuposição”, tomemos a definição de Stalnaker (1978, p. 321), segundo a qual ela é aquilo “que é tomado pelo falante como *Common Ground* entre os participantes da conversa, o que é tratado como conhecimento comum ou mútuo”<sup>6</sup>. Ou seja, é a informação discursiva ou extralinguística que o falante assume como parte do fundo conversacional compartilhado entre os participantes da conversa. Parece haver um consenso nesses trabalhos de que o QU-*in situ* precisa ser licenciado por um contexto específico, ou um *Common Ground* (doravante CG) entre

<sup>5</sup> Na proposta de Reinhart (1998), um sintagma-QU, em uma pergunta com QU-*in situ*, é ligado em sua própria posição canônica por meio de um operador-Q e essa ligação é realizada via *Unselective binding*. Em seu sistema, o operador-Q ligaria todas as variáveis *in situ* de mesma função de forma não seletiva.

<sup>6</sup> Neste trabalho, estamos lidando com uma abordagem pragmática do conceito de pressuposição, em que “uma enunciação A pressupõe pragmaticamente uma proposição C se e somente se A for adequada apenas no caso de B ser mutuamente conhecida pelos participantes” (Levinson, 2007, p. 260). Nesse caso, a relação pela qual se preza é a que se estabelece entre os participantes da conversa e, por isso, uma informação pressuposta é aquela que todos os membros da conversa compartilham entre si e sabem que compartilham entre si. Logo, uma falha na pressuposição pragmática resulta na infelicidade da sentença naquele determinado contexto.

os participantes da conversa, enquanto o *ex situ* é mais livre para ocorrer em quaisquer contextos, mesmo em situações *out-of-the-blue*, em que a princípio não haveria compartilhamento de informação prévia com relação à pergunta realizada. No exemplo (10), a criança e o avô recordam um momento em que observaram uma garça mergulhando na água para pescar. A pergunta com QU-*in situ* é viável porque ambos possuem a informação de que “a garça estava fazendo algo” na situação que vivenciaram:

(10) Avô: como que chamava aquela aquela ave? (corpus Túlio, 3;01.16)

Criança: humm?

Avô: garça!

Avô: a garça estava lá fazendo o que?

Criança: xxx [= incompreensível] peixinho.

Avô: e ela tchibum dentro...pegava o peixe com a boca e...

Criança: comeu.

Entre os diversos autores que já trataram do tema, ainda não há consenso quanto ao que deve fazer parte do *background* conversacional para que o uso do QU-*in situ* seja apropriado. Para Pires & Taylor (2007, p. 5), por exemplo, o CG deve conter um conjunto com as possíveis respostas à pergunta, definidas discursiva ou extralinguisticamente. Já para DeRoma (2011, p. 15) as respostas possíveis não são tão cruciais como a presença (mesmo que pressuposta) da porção não-QU da própria pergunta. Para a autora, exemplos como em (11) enfraquecem a análise de Pires & Taylor, visto que não se pode recuperar, a partir do contexto, um conjunto de valores possíveis que sirvam como resposta à pergunta da mãe em (11a) ou de A em (11b):

(11) a. Filha: Ah, mãe, não vou lá não!

Mãe: Mas não vai por quê?

Filha: Que mico!

b. [A começa a falar com B e vê B fazendo as malas]

A: Tônia, tô indo almoçar. Que que é isso? Vai para onde?

B: Búzios.

Exemplos retirados de DeRoma (2011, p. 113)

Por outro lado, em (11a) a porção não-QU da pergunta “mas não vai \_\_\_” já foi inserida no discurso anterior pela filha, quando esta diz “não vou lá não”; e em (11b), a porção não QU “vai para \_\_\_” já está pressuposta no contexto, dado que A vê B fazendo as malas e, usualmente, uma pessoa faz as malas quando planeja ir para algum lugar. Para DeRoma, o QU-*in situ* é também menos aceitável em contexto *out-of-the-blue*, como em (12), em que o interlocutor B nega a pressuposição de que “haverá uma eleição”, causando o estranhamento que se observa.

(12) A: Você vai votar em quem?

B: # Tem eleição?

Exemplo retirado de DeRoma (2011, p. 120)

É necessário notar, contudo, que essa construção seria possível caso A e B estivessem em época de eleições no Brasil, e B tivesse mais de 18 anos e, portanto, fosse obrigado a votar. O uso do QU-*in situ* seria, nesse caso, aceitável pois o contexto cultural o permitiria. Esse é um dos pontos levantados por Oushiro (2010), para quem a construção depende de um *Common Ground*, que, além de discursivo e extralinguístico, também pode conter pressuposições de caráter cultural ou universal. Essas, por sua vez, permitiriam perguntas como (12) feitas de supetão, pois estariam ancoradas a informações compartilhadas culturalmente.

Precisamos, todavia, ter em mente que, em se tratando de uma abordagem pragmática, o uso da estratégia com QU-*in situ* em contextos *out-of-the-blue* nunca será impossível ou resultará em falsidade da proposição, mas apenas será menos aceitável que a opção *ex situ*. Em (13) o falante parece tomar como certo que seu interlocutor “comeu alguma coisa hoje” ou “foi para algum lugar” e ele espera que ambos compartilhem essas informações.

(13) a. Você comeu o quê hoje?

b. Você foi para onde hoje?

Em síntese, nenhuma das análises discutidas parece explicar completamente o uso do QU-*in situ* no PB. A proposta de Pires & Taylor (2007) parece ser limitadora para abarcar o fenômeno na língua. Por outro lado, a análise de DeRoma (2011), que sugere que a parte não-QU deve estar explícita ou pressuposta no discurso para que o QU-*in situ* seja viável, à primeira vista, parece bastante adequada, pois abrange uma ampla gama de ocorrências da interrogativa no português brasileiro. No entanto, ela falha ao lidar com casos de pressuposição universal e cultural, a que Oushiro (2010) nos chama a considerar. Combinando as análises anteriores, temos agora três categorias que, pensamos, podem abarcar os contextos que licenciam QU-*in situ* no PB: (i) contexto extralinguístico, (ii) antecedente discursivo e (iii) conhecimento prévio, seja ele proveniente de interações vivenciadas pelos sujeitos envolvidos em situações anteriores à conversa, ou de pressupostos universais e culturais.

Por fim, consideremos novamente a análise de DeRoma (2011), atualizada por Lacerda (2020), para os quais temos uma associação do elemento-QU à projeção de Foco que, no QU-*in situ*, poderia ser realizada via *Unselective Binding*. Nesse contexto, um sintagma-QU funcionaria como uma instância de foco, enquanto a parte não-QU poderia ser informação já conhecida (nos casos de foco estreito) ou nova (nos casos de foco largo). A principal distinção entre as duas formas é que o QU-*in situ* ocorreria apenas junto à informação dada, enquanto a opção *ex situ* poderia também ser utilizada em contextos de informação totalmente nova, produzida em contexto *out-of-the-blue* (Hamlou, 2011).

### 3 Aquisição do QU-*in situ* em PB

Conforme discutido na introdução, no PB observa-se uma disparidade entre os dados de adultos e crianças em relação à produção do QU-*in situ*. Nos dados adultos, por exemplo, Lopes-Rossi (1996, p. 68) reportou, para um *corpus* televisivo, uma frequência de 32,4% de QU-*in situ* contra 67,6% de estratégias com QU-*ex situ* (de um total de 279 ocorrências) e, para o *corpus* do NURC/SP, uma frequência de 12,5% de QU-*in situ* contra 87,5% de perguntas com QU-*ex situ* (das 72 ocorrências)<sup>7</sup>. Já Oushiro (2012, p. 57) encontrou 22,7% de QU-*in situ* e 77,3% de QU-*ex situ* em um *corpus* sociolinguístico com 53 gravações do português paulistano contendo 999 perguntas-QU. Esses dados, exceto talvez os do NURC/SP (Lopes-Rossi, 1996), sugerem uma utilização relativamente frequente do QU-*in situ* na fala adulta, o que, como veremos, não se reflete na fala infantil. No que se segue, discutimos trabalhos anteriores da literatura em aquisição baseados em dados espontâneos (seção 3.1) e experimentais (seção 3.2).

#### 3.1 Dados espontâneos

Analisando os dados longitudinais de uma criança (G.) entre 2;4 e 3;10 anos, Sikansi (1999, p. 99) não registrou nenhuma ocorrência de QU-*in situ*. Em contraste, o QU-movido<sup>8</sup> foi utilizado em 97,5% dos casos e o QU-(é)-que em 2,5% (de um total de 124 ocorrências) a partir de 2;08 anos. Na amostra de fala adulta analisada pela autora, ela encontrou 22,5% de QU-movido, 73,75% de QU-(é)-que e 3,75% de QU-*in situ*. Para ela, a ausência de QU-*in situ* na fala de G. sugere que essa estrutura é adquirida posteriormente pela criança. Grolla (2000, p. 41), por sua vez, observou os dados espontâneos de uma criança (N.) com idade entre 2;0 e 4;0 anos e, das 520 perguntas-QU produzidas por ela, apenas 1,7% das ocorrências foi com QU-*in situ* (8 ocorrências), enquanto o QU-movido foi realizado em 66% dos casos (344 ocorrências), o QU-que em 29,4% (153 ocorrências) e o QU-é-que em 2,9% (15 ocorrências). A autora também notou que a emergência do QU-*in situ* nos dados infantis ocorreu tardiamente, com as poucas ocorrências aparecendo apenas a partir de 3;9 anos, apesar de a criança já demonstrar domínio de construções mais complexas a partir de 2 anos e meio, como estruturas de tópico-comentário, sentenças clivadas e orações relativas.

Lessa-de-Oliveira (2003) analisou as perguntas-QU de duas crianças em processo de aquisição do dialeto falado em Vitória da Conquista/Bahia, com idades entre 18 e 30 meses, e observou um padrão significativamente diferente do encontrado em São Paulo. Enquanto no dialeto paulista, observado por Grolla (2000) e Sikansi (1999), as crianças produziram o QU-*in situ* tardiamente, no dialeto analisado por Lessa-de-Oliveira, a construção emergiu nos dados entre 1;07 e 1;09. Para a autora, a frequência com que a construção é usada no *input* desempenha papel

---

<sup>7</sup> Os dados analisados por Lopes-Rossi (1996) foram coletados a partir de falas espontâneas em programas de televisão, incluindo entrevistas e debates ao vivo, além de gravações de oito informantes do Projeto NURC/SP.

<sup>8</sup> Na descrição dos trabalhos, optamos por manter a nomenclatura original utilizada pelos autores. Lembremos, no entanto, que a opção convencionalizada por parte da literatura como QU-movido (e.g. "O que o gato fez\_\_?") aqui estamos chamando de QU-movido simples para não haver confusões com as estratégias QU-que e QU-é-que, que também envolvem deslocamento do elemento-QU.

fundamental na definição da ordem de emergência das estratégias, tendo em vista que o QU-*in situ*, primeira forma interrogativa empregada por L., foi também a forma mais frequente em seu *input*, totalizando 81,7%.

Embora concordemos que o *input* desempenha um papel essencial nesse contexto e pretendamos investigar essa hipótese mais adiante, essa não é a única consideração a ser feita. A relação entre *input* e fala infantil nem sempre é linear<sup>9</sup>: as crianças analisadas por Sikansi (1999) e Grolla (2009), por exemplo, empregaram mais o QU-movido, mesmo com uma maior presença de QU-que em seu *input*. De fato, até mesmo os dados de Lessa-de-Oliveira (2003) corroboram essa ideia: no *input* de uma das crianças, o QU-*in situ* foi a estrutura mais utilizada (em 53,5% dos casos), mas não foi a primeira interrogativa a emergir na fala infantil.

Grolla (2009), mais tarde, observou a fala espontânea de mais uma criança com idade entre 1;10 e 5;6 anos e adquirindo o dialeto paulistano do PB. Os resultados sugerem novamente uma baixa frequência do QU-*in situ* no *corpus* infantil, totalizando 1,7% (das 412 perguntas-QU), enquanto em seu *input*, a construção foi empregada em 9,3% (das 719 perguntas-QU). Ademais, a estratégia foi novamente a última a emergir: aos 3;11 anos, enquanto o QU-movido surgiu já aos 2;0 anos. Essa aquisição tardia foi, mais tarde, corroborada por Silveira (2011), que observou amostras transversais de 17 crianças e longitudinais de 6 crianças, com faixa etária entre 1;10 e 5;0 anos de idade. O autor notou que perguntas com QU-*in situ* emergiram apenas após todas as outras estratégias de pergunta e estruturas A-barras mais sofisticadas. Seguindo Grolla (2000, 2009), Silveira sugere que a aquisição de perguntas-QU pelas crianças segue a seguinte ordem: “QU-movido > QU-que > QU-é-que/QU-*in situ*” (Silveira, 2011, p. 185).

Em suma, há uma clara diferença dialetal: enquanto as crianças paulistas produziram muito pouco (ou quase nada) de QU-*in situ*, esta estratégia foi uma das primeiras a emergir no dialeto baiano. Quanto a isso, Grolla (2009) argumenta que não devemos comparar os resultados dos dois dialetos, visto que, na língua adulta, há diferenças importantes nas duas variedades, que inclusive licenciam diferentes estruturas: (i) no dialeto baiano, aceita-se a construção *non-D-linked* “que diabo” em posição *in situ*, já no paulista, isso é impossível; (ii) Lessa-de-Oliveira relata também a ocorrência da estrutura “é-QU-que”, como em “é o que que ele quer?”, não encontrada no dialeto paulista.

### 3.2 Dados experimentais

Vieira & Grolla (2020) realizaram um estudo com 52 crianças entre 4;6 e 5;6 anos, além de 60 adultos falantes de PB, utilizando uma tarefa de produção eliciada para examinar o uso de perguntas-QU, com foco especial na posição do elemento-QU, se movido ou *in situ*. O estudo tinha duas condições: uma envolvendo *Common Ground* (CG)

---

<sup>9</sup> Pesquisas em psicolinguística argumentam que uma alta frequência de uma dada construção no *input* pode orientar as produções infantis nos estágios iniciais de aquisição (Diessel; Tomasello, 2000; Bybee; Hopper, 2001; Huttenlocher et al., 2002; Bod; Hay; Jannedy, 2003). Observamos essa correlação em diversos fenômenos e estudos, como evidenciado no caso de uma das crianças analisadas por Lessa-de-Oliveira (2003). Entretanto, Yang (2015), com base em um robusto *corpus* de inglês infantil contendo 3 milhões de palavras, notou que palavras irregulares frequentes são tão passíveis de erros de sobre-regularização quanto aquelas menos frequentes: por exemplo, *blew* foi a palavra mais sobre-regularizada, apesar de ser bastante comum no *input* da criança. Portanto, embora o *input* tenha um papel fundamental no processo de aquisição, e muitas vezes a ordem de emergência e frequência das estruturas no *input* pareçam estar interligadas, essa relação nem sempre é garantida, o que indica que a frequência no *input* não pode ser a única explicação para os dados infantis aqui discutidos.

proeminente e outra com CG menos proeminente. Na condição proeminente, os participantes precisavam montar uma carta igual à de um fantoche, fazendo perguntas baseadas em suas instruções para estabelecer um CG discursivo. E na condição com CG menos proeminente, eles criavam um cenário e o fantoche tentava adivinhá-lo por meio de suas perguntas. A previsão geral era que a produção de QU-*in situ* seria favorecida em contextos com CG mais proeminente; e desfavorecida em contextos em que o CG era menos evidente.

Os achados do estudo corroboraram as previsões, sugerindo que a criação de um CG influenciou a formulação das estruturas interrogativas, tanto na fala de adultos quanto na de crianças. Ambos os grupos empregaram mais QU-*in situ* na condição com CG proeminente ( $\chi^2$ , p-valor < 0,001). Os adultos usaram essa construção em 49,5% das perguntas-QU na condição com CG proeminente, enquanto na condição menos proeminente, essa taxa caiu para 33,5%. As crianças, por sua vez, empregaram a estrutura *in situ* em 26% das perguntas-QU na condição proeminente e em 15,7% na condição não proeminente. Isso indica que, em um contexto apropriado para a emergência da estrutura, as crianças a utilizaram de maneira produtiva. No entanto, os adultos a empregaram em aproximadamente o dobro das ocasiões que as crianças. Assim, os resultados sugerem que as crianças analisadas, entre 4;6 e 5;6 anos (mais velhas que as participantes de estudos longitudinais anteriores), já possuem a estrutura em sua gramática, mas ainda a utilizam em menor frequência em comparação aos adultos, preferindo estratégias que envolvam o movimento do elemento-QU.

## 4 Dados da pesquisa atual

### 4.1 *Corpus*

Os dados analisados neste estudo foram extraídos do *corpus* Santana-Santos, organizado pela Profa. Dra. Raquel Santana Santos (USP). Ele consiste em amostras longitudinais de quatro crianças - Leonardo, Luana, Maria Eliza e Túlio - nascidas e criadas em São Paulo, com idades entre 1;02.28 e 5;06.07, além do *input*. As interações foram registradas em 308 sessões, cada uma com duração média de 30 minutos, totalizando cerca de 154 horas de gravação. O *input* consistiu nas falas de 23 adultos, incluindo familiares e conhecidos, como pais, tios, avós e cuidadores, com quem as crianças já mantinham laços afetivos. A tabela abaixo sumariza a distribuição dos dados:

Nome	Idade inicial	Idade final	Nº de sessões	Tempo total de gravação	# Perguntas infantis	# Perguntas do input
Luana	1;07.12	5;06.07	167 <sup>10</sup>	83h30m	2832	9186
Leonardo	2;00.14	4;08.27	36	18h	365	1491

<sup>10</sup> Para endereçar a diferença considerável no tamanho das amostras, verificamos o equilíbrio interno de cada uma delas. Calculamos a proporção de perguntas produzidas pela criança em relação ao total de perguntas da diáde (adulto-criança). Observamos uma consistência entre três das quatro crianças: Luana (23,6%), Túlio (26,5%) e Leonardo (19,7%). A taxa menor observada em Maria E. (5,2%) reflete sua idade inicial mais precoce (1;02), momento em que há menor produtividade linguística da criança. Além disso, as análises individuais apontam a uma mesma tendência: a preferência pelo QU-*ex situ* por todas as crianças. Portanto, consideramos que a disparidade no tamanho das amostras não introduz um viés na análise, já que a amostra mais robusta de Luana corrobora as mesmas tendências de aquisição verificadas nas amostras menores.

Túlio	2;00.07	4;11.12	70	35h	1078	2992
Maria E.	1;02.28	4;02.07	35	17h30m	141	2577
Total	-	-	308	154h	4416	16246

TABELA 1 – Descrição do corpus espontâneo longitudinal.

Fonte: elaborada pelos autores.

As perguntas coletadas foram extraídas de transcrições de gravações de áudio, produzidas durante situações cotidianas em que as crianças interagem com adultos e crianças mais velhas. Essas interações geralmente ocorriam durante atividades como leitura de livros, brincadeiras e refeições, sem que sua fala tenha sido direcionada para assuntos específicos ou para a produção de estruturas ou palavras específicas. Neste estudo, portanto, foi considerada tanto a fala espontânea das crianças como a fala dirigida a elas.

#### 4.2 Metodologia

Para as análises a seguir, do total de perguntas extraídas do *corpus* na Tabela 1, incluímos apenas as perguntas simples, diretas e gramaticais produzidas pelas crianças e seus interlocutores. Excluímos todas as perguntas com entoação eco, com elisão do verbo, perguntas “formulaicas”, como expressões cristalizadas da língua, (e.g. “que foi?”), retóricas, múltiplas, imitadas de uma sentença anterior, perguntas com “cadê”<sup>11</sup> e perguntas de sujeito sem o acréscimo do complementizador e/ou cópula (‘que’ ou ‘é-que’)<sup>12</sup>. Por fim, também excluímos as perguntas construídas a partir de fragmentos do discurso imediatamente anterior (e.g. A: falei.../ B: falou o quê?). Ao excluí-las, visamos restringir nossa análise a contextos de genuína opcionalidade sintática, nos quais tanto o QU-*in situ* quanto o QU-*ex situ* são candidatos viáveis.

Além disso, classificamos cada ocorrência interrogativa seguindo critérios sintáticos e pragmáticos:

- a) Função do constituinte interrogativo: classificado conforme a função sintática exercida pelo constituinte-QU (sujeito, objeto direto, objeto indireto, objeto de preposição, adjunto, predicativo);
- b) Palavra-QU: classificado conforme o pronome empregado na pergunta-QU (como, quando, onde, por que, o que, quem, qual);

<sup>11</sup> As perguntas com “cadê” também foram retiradas da análise por não conterem um verbo finito e por não possuírem uma contraparte *in situ* (“cadê o livro?” vs. “o livro cadê?”).

<sup>12</sup> Em perguntas de sujeito sem complementizador ou cópula, o movimento para a periferia esquerda não pode ser observado independentemente. Por exemplo, na pergunta (i) abaixo, ‘quem’ pode ter se movido para Spec,CP ou pode ter permanecido dentro de TP. Não há nada na estrutura que nos permita analisá-la de forma não ambígua. Já em (ii), a presença de ‘que’ seguindo ‘quem’ deixa claro que houve movimento para Spec,CP:

- (i) Quem fez isso?
- (ii) Quem [<sub>C</sub> que] fez isso?

Casos como (ii) foram contabilizados como envolvendo movimento ao passo que casos como (i) não foram incluídos na análise. Além disso, foram também suprimidas as perguntas sem contexto determinado, como aquelas realizadas em início de gravação.

- c) Complexidade do constituinte interrogativo: (i) baixa, sem NP e sem preposição, e.g. “ele fez o quê?”, ou seja, estruturas em que temos um constituinte não ramificado; (ii) intermediária, com estruturas em que há uma ramificação (como um NP seguindo a palavra-QU ou uma palavra-QU com preposição), e.g. “ele fez que lanche?” ou “ele pôs em qual?”; e (iii) alta, em estruturas com NP e com preposição, e.g. “ele colocou em qual estante?”, ou seja, estruturas em que existe uma ramificação adicional;
- d) Tipo de verbo: lexical (“o que o menino comeu?”) ou cópula (“o que é aquele brinquedo?”)
- e) Estratégia interrogativa utilizada, como QU-movido simples, QU-que, QU-é-que, QU-*in situ*;
- f) Contexto: (i) antecedente discursivo (perguntas cuja porção não-QU está presente, explícita ou implicitamente, no discurso imediatamente anterior); (ii) extralinguístico (perguntas vinculadas diretamente ao ambiente que cerca os falantes e/ou a ações não linguísticas dos sujeitos envolvidos); (iii) conhecimento prévio (perguntas vinculadas a pressuposições universais, culturais ou a algum conhecimento previamente compartilhado pelos envolvidos na conversa, como uma referência a alguma situação que eles experienciaram juntos); (iv) *out-of-the-blue* (perguntas feitas de supetão); (v) contexto indisponível (perguntas realizadas no início da gravação ou para as quais, por qualquer motivo, não pudemos recuperar o contexto).

Para a análise quantitativa, empregamos Modelos Lineares Generalizados Mistos (GLMM) com distribuição binomial e função *logit*, dado que nossa variável resposta é binária (presença ou ausência de movimento-QU aberto) e para controlar a variabilidade individual dos participantes por meio de efeitos aleatórios. Especificamente para o grupo infantil, devido à distribuição dos dados que impediu a convergência de um modelo misto, optou-se por um Modelo Linear Generalizado (GLM) simples. Modelos separados para cada grupo testaram os preditores morfosintáticos (Palavra-QU, Função Sintática, Complexidade e, para os adultos, tipo de verbo). Um segundo modelo investigou a influência pragmática na produção das perguntas-QU, tendo como variável independente o tipo de contexto (classificado conforme acima), e testou sua interação com o grupo (adulto vs. criança).

## 5 Resultados

De modo geral, as crianças e os adultos produziram um grande número de perguntas-QU. Antes de nos debruçarmos sobre as taxas de produção, fornecemos, nos exemplos que se seguem, as primeiras produções infantis de cada tipo em nosso *corpus*<sup>13</sup>:

---

<sup>13</sup> Os dados apresentam apenas as transcrições ortográficas das falas infantis, deixando de lado as minúcias fonológicas, como o apagamento da marca de infinitivo em ‘tirar’ ou de ‘vo’ em ‘você’, que, para os fins desta pesquisa, não são relevantes.

- (14) Primeiras perguntas com QU-movido simples:
- a. que é isso aqui? (Leonardo, 2;00.14)
  - b. que você quer, mãe? (Luana, 2;01.04)
  - c. que é esse? (Maria, 1;05.17)
  - d. o que ele está fazendo? (Túlio, 2;00.07)
- (15) Primeiras perguntas com QU-que:
- a. que que você está comendo? (Leonardo, 2;00.14)
  - b. que que é esse? (Luana, 1;07.12)
  - c. como que faz? (Maria, 1;08.24)
  - d. que que tem aqui? (Túlio, 2;00.07)
- (16) Primeiras perguntas com QU-é-que:
- a. como é que ela aparece aí? (Leonardo, 3;08.16)
  - b. que é que está fazendo esse? (Luana, 2;10.00)
  - c. como é que eu vou tirar? (Túlio, 3;02.11)
- (17) Primeiras perguntas com QU-*in situ*:
- a. está aonde? (Leonardo, 2;03.01)
  - b. é que cor? (Luana, 2;05.07)
  - c. foram tintas que cor? (Maria, 2;11.30)
  - d. tira o quê? (Túlio, 2;10.11)

Após aplicados os critérios de exclusão, chegamos aos seguintes resultados: foram observadas 17.457 perguntas-QU, dentre as quais 3.255 foram produzidas pelas crianças e 14.202 pelos adultos, como detalhado na tabela abaixo:

Grupo	corpus Luana	corpus Túlio	corpus Leonardo	corpus Maria E.	Total
Criança	1.217	386	120	50	1.773
Input	4.710	1.830	746	1.393	8.679
Total	5.927	2.216	866	1.443	10.452

TABELA 2 – Perguntas-QU produzidas por crianças (N=4) e adultos (N=23) após critérios de exclusão.

Fonte: elaborada pelos autores.

As crianças produziram 26 perguntas com QU-*in situ*, totalizando 1,5% de seus dados, enquanto os adultos produziram 850, totalizando 9,8%. Em contrapartida, foram produzidas 1.747 perguntas com o QU-*ex situ* na fala infantil, totalizando 98,5%, enquanto os adultos produziram 7.829 dessas estruturas, ou 90,2% das perguntas-QU. O gráfico abaixo ilustra a clara preferência tanto de crianças quanto de adultos pelo movimento do elemento-QU:

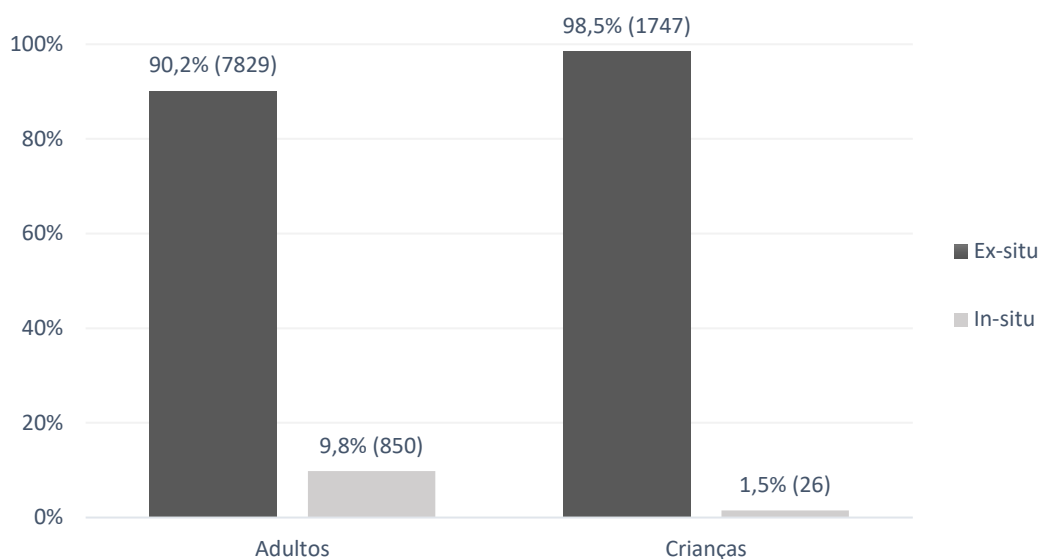


GRÁFICO 1 – Comparação entre QU-ex situ e QU-in situ na fala adulta (N=23) e infantil (N=4).  
 Fonte: elaborada pelos autores.

Analizamos, a seguir, todas as estratégias-QU separadamente: separamos o QU-ex situ em QU-movido simples (“Quem o gato viu?”), QU-que (“Quem que o gato viu?” e QU-é-que (“Quem é que o gato viu?”), contrapondo-os ao QU-in situ (“O gato viu quem?”). O gráfico abaixo revela que a estratégia mais utilizada na fala das crianças foi o QU-movido simples, com 75,5% de todas as perguntas-QU, seguida pelo QU-que, que representou 21,4%. As estratégias menos utilizadas foram o QU-é-que, com 1,6% e o QU-in situ, com 1,5%. Já os adultos produziram mais QU-que, totalizando 54,1% das perguntas-QU, seguido pelo QU-movido simples, com 24,6%. As estratégias menos produtivas na fala adulta foram o QU-é-que, com 11,5%, e o QU-in situ, com 9,8%.

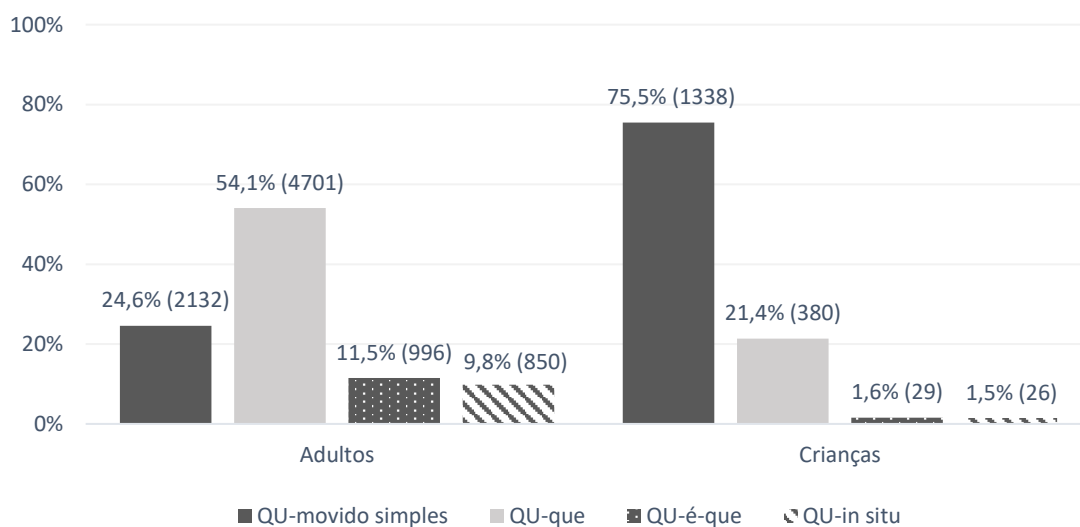


GRÁFICO 2 – Comparação entre estratégias-QU na fala adulta (N=23) e infantil (N=4).  
 Fonte: elaborada pelos autores.

A primeira pergunta com QU-*in situ* apareceu no *corpus* aos 2;03.01, na fala de Leonardo. O QU-movido simples, por outro lado, surgiu já aos 1;05.17 na fala de Maria. O QU-que também foi registrado relativamente cedo, a partir dos 1;07.12 na fala de Luana (inclusive, nos dados desta criança, apareceu antes do QU-movido simples). O QU-é-que, por fim, foi a última estratégia a ser observada, somente a partir dos 2;10 anos, na fala de Luana. Em relação à ordem de emergência das estruturas na fala de Leonardo e de Túlio, o QU-movido simples e o QU-que apareceram em suas primeiras sessões (aos 2;00.14 e 2;00.07 respectivamente):

Estratégias	Leonardo	Luana	Maria	Túlio
QU-movido simples	2;00.14	2;01.04	1;05.17	2;00.07
QU-que	2;00.14	1;07.12	1;08.24	2;00.07
QU-é-que	3;08.16	2;10.00	-	3;02.11
QU- <i>in situ</i>	2;03.01	2;05.07	2;11.30*	2;10.11*

\*perguntas sem contraparte movida possível – exemplos em (17)

TABELA 3 – Idade de emergência das estratégias-QU na fala das crianças.

Fonte: elaborada pelos autores.

Conforme mencionado na seção 1, o QU-*in situ* pode aparecer em estruturas em que uma contraparte movida é possível, como em “o João comeu o quê?”, e em estruturas em que é impossível, como em “o dodói de quem?”. A partir dessa distinção, classificamos as do primeiro tipo como QU-*in situ* opcional e as do segundo como QU-*in situ* obrigatório. A versão opcional foi usada por apenas duas crianças, Leonardo e Luana. Já o obrigatório surgiu na fala de Maria e de Túlio, aos 2;11.30 e 2;10.11 anos. Essas idades indicam que o QU-*in situ* é, de fato, uma estrutura tardia, surgindo em nossos dados ~10 meses após o QU-movido simples.

Em síntese, notamos que as crianças tendem a seguir o padrão geral dos adultos, produzindo predominantemente a construção mais frequente no *input*, o QU-*ex situ*. No entanto, não seguiram essa tendência de forma completa, já que, apesar de terem demonstrado também uma “preferência” pelo movimento-QU, desfavoreceram o QU-*in situ* mais que os adultos. Em relação às estratégias-QU como um todo, os resultados sugerem que as crianças começam produzindo o QU-movido simples, aos 1;05, sendo essa a estratégia mais frequente em seus dados, com 75,5% de todas as perguntas-QU. O QU-que surge logo em seguida, aos 1;07, representando 21,4% dos dados. O QU-*in situ* aparece no *corpus* aos 2;03, sendo produzido em sua forma opcional por duas das quatro crianças e raramente, com apenas 1,5% das perguntas-QU. Por fim, o QU-é-que é a última estratégia a emergir nos dados, somente a partir dos 2;10 anos e totalizando 1,6% das perguntas. Portanto, a ordem de emergência observada foi: QU-movido simples ~ QU-que > QU-*in situ* > QU-é-que, i.e., o QU-movido simples e QU-que, quase simultaneamente, seguidos após um intervalo de 8 meses pelo QU-*in situ* e, finalmente, pela estrutura mais tardia, o QU-é-que.

De maneira individual, não houve grandes diferenças nas duas primeiras preferências infantis: todas as crianças empregaram mais o QU-movido simples e, em segundo lugar, o QU-que. Túlio e Maria não produziram QU-*in situ* em sua versão opcional, sugerindo uma aquisição mais tardia em comparação com as outras duas crianças. Apenas Luana e Leonardo utilizaram a versão opcional, sendo a construção mais frequente na fala de Leonardo (7,5%) que, curiosamente, foi também aquele que mais a ouviu em seu contexto familiar (12,3%).

Já os adultos produziram, em números agregados, mais QU-que, totalizando 54,1% das perguntas-QU, seguido pelo QU-movido simples, com 24,6%. Essa preferência pelo QU-que no *input* confirmou-se em três das

quatro amostras analisadas (Luana, Túlio e Maria E.), havendo uma inversão apenas no *corpus* de Leonardo, no qual os adultos produziram ligeiramente mais QU-movido simples (40,8%) do que QU-que (34,5%). As estratégias-QU menos frequentes na fala adulta foram o QU-é-que, com 11,5%, e o QU-*in situ*, com 9,8%. A maior diferença entre crianças e adultos (com exceção do *corpus* adulto ao qual Leonardo foi exposto) foi a maior produtividade do QU-movido simples na fala infantil, apesar da preferência pelo QU-que no *input*. Além disso, a baixa frequência de QU-é-que e QU-*in situ*, mesmo considerando sua emergência tardia, não se equipara ao *input* observado. Em outras palavras, se as crianças estivessem seguindo rigorosamente os dados que recebem de seu contexto, esperaríamos uma produção semelhante de QU-é-que e QU-*in situ*, que foram empregadas em 11,5% e 9,8% na fala adulta, respectivamente, mas que, em contrapartida, foram consideravelmente menos frequentes na fala infantil, totalizando 1,6% e 1,5% das perguntas-QU. Ademais, seria de se esperar que a primeira estratégia a emergir nos dados infantis e a mais frequente fosse o QU-que, uma vez que esta é a mais produtiva nos dados adultos.

### 5.1 Influência dos Fatores Morfossintáticos

A seguir, analisamos a influência dos fatores morfossintáticos na escolha do tipo de interrogativa, QU-*in situ* ou *situ*. Para tanto, observamos as perguntas quanto ao emprego de palavras-QU, função sintática, o nível de complexidade do constituinte-QU e o tipo de verbo.

Para avaliar o quanto cada uma dessas variáveis pode prever a produção de QU-*in situ*, submetemos os dados a uma Regressão Logística de Efeitos Mistos (GLMM). A variável dependente corresponde à presença ou ausência de movimento nas perguntas-QU, enquanto as variáveis independentes são os fatores morfossintáticos mencionados anteriormente e o falante foi incluído como efeito aleatório. Os resultados devem ser interpretados com referência ao QU-*in situ*, ou seja, devem ser analisados em termos do que aumenta ou diminui a probabilidade dessa construção. A Sintaxe no R para o modelo adulto foi: `glmer(Classificacao ~ Palavra.QU + Funcao.QU + Verbo + Complexidade + (1|Falante), family = binomial, data)`. O output obtido para o adulto está apresentado na tabela a seguir<sup>14</sup>:

Variável	Nível (vs. Referência)	Odds Ratio (OR) <sup>15</sup>	IC 95%	Valor-p
<b>Intercepto</b> <sup>16</sup>	Ref: Como / Adjunto / Peso 1 / Lexical	0.021	[0.012 – 0.035]	< 0.001
<b>Palavra-QU</b>	O que	17.06	[9.65 – 30.17]	< 0.001
	Onde	9.37	[5.89 – 14.88]	< 0.001
	Por que	2.53	[1.42 – 4.52]	0.002
	Quem	6.15	[3.26 – 11.62]	< 0.001

<sup>14</sup> Seguindo indicação dos pareceristas, testamos também se haveria ortogonalidade entre as variáveis (como entre certas palavras-QU e suas funções sintáticas) utilizando o Fator de Inflação da Variância (VIF). Os resultados confirmaram que todas as variáveis são suficientemente independentes para entrar no mesmo modelo, com valores de VIF padronizados sempre abaixo de 2 (Palavra-QU (1,16), Função Sintática (1,64), Complexidade (1,33) e Tipo de Verbo (1,78)).

<sup>15</sup> As razões de chances (*odds ratios*) apresentadas refletem a probabilidade relativa de ocorrência de QU-*in situ*. Valores acima de 1 indicam que a estrutura é favorecida em comparação ao nível de referência (ex: Peso 2 *versus* Peso 1), enquanto valores abaixo de 1 indicam desfavorecimento.

<sup>16</sup> O intercepto atua como uma linha de referência na análise de regressão e os outros coeficientes indicam desvios em relação a ela.

	Quanto	2.84	[1.12 – 7.19]	0.028
	Qual	1.64	[0.81 – 3.32]	0.17
	Quando	< 0.001	[0.00 – 15.01]	0.99
	Que	1.14	[0.62 – 2.10]	0.66
<b>Função Sintática</b>	Obj. Direto	0.56	[0.36 – 0.88]	0.012
	Obj. Indireto	0.49	[0.33 – 0.74]	< 0.001
	Obj. Preposicionado	3.00	[1.80 – 5.01]	< 0.001
	Predicativo	1.38	[0.85 – 2.24]	0.19
<b>Complexidade</b>	Peso 2 (vs. Peso 1)	2.16	[1.61 – 2.91]	< 0.001
	Peso 3 (vs. Peso 1)	9.11	[3.33 – 24.95]	< 0.001
<b>Verbo</b>	Cópula (vs. Lexical)	0.50	[0.37 – 0.68]	< 0.001

TABELA 4 – Resultados do modelo de regressão logística de efeitos mistos (GLMM) para a produção de QU-*in situ* nos dados adultos. A variável resposta é a presença (1) ou ausência (0) de *in situ*. O modelo inclui interceptos aleatórios por falante. Acurácia do Modelo: R<sup>2</sup> Marginal (efeitos fixos) = 0,34; R<sup>2</sup> Condicional (fixos + aleatórios) = 0,38..

Fonte: elaborada pelos autores.

Na análise das palavras interrogativas (Palavra.QU), o coeficiente para o pronome ‘como’<sup>17</sup> serve como referência (intercepto) para as comparações. O pronome ‘o que’ aumenta significativamente a chance da ocorrência de uma pergunta *in situ*, com um aumento de 17.06 vezes nas chances em comparação ao intercepto. Ou seja, mantendo as outras variáveis do modelo constantes, a chance de ocorrer uma pergunta *in situ* é 17.06 vezes maior quando há o pronome ‘o que’ do que quando ‘como’ é utilizado. Já ‘onde’ proporciona um aumento de cerca de 9.37; ‘quem’, um aumento de aproximadamente 6.15; e ‘por que’ um aumento de 2.53. O pronome ‘quanto’ também se mostrou significativo, com um aumento de 2.84 vezes. As palavras ‘qual’, ‘quando’ e ‘que’ não demonstraram diferenças estatisticamente significativas em suas chances de ocorrência na posição *in situ* quando comparados a ‘como’. Quanto à função sintática dos constituintes-QU, com o adjunto servindo como referência, observou-se uma redução na chance de uso do QU-*in situ* com ‘objeto direto’ e ‘objeto indireto’ em comparação ao intercepto. Especificamente, o ‘objeto direto’ reduz essa chance em 0.56 vezes e o ‘objeto indireto’ em 0.49 vezes. Já o ‘objeto de preposição’ aumenta essa chance em cerca de 3 vezes. Todas as funções sintáticas, exceto ‘predicativo’, têm probabilidades de ocorrência significativamente distintas em relação a adjuntos. Em relação ao nível de complexidade do constituinte, observou-se que constituintes classificados como de complexidade intermediária (nível 2) aumentam em 2.16 vezes as chances de ocorrência de uma pergunta *in situ* em comparação aos de complexidade baixa (nível 1), enquanto aqueles de complexidade alta (nível 3) proporcionam um aumento substancial, com 9.11 vezes mais chances de ocorrência *in situ*. A razão de chance para ambos foi significativa. Por fim, no que tange ao tipo de verbo, constatou-se que a presença da cópula está associada a uma diminuição na probabilidade de ocorrência de QU-*in situ*, com uma razão de chance significativa de 0.5 vezes menor do que quando um verbo lexical está presente.

Para o grupo infantil, tentamos a princípio aplicar o mesmo modelo misto utilizado para os adultos, atendendo à sugestão dos pareceristas de incluir o falante como variável aleatória. No entanto, isso se revelou inviável

<sup>17</sup> Mantivemos ‘como’ e ‘adjunto’ como referência para facilitar a leitura. Como eles apresentam menos casos de QU-*in situ*, usá-los como base faz com que os resultados mostrem o quanto os outros fatores aumentam a chance dessa ocorrência (resultando em números positivos e maiores que 1), o que é mais intuitivo de interpretar do que reduções.

devido à distribuição dos dados infantis: diferente dos adultos, as crianças raramente produziram o QU-*in situ*, o que gerou lacunas nos dados (e.g. nenhuma ocorrência de ‘quando’ *in situ* para determinada criança) e inviabilizou os mesmos cálculos estatísticos. Diante da limitação da amostra, optamos por manter um modelo mais simples (GLM) incluindo apenas as variáveis que apresentaram dados suficientes para serem analisados, excluindo a variável ‘tipo de verbo’. A sintaxe no R para o modelo foi: `glm(Classificacao ~ Palavra.QU + FunPalavra.QU + Complexidade, family = binomial, data)`, resultando no *output* abaixo:

Variável	Nível (vs. Referência)	Odds Ratio (OR)	IC 95%	Valor-p
<b>(Intercepto)</b>	Ref: Como / Adjunto / Peso 1	< 0,001	[Incalculável]	0,986
<b>Palavra-QU</b>	O que	> 1.000	[0,00 – ∞]	0,988
	Onde	> 1.000	[0,00 – ∞]	0,989
	Por que	> 1.000	[0,00 – ∞]	0,990
	Qual	> 1.000	[0,00 – ∞]	0,989
	Quando	1,02	[0,00 – ∞]	0,999
	Quanto	0,37	[0,00 – ∞]	0,999
	Que	> 1.000	[0,00 – ∞]	0,989
	Quem	> 1.000	[0,00 – ∞]	0,990
<b>Função Sintática</b>	Obj. Direto	0,17	[0,016 – 2,47]	0,165
	Obj. Indireto	1,15	[0,122 – 9,15]	0,896
	Obj. Preposicionado	5,64	[0,376 – 96,0]	0,213
	Predicativo	0,28	[0,031 – 3,62]	0,287
<b>Complexidade</b>	Peso 2 (vs. Peso 1)	6,24	[1,40 – 28,1]	0,015

TABELA 5 – Resultados do modelo simplificado (GLM) para os dados infantis. Devido à baixa frequência de QU-*in situ*, as estimativas para Palavra-QU são instáveis (valores de p não significativos e intervalos indeterminados). Apenas a variável Complexidade mostrou-se um preditor confiável e significativo. Acurácia do Modelo:  $R^2 = 0.89$ .

Fonte: elaborada pelos autores.

As variáveis Palavra.QU e função sintática do constituinte-QU não se correlacionam significativamente com a ocorrência de QU-*in situ* ou *ex-situ* devido à baixa variabilidade dos dados. No que se refere ao nível de complexidade do constituinte, observou-se que o nível de complexidade (2) apresentou um aumento de 6.24 vezes na probabilidade de emprego da estratégia QU-*in situ* em comparação ao nível de complexidade baixo (1). Contudo, é crucial notar que a probabilidade base prevista pelo modelo para o uso de *in situ* é extremamente baixa (próxima de 0%), indicando uma preferência robusta e quase categórica pela estratégia QU-*ex situ*. No que concerne à função sintática, a análise inferencial não apontou significância, reforçando a tendência geral de movimento. Portanto, os resultados do modelo indicam que somente o nível de complexidade do constituinte-QU é um fator estatisticamente significativo, mas mesmo com o aumento proporcionado pela maior complexidade, a probabilidade absoluta de ocorrência de *in situ* permanece inferior a 1%.

Em termos de estatística descritiva, registrou-se um aumento de QU-*in situ* nas perguntas com ‘objeto indireto’ (de todas as perguntas com essa função, o QU-*in situ* foi empregado em 10,6% na fala infantil e em 22,4% na fala adulta) e um aumento ainda mais expressivo nas que envolvem ‘objeto de preposição’ (nas perguntas com essa função, houve 49,5% de QU-*in situ* nos dados adultos e 34,8% nos dados infantis). Uma possível explicação para essa preferência numérica é que ambas as funções sintáticas incluem preposição, o que pode tornar menos custoso manter o elemento-QU *in situ* ao invés de movê-lo, realizando *pied-piping*. Além disso, ao observar o nível de complexidade do sintagma interrogativo, notamos uma tendência crescente em favor do QU-*in situ* à medida

que o constituinte se torna mais complexo. Em outras palavras, quanto maior a complexidade do constituinte, maiores são as chances de que ele permaneça *in situ*. Por exemplo, a probabilidade de que a sentença “ele colocou [em qual estante]?” (PP e QU+NP) possua o sintagma QU *in situ* é maior do que a da sentença “ele colocou [o quê]?” (sem PP e sem QU+NP).

## 5.2 Influência dos Fatores Pragmáticos

A seguir, analisamos a influência de fatores pragmáticos na escolha da interrogativa, QU-*in situ* ou *ex situ*. Para tanto, separamos as perguntas conforme o contexto em que era produzida: discursivo (porção não-QU implícita ou explícita), extralinguístico, conhecimento prévio (interações prévias, pressuposições culturais e universais) e *out-of-the-blue*.

Para analisar o impacto de cada variável na produção de QU-*in situ* e verificar se há interação com o grupo etário, aplicamos uma Regressão Logística de Efeitos Mistos (GLMM), o que permite controlar a variabilidade individual dos participantes (efeito aleatório). A variável dependente foi a presença ou ausência de movimento nas perguntas-QU, enquanto as variáveis independentes incluíram o contexto e o grupo etário (Adulto vs. Criança). A Sintaxe no R foi: `glmer(Classificacao ~ Contexto * Grupo + (1|Participante), family = binomial, data)`. Os resultados, apresentados na tabela 6, devem ser interpretados em termos de como aumentam ou diminuem a probabilidade de ocorrência do QU-*in situ*. Com esse modelo, obtivemos o *output* na tabela abaixo:

Variável	Nível (vs. Referência)	Odds Ratio (OR)	IC 95%	Valor-p
<b>Intercepto</b>	Adulto / Discursivo	0.16	[0.11 – 0.24]	< 0.001
<b>Contexto</b>	Conhecimento Prévio	0.39	[0.29 – 0.51]	< 0.001
	Extralinguístico	0.34	[0.30 – 0.40]	< 0.001
	<i>Out-of-the-blue</i>	< 0.001	[0.00 – ∞]	0.97
<b>Grupo</b>	Criança (vs. Adulto)	0.29	[0.12 – 0.68]	0.004
<b>Interação</b>	Criança * Conhecimento Prévio	2.09	[0.26 – 16.6]	0.48
	Criança * Extralinguístico	0.82	[0.42 – 1.60]	0.55
	Criança * <i>Out-of-the-blue</i>	4.78	[0.00 – ∞]	0.99

TABELA 6 – Resultados do Modelo Linear Generalizado Misto (GLMM) avaliando a influência do contexto pragmático e do grupo etário. A variável resposta é a presença (1) ou ausência (0) de *in situ*. O modelo inclui interceptos aleatórios por participante (Adulto/Criança).

Acurácia do Modelo: R<sup>2</sup> Marginal (efeitos fixos) = 0,33; R<sup>2</sup> Condicional (fixos + aleatórios) = 0,40.

Fonte: elaborada pelos autores.

Na fala adulta, a chance base de produção de QU-*in situ* é de 0.16 vezes menor em relação ao *ex situ*, tomando o contexto discursivo como referência. Em comparação a esse nível base (Intercepto), a chance de ocorrência diminui significativamente quando o Conhecimento Prévio está presente (OR = 0.39, p<0.001) ou sob influência de fatores extralinguísticos (OR = 0.34, p<0.001). Ou seja, a presença desses contextos reduz significativamente a probabilidade de escolha do QU-*in situ* em relação ao contexto discursivo.

Na fala infantil, o efeito principal de Grupo indica que as crianças produzem significativamente menos QU-*in situ* do que os adultos no contexto de referência (OR = 0.29, p = 0.004). Estatisticamente, não houve interação significativa entre grupo etário e contexto, o que poderia sugerir tendências similares. Contudo, dada a frequência extremamente baixa de QU-*in situ* nas crianças, é plausível interpretar que o contexto pragmático ainda não atua

como um gatilho robusto para elas, diferentemente do que ocorre com os adultos. Esse dado torna-se especialmente interessante quando contrastado com os resultados da seção anterior: enquanto a complexidade sintática mostrou-se um preditor significativo para as crianças (Tabela 5), os fatores pragmáticos não parecem exercer a mesma força. Isso parece apontar para a hipótese de que a criança adquire primeiro as restrições gramaticais, sendo a sensibilidade às restrições pragmáticas um desenvolvimento mais tardio.

Os resultados sugerem que, ainda que o QU-*in situ* pareça estar restrito a situações em que há um CG estabelecido previamente, nunca ocorrendo em contextos *out-of-the-blue*, o QU-*ex situ* continua sendo a tendência geral na fala dos dois grupos, independente do contexto. Em outras palavras, mesmo quando as condições discursivas favorecem o uso de QU-*in situ* na fala dos adultos, isso não implica que eles vão preferir essa construção em detrimento da sua contraparte movida. Assim, podemos concluir que o QU-*ex situ* é a construção interrogativa padrão da língua, enquanto o uso de QU-*in situ* constitui uma alternativa mais marcada da língua e consideravelmente mais restrita.

### 5.3 Discussão dos Resultados

Para discutirmos os dados reportados acima, recordemos as premissas que motivaram o presente estudo: (i) com base nos dados infantis e adultos reportados na literatura, parece haver uma assimetria entre os dados espontâneos adultos (Lopes-Rossi, 1996; Oushiro, 2012) e infantis (Sikansi, 1999; Grolla, 2000, 2009); (ii) o QU-*in situ*, mesmo sendo produtivo na fala adulta (Lopes-Rossi, 1996; Oushiro, 2012), é uma construção rara e tardia na fala infantil (Sikansi, 1999; Grolla, 2000, 2009), apesar de ser mais econômica do que sua contraparte movida, na análise aqui adotada.

A respeito da assimetria entre fala adulta e infantil observada anteriormente no dialeto paulista, ela pode ter ocorrido devido à natureza dos *corpora* comparados. Argumentamos aqui que não é possível comparar as taxas encontradas na fala infantil às encontradas em estudos como o de Lopes-Rossi (1996) e Oushiro (2012), pois estes envolvem dados que são de natureza diferente daqueles de um *input* em potencial. Os trabalhos de Grolla (2009) e Sikansi (1999a) já sugeriam uma diferença no emprego do QU-*in situ* na fala dirigida à criança (CDS) e naquela entre adultos: nos dois trabalhos, o QU-*in situ* foi a forma interrogativa menos utilizada tanto pelas crianças como pelos adultos com quem elas interagiam.

Oushiro e Lopes-Rossi trabalharam com entrevistas sociolinguísticas ou para a TV, em que os adultos conversam entre si e, nesses contextos, novos questionamentos são recorrentes. Sabemos, no entanto, que, nesse tipo de contexto, o QU-*in situ* é mais feliz mesmo em línguas como o inglês, com movimento-QU predominante (Quirk et al., 1985; Huddleston; Pullum, 2002). Isso porque, consoante Biezma (2018), essas perguntas atuam como movimentos de *follow-up*, promovendo a continuidade do questionamento. É possível pensar, então, que os *corpora* analisados por Oushiro e Lopes-Rossi tenham oferecido mais oportunidades para emprego do QU-*in situ* do que interações cotidianas como as gravações espontâneas com crianças usadas em nosso estudo. Além disso, é possível também que haja uma diferença entre a fala dirigida à criança e a fala entre adultos com respeito ao grau de pressuposições compartilhadas. Por exemplo, caso o QU-*in situ* seja de fato facilitado em contextos de 'forte pressuposição', é possível que, na fala com outros adultos, mais pressuposições sejam compartilhadas do que na fala com crianças, já que estas ainda estão construindo seu repertório de pressuposições culturais e universais. Isso

tornaria o QU-*in situ* mais produtivo na fala voltada a adultos do que àquela dirigida a crianças. Outras duas considerações que nunca devem ser subestimadas são a variação individual e o acaso, variabilidade esta que buscamos controlar estatisticamente ao inserir o falante como efeito aleatório nos modelos apresentados. Em outras palavras, a construção pode, por acaso ou por variação entre indivíduos, não ter sido utilizada com tanta frequência no *corpus* analisado aqui, em comparação com os estudos de Oushiro (2012) e Lopes-Rossi (1996).

Interpretamos a assimetria entre fala adulta e infantil como fruto de uma diferença crucial entre a fala entre adultos, de um lado, e a fala dirigida à criança, de outro. O fato é que os dados adultos aqui reportados correspondem à fala dirigida à criança e estudos anteriores olharam para a fala entre adultos, que, provavelmente em razão da situação comunicativa, obtiveram taxas superiores de QU-*in situ*. A comparação feita aqui parece-nos, portanto, mais apropriada, visto que lida exclusivamente com fala dirigida à criança, i.e., com uma amostra do que as crianças ouvem em seu cotidiano. Esses resultados sugerem que o QU-*in situ* não é tão produtivo na fala adulta quanto se poderia supor inicialmente: o QU-*ex situ* é utilizado em mais de 90% das oportunidades em que ambas as estruturas são licenciadas.

Quanto à produção infantil de QU-*in situ*, nossos dados corroboram os de Grolla (2000, 2009) e Silveira (2011), segundo os quais o QU-*in situ* é uma construção tardia e rara na fala infantil. Isso não é surpreendente, considerando a distribuição desproporcional das estratégias envolvendo movimento-QU em relação ao QU-*in situ* no *input* recebido pelas crianças. Ou seja, visto que apenas raramente as crianças têm contato com o QU-*in situ*, não é inesperado que elas também pouco produzam a construção e esta surja tardiamente em seus dados. Note que a mesma explicação também poderia ser estendida para o QU-*é-que*, construção também tardia e rara nos dados infantis. Quanto à fala adulta, não observamos uma produtividade comparável àquela relatada por Lopes-Rossi (1996) e Oushiro (2012) e nossos resultados estão em concordância com os de Grolla (2009), que reportou 9,3% de QU-*in situ* (67 de 719 perguntas-QU) no *input* de L.

No tocante à controvérsia entre o menor custo derivacional do QU-*in situ* e sua baixa proporção, notemos que isso só representaria um problema para uma abordagem que privilegiasse o Princípio da Economia como motor do desenvolvimento linguístico inicial das crianças (Zuckerman, 2001). Nesse contexto, seria necessário explicar por que em línguas como o inglês as crianças nunca produzem o QU-*in situ* em suas primeiras construções, mesmo sendo uma opção mais econômica e derivacionalmente licenciada pela gramática. O fato é que outros fatores parecem estar em jogo no processo de aquisição. Em inglês, por exemplo, podemos argumentar que as crianças marcam o parâmetro [+movimento-QU], produzindo consistentemente o QU-*ex situ* desde cedo (Guasti, 2014). Mais tarde, em apenas casos bastante restritos, como *quizzes* e interrogatórios, a língua permitiria o uso da opção *in situ* (Nguyen; Legendre, 2021). Já no PB, sendo uma língua de movimento-QU opcional, a criança parece se guiar por seu *input*.

Há evidências sólidas que ressaltam a importância do *input* no processo de aquisição de linguagem. Mesmo dentro de uma abordagem racionalista como a que adotamos, não devemos subestimar a influência da experiência na seleção das informações relevantes para o desenvolvimento linguístico. Lidz & Gagliardi (2015) argumentam que, nesse modelo, o falante deve buscar em seu *input* as pistas que o auxiliem a escolher entre representações abstratas dentro de um conjunto de gramáticas possíveis, que, por sua vez, são inatas. Dessa forma, a experiência “fornece as informações relevantes para a identificação dessas representações” (p. 12.2, tradução nossa),

permitindo que o falante selecione as características da gramática de uma língua específica e, posteriormente, extrapole seu *input* para realizar generalizações. Embora a frequência no *input* por si só não explique todo o comportamento infantil, – conforme sugerido pela preferência da criança pelo QU-movido simples (75,5% das perguntas-QU) em contraste com a maior produtividade de QU-que em seu *input* (54,1% das ocorrências) – há casos em que as pistas estatísticas fornecidas pelo *input* são tão marcantes que orientam a criança em seu processo de aquisição.

Perkins, Feldman & Lidz (2017) apresentaram, por exemplo, um modelo computacional de aquisição da estrutura argumental, fundamentados na ideia de um “filtro do *input*”. Ele foi capaz de categorizar 2/3 dos verbos mais frequentes, filtrando construções não básicas como “ruído”. Essa mesma ideia foi retomada por Schneider, Perkins & Feldman (2020) que modelaram a aquisição da concordância em inglês, considerando o *input* como uma mistura entre “sinal” e “ruído”. Ao filtrar o que seria classificado como “ruído”, ou seja, filtrando construções não básicas da língua, o modelo foi capaz de fazer as regularizações necessárias para estabelecer uma gramática consistente da língua alvo. Nguyen & Legendre (2021), ao aplicarem essa mesma hipótese, observaram que a produção das perguntas de sondagem<sup>18</sup> em inglês, na forma com QU-*in situ*, pode ser resultado da atuação desse mecanismo nos primeiros estágios de aquisição: esse tipo de pergunta é utilizado em cerca de 10% das perguntas-QU na fala direcionada a crianças e, quando comparada a uma variante mais frequente, como o QU-*ex situ* (>80%), pode ser filtrada como “erro” pela gramática infantil. As autoras destacam que as crianças tendem a regularizar em direção à variante mais frequente, considerando a menos frequente como “ruído”.

Ao aplicar esse conceito aos nossos dados, notamos que as crianças são expostas a aproximadamente 10% de QU-*in situ* e cerca de 90% de QU-*ex situ*. Essa diferença desproporcional pode servir como uma pista estatística que leva a criança a classificar o QU-*in situ* que ouve como “ruído”, de maneira semelhante ao que acontece em inglês. Ela observa que, em sua língua, 90% das perguntas-QU são produzidas com movimento-QU, sendo o restante analisado como “ruído” e, com isso, produz massivamente a variante mais frequente em seu *input*, o QU-*ex situ*. Essa “escolha” seria realizada a despeito de o QU-*in situ* ser mais econômico do que o *ex situ*, pois essa pista estatística teria maior impacto do que cálculos sobre custo derivacional, fazendo com que ela opte pelo QU-*ex situ* justamente por não ter, em seu *input*, evidências robustas de que em sua língua o movimento-QU é opcional. Essa filtragem seria realizada nos primeiros estágios de aquisição, antes que a criança perceba que as duas variantes não são totalmente opcionais. Posteriormente, ao notar que o QU-*in situ* envolve particularidades pragmáticas e sintáticas, a criança começa a utilizá-lo com maior frequência, especialmente em situações de entrevistas e inquéritos, onde essa estrutura é mais apropriada, alcançando taxas semelhantes às dos adultos.

Neste trabalho, argumentamos que a criança adquirindo uma língua conta com diversos mecanismos que a guiam nesse intrincado processo. Nenhum fator isoladamente pode explicar o desafio que representa a aquisição

<sup>18</sup> As perguntas de sondagem são um tipo específico de pergunta-QU, que ocorre principalmente em contextos de sala de aula ou *quizzes* e interrogatórios. Nesses casos, a palavra-QU complementa um sintagma nominal e é utilizada quando o falante busca acessar o conhecimento de seu interlocutor, deixando uma lacuna com o elemento-QU na pergunta.

Teacher: *I need to know about displacements. They have a what?*  
 eu preciso saber sobre deslocamentos. Eles têm uma o que  
 “Eu preciso saber sobre deslocamentos. Eles têm uma o quê?”

Student: *Distance.*

“Distância”

(Nguyen; Legendre, 2021, p. 7)

das perguntas-QU. Em nosso estudo, as crianças produziram menos QU-*in situ*, alinhando-se ao seu *input*, e, ao mesmo tempo, empregaram mais QU-movido simples em lugar do QU-que, contrariando a preferência dos adultos com quem interagiram. Ou seja, o filtro do *input* daria conta de explicar o desfavorecimento de QU-*in situ* por parte da criança, dada a escassez de evidências de que a língua permite alternativas ao deslocamento do elemento-QU (QU-movido simples, QU-que e QU-é-que). Mas o que explicaria a maior frequência de QU-movido simples em lugar de QU-que, apesar do que se nota na fala adulta? Nesse caso, a economia derivacional parece ser uma pista mais significativa para a criança do que a frequência: visto que ambas as estruturas envolvem movimento-QU, a construção mais econômica, o QU-movido simples, será preferida em lugar da mais frequente, o QU-que, que envolve um maior custo derivacional associado à inserção do complementizador “que”. Importante ressaltar que, diferentemente do que ocorre com a complexidade sintática, a criança ainda não parece utilizar as pistas pragmáticas como um gatilho robusto para o QU-*in situ*, sugerindo que as restrições de natureza pragmática são de aquisição mais tardia que as de natureza gramatical.

## Conclusão

Este trabalho explora a complexidade da aquisição do QU-*in situ* no português brasileiro, enfatizando os aspectos sintáticos e pragmáticos que diferenciam as estruturas e influenciam em sua produção. Discutimos aqui a aparente assimetria entre a fala infantil e adulta: enquanto as crianças nunca (ou quase nunca) produzem a construção, temos evidências na literatura de taxas produtivas na fala adulta. Entretanto, a partir da investigação de dados espontâneos, notamos que, na fala dirigida à criança, o QU-*in situ* não é tão produtivo quanto na fala entre adultos. Ou seja, a assimetria atribuída ao PB é fruto da comparação entre *corpora* de naturezas diferentes, que não poderiam, a princípio, ser diretamente comparados.

Os resultados sugerem que, embora as crianças tendam a seguir o padrão geral dos adultos favorecendo o QU-*ex situ*, elas não replicam completamente esse comportamento, pois há outros fatores que contribuem para suas primeiras produções linguísticas. Vimos, por exemplo, que o custo derivacional parece fundamental na escolha da criança entre QU-movido simples e QU-que, uma vez que, mesmo ouvindo em seu *input* um maior número de instâncias de QU-que, ela opta pela opção mais econômica, o QU-movido simples, que, na análise adotada aqui, envolve apenas o movimento do pronome interrogativo para a posição inicial, sem a inserção adicional de elementos. Além disso, os dados apontam para uma interessante distinção com relação à influência que fatores morfossintáticos e pragmáticos desempenham na aquisição dessas estruturas: o QU-*in situ*, apesar de pouco utilizado, foi favorecido por contextos sintáticos – por exemplo, diante de constituintes-QU mais complexos. Já no domínio pragmático, embora a criança pareça respeitar a restrição absoluta (nunca produzindo em contextos *out-of-the-blue*), a sensibilidade às nuances discursivas parece ser adquirida mais tardiamente, não alcançando ainda a mesma força preditiva observada nos adultos. Esses achados contribuem para debates mais amplos na literatura, especialmente no que se refere ao papel do *input* e da interação entre diferentes fatores que guiam os primeiros estágios do desenvolvimento linguístico das crianças.

## Informações complementares

Avaliação e resposta dos autores

Avaliação: <https://doi.org/10.25189/rabralin.v24i1.2324.R>

Resposta dos autores: <https://doi.org/10.25189/rabralin.v24i1.2324.A>

Editora

Raquel Meister Ko. Freitag

Afiliação: Universidade Federal de Sergipe

ORCID: <https://orcid.org/0000-0002-4972-4320>

RODADAS DE AVALIAÇÃO

Avaliador 1: Livia Oushiro

Afiliação: Universidade Estadual de Campinas

ORCID: <http://orcid.org/0000-0002-2165-3305>

Avaliador 2: Ronaldo Manguiera Lima Jr.

Afiliação: Universidade de Brasília

ORCID: <https://orcid.org/0000-0002-8610-0306>

AVALIADOR 1

O artigo “Pistas estatísticas e a aquisição de perguntas QU-in situ e QU-ex situ em Português Brasileiro” apresenta uma análise original, muito bem motivada e embasada, que discute incongruências entre dados de produção de QU-in situ na fala adulta e de crianças reportados previamente na literatura. Ancoradas em um corpus robusto de 154 horas de gravação, as autoras estabelecem hipóteses e questões claras, e as perguntas são respondidas por meio de minuciosa análise estatística e descritiva. O texto encontra-se muito bem organizado e claramente redigido, e a discussão se baseia em sólidos dados empíricos. Por esses motivos, sou favorável à publicação do artigo.

Chamo a atenção para três aspectos da análise e da apresentação dos resultados que merecem revisão:

1) Metodologia: A apresentação da metodologia levanta questões quanto aos dados que foram efetivamente incluídos nas análises quantitativas; a maior parte dos questionamentos é sanada na seção de resultados, mas tais questões podem ser evitadas se os casos de exclusão já forem apresentados na seção de metodologia;

2) Modelos de regressão: Todos os modelos de regressão incluem apenas efeitos fixos, mas já é praxe na área de Linguística incluir efeitos aleatórios. Recomendo fortemente que os modelos sejam refeitos com a inclusão do

falante individual como um efeito aleatório, a fim de que se possa evitar o questionamento de possíveis efeitos idiossincráticos. Além disso, recomendo testar se todas as variáveis incluídas nos modelos são de fato ortogonais entre si; em especial, é provável que Palavra.QU e Função.QU não sejam totalmente independentes. Os modos de implementar esses refinamentos estão descritos no manuscrito.

### 3) Interpretação de resultados estatísticos:

a. Na leitura das estimativas, dá-se a entender frequentemente que os valores-p para cada preditor estão sendo interpretado *por si*, e não em relação ao valor do Intercepto. Um fator não tem significância, mas sim a diferença de estimativa em relação a outros fatores da variável – o que resulta, em última instância, na correlação significativa entre a variável dependente e a variável independente. Sugeri reescritas de vários trechos para evitar esse erro de interpretação.

b. Na Tabela 6 (p.21), em que se testa a interação entre Contexto \* Falante, as estimativas devem ser lidas em relação ao Intercepto = Adulto:discursivo. Dois resultados são lidos incorretamente:

(i) “todos os contextos significativos desfavoreceram a utilização do QU-*in situ* em comparação ao *ex situ* na fala de ambos os grupos” (p.21): os fatores que apresentaram diferença significativa em relação ao intercepto de fato são todos desfavorecedores (estimativa < 1 em odds-ratio), mas a comparação dessas estimativas não é entre QU-*in situ* vs. QU-*ex situ*, mas sim entre cada termo (p.ex., Adultos: conhecimento prévio) vs. intercepto (adulto:discursivo). A comparação entre QU-*in situ* e QU-*ex situ* se encontra apenas no Intercepto: a estimativa de 0,228 (abaixo de 1) indica que QU-*in situ* é despreferido em relação a QU-*ex situ*.

(ii) “[...] pode-se inferir que o contexto discursivo é um preditor robusto para a utilização de QU-*in situ* tanto por adultos quanto por crianças” (p.21). A estimativa criança:discursivo = 0,195 indica que as crianças desfavorecem QU-*in situ* em relação aos *adultos*. A tabela não permite inferir que o contexto discursivo é um preditor robusto para as crianças, mas apenas para os adultos. Para saber se o contexto discursivo favorece QU-*in situ* para as crianças, é necessário criar um modelo à parte, apenas com os dados das crianças, ou mudar o valor de referência do modelo da Tabela 6 para criança:discursivo. Esta interpretação equivocada será repetida adiante na discussão e na conclusão do artigo, o que precisará ser revisado. Entretanto, os resultados da Tabela 6 permitem chegar a outra conclusão, tão interessante quanto: pela baixa probabilidade geral de ocorrência de QU-*in situ* na fala das crianças, é bem possível que o contexto pragmático não seja um fator relevante, diferentemente do que ocorre com os adultos. Estes apresentam restrições gramaticais e pragmáticas, ao passo que as crianças parecem ter apenas restrições gramaticais (preferência por QU-*in situ* em construções complexas). Desse modo, a criança parece adquirir primeiro as restrições gramaticais, e a aquisição das restrições pragmáticas é mais tardia. Isso, me parece, tem implicações interessantes para o processo de aquisição.

Os pontos (1) e (3a) são de fácil resolução. Embora os pontos (2) e (3b) envolvam novas análises estatísticas e reescrita de certos trechos, é pouco provável que os resultados se modifiquem substancialmente. Os apontamentos acima e outros comentários menores no manuscrito têm o objetivo de contribuir para a argumentação do artigo, que apresenta resultados muito interessantes e elucida o processo de aquisição de perguntas-QU.

AVALIADOR 2

O artigo apresenta uma investigação relevante para a área de aquisição da linguagem, com perguntas de pesquisa claras, hipóteses coerentes e uma análise robusta, tanto em termos de quantidade de dados quanto de rigor científico no desenho experimental e na análise. Os principais resultados advêm de uma análise descritiva, uma vez que o uso de perguntas *in situ* é muito baixo tanto na fala das crianças quanto na fala dos adultos dirigida a elas, o que, por si só, já fornece o insumo necessário para as discussões. Ainda assim, as autoras realizaram uma análise inferencial por meio de modelos de regressão logística, decisão que considero acertada e que, certamente, adiciona rigor metodológico à investigação.

No entanto, há necessidade de alguns esclarecimentos e de correções pontuais na interpretação dos coeficientes do modelo, todos indicados nos comentários que deixei ao longo do texto devolvido em anexo. Ressalto também a importância de reportar algumas das probabilidades previstas pelo modelo (transformadas a partir dos *odds ratios*) e de incluir os participantes como efeitos aleatórios.

Parabenizo as autoras por conduzirem uma discussão aprofundada dos resultados, articulando-os às perguntas de pesquisa e às hipóteses propostas inicialmente. Trata-se de um trabalho de impacto e relevância, que considero apto para publicação após as devidas correções e esclarecimentos.

#### RESPOSTA DOS AUTORES

Prezados Editores e Pareceristas da Revista ABRALIN,

Gostaríamos de expressar nosso sincero agradecimento pela leitura atenta e pelos valiosos comentários sugeridos para o aperfeiçoamento do manuscrito. As observações contribuíram significativamente para aumentar o rigor metodológico das análises estatísticas e a clareza da argumentação teórica. Informamos que todas as sugestões foram cuidadosamente consideradas e incorporadas à nova versão do texto. Nos casos em que não foi possível implementar exatamente o que foi solicitado (devido a limitações intrínsecas aos dados), apresentamos uma justificativa técnica detalhada e adotamos a melhor alternativa disponível.

Abaixo, detalhamos ponto a ponto as principais alterações realizadas.

#### 1. Distribuição do Corpus

Foi sugerido que os números absolutos da Tabela 2 fossem trocados pelos números após a aplicação dos critérios de exclusão, o que atendemos integralmente.

#### 2. Esclarecimentos Metodológicos

Foi sugerido também que deixássemos mais claros os critérios de exclusão, o que atendemos também integralmente. Explicitamos no texto a exclusão do item 'cadê' e deixamos mais claro que apenas as perguntas em que tanto o QU-*in situ* quanto o *ex situ* fossem alternativas possíveis foram consideradas na análise. Ou seja, apenas perguntas com movimento-QU de fato opcional entraram na análise.

Também explicamos por que optamos pelo "como" para intercepto na análise estatística. De fato, ele é o default do modelo por ordem alfabética, mas não achamos conveniente mudar por uma razão: como eles apresentam menos casos de QU-*in situ* nos nossos dados, usá-lo como base faz com que os resultados mostrem o quanto os outros fatores aumentam a chance dessa ocorrência (resultando em números

positivos e maiores que 1), o que nos parece mais intuitivo de interpretar do que reduções. O mesmo vale para intercepto adjunto.

### 3. Visualização dos Gráficos

Conforme muito bem apontado, a visualização dos gráficos em tons de cinza dificultou a distinção entre as categorias, principalmente no Gráfico 2. Como as diretrizes da Revista pedem que os gráficos sejam preferencialmente representados em tons de cinza, optamos pela manutenção desses tons, com adição de texturas. Cremos que assim a visualização não será comprometida.

### 4. Análise Estatística dos Dados Adultos (Tabela 4)

Foi sugerido o uso de modelos de efeitos mistos para controlar a variação individual dos falantes e garantir a independência das observações. Para atender à sugestão, substituímos o modelo anterior por um Modelo Linear Generalizado Misto (GLMM), incluindo interceptos aleatórios para cada falante (1 | Falante).

Além disso, calculamos para o modelo adulto o Fator de Inflação da Variância (VIF) para todas as variáveis preditoras. Os resultados confirmaram a ausência de colinearidade prejudicial (todos os VIFs < 2), validando a inclusão simultânea de Palavra-QU, Função Sintática, Complexidade e Tipo de Verbo no mesmo modelo.

Refinamos os níveis das variáveis, excluindo a categoria “Outras” da função sintática para evitar distorções estatísticas (separação completa) e garantir estimativas mais estáveis para as demais variáveis, já que esta categoria incluía algumas perguntas não necessariamente opcionais como “isso aqui é um o quê?” ou “(es)tá um pouquinho o quê, filha?” a que chamamos de Probe Questions (Nguyen & Legendre, 2021). Elas são perguntas fragmentadas, com um DP parcialmente preenchido e, por isso, o QU-in situ é a opção preferida nesse caso. Os casos são interessantes, mas recentemente submetemos esses resultados a outra revista, por isso optamos por tirá-las dessa análise e deixar apenas as perguntas em que tanto a contraparte movida como in situ eram alternativas viáveis.

Por fim, alteramos toda a visualização das tabelas atendendo às sugestões. Reportamos os resultados completos com Odds Ratios, Intervalos de Confiança de 95%, Pvalores e R, conforme solicitado, conferindo maior transparência e legibilidade dos números.

### 5. Análise Estatística dos Dados Infantis (Tabela 5)

Foi sugerida a aplicação da mesma modelagem robusta (GLMM) para os dados das crianças. No entanto, devido à natureza dos dados de aquisição (amostra com alta esparsidade e muitas células com frequência zero para a construção com QU-in-situ em diversas combinações de variáveis), o modelo misto apresentou problemas de convergência.

Por isso, optamos por manter um Modelo Linear Generalizado (GLM) simples para as crianças. No texto (Seção 4.2), incluímos uma explicação clara sobre essa limitação metodológica imposta pela distribuição dos dados, assegurando ao leitor que essa foi a escolha estatística mais honesta e adequada para representar o comportamento infantil. E o motivo para não termos conseguido implementar o modelo é

justamente nossa argumentação: a criança produz quase categoricamente o QU-ex situ e apenas muito raramente o QU-in situ.

Quanto ao diagnóstico de colinearidade, não aplicamos o cálculo do VIF para o modelo infantil pela mesma razão da esparsidade dos dados (a falta de variância em células específicas inviabiliza o cálculo). No entanto, como as variáveis linguísticas testadas são as mesmas do modelo adulto (em que o VIF confirmou a independência), assumimos razoavelmente que não há violação de ortogonalidade entre os preditores também para as crianças.

### 6. Análise dos Fatores Pragmáticos (Seção 5.2 / Tabela 6)

Foi sugerido o refinamento da análise da interação entre Grupo e Contexto, aplicação do modelo com efeitos mistos, e ajuste de um ponto da interpretação teórica, ao que atendemos integralmente.

Implementamos um GLMM que testou explicitamente a interação entre ‘Grupo’ (Adulto/Criança) e ‘Contexto Pragmático’. A interpretação dos resultados foi reescrita para clarificar a distinção entre restrições gramaticais e nuances pragmáticas. Argumentamos agora, com base nos novos números, que embora as crianças respeitem a restrição absoluta (não usar QU-in-situ em contextos out-of-the-blue), elas ainda não possuem a “sensibilidade às nuances discursivas” que os adultos demonstram. Isso fortaleceu a discussão sobre a hipótese de “Gramática antes da Pragmática”, o que nos pareceu extremamente interessante. Tão interessante, que optamos por trocar o título do artigo, se possível, para refletir este novo achado.

Acreditamos que essas modificações tenham elevado substancialmente a qualidade do artigo. Agradecemos, portanto, aos pareceristas Profa. Dra. Livia Oushiro e Prof. Dr. Ronaldo Manguera Lima Jr pela leitura atenta e pelo cuidado que demonstraram na inspeção das análises estatísticas que não só fortaleceram a argumentação, mas também a deixaram mais precisa, honesta e robusta.

Colocamo-nos à disposição para quaisquer outros esclarecimentos.

Atenciosamente,

Clariana Vieira e Elaine Grolla

### Conflito de Interesse

As autoras não têm conflitos de interesse a declarar.

## Protocolo e Pré-Registro de Pesquisa

Avaliando os roteiros propostos pela Equator Network, consideramos que nenhum deles se mostra relevante para a pesquisa em tela. Também informamos que a pesquisa desenvolvida não foi pré-registrada em repositório institucional independente.

## Declaração de Disponibilidade de Dados

Por se tratar de dados infantis não anonimizados (*corpus* Santana-Santos), os dados e materiais que suportam os resultados deste estudo estão disponíveis para consulta sob demanda em drive institucional de responsabilidade das autoras.

## Pesquisa com seres humanos

O *corpus* utilizado não requer aprovação do Comitê de Ética em Pesquisa por estar em conformidade com a Resolução nº 510, de 7 de abril de 2016, artigo 1º, parágrafo único, inciso V. Em atendimento à Resolução, a identidade de todos os participantes desta gravação foi mantida em sigilo. Seus nomes e apelidos foram substituídos por pseudônimos.

## Fontes de financiamento

Esta pesquisa foi financiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – processo 140447/2019-9 (atribuído à primeira autora) e processo 310213/2023-1 (atribuído à segunda autora).

## REFERÊNCIAS

AMBAR, M. How do wh-exclamatives and in situ wh-questions correlate? **Comunicação apresentada ao XXVI Incontro di Grammatica Generativa**, Roma, 2000.

BAUNAZ, L.; BONAN, C. Activation levels: A fresh perspective on French wh in-situ. **Isogloss: Open Journal of Romance Linguistics**, Barcelona, vol. 9, n.1, p. 1-43, 2023. DOI <https://doi.org/10.5565/rev/isogloss.305>.

BIEZMA, M. Givenness and the difference between WH-fronted and WH-*in-situ* questions in Spanish. *In: Romance Languages and Linguistic Theory: Selected papers for Going Romance 29*, vol. 29, p. 21-39, 2018. DOI <https://doi.org/10.1075/rllt.13.03bie>.

BOD, R.; HAY, J.; JANNEDY, S. (Eds.). **Probabilistic Linguistics**. Cambridge: The MIT Press, 2003. Disponível em: <https://mitpress.mit.edu/9780262523387/probabilistic-linguistics/>.

BOECKX, C. Decomposing French Questions. *In: Proceedings of the 23rd Annual Penn Linguistics Colloquium*, J. Alexander, N.R. Han and M. Minnick Fox (eds.), Philadelphia, University of Pennsylvania, p. 69–80, 1999.

Disponível em: <https://repository.upenn.edu/handle/20.500.14332/45490>.

BONAN, C. From northern Italian to Asian wh-in situ: A theory of low focus movement. **Isogloss: Open Journal of Romance Linguistics**, v. 7, p. 1-59, 2021. DOI <https://doi.org/10.5565/rev/isogloss.108>.

BYBEE, J.; HOPPER, P. (Eds.) **Frequency and the Emergence of Language Structure**. Amsterdam: John Benjamins, 2001. DOI <https://doi.org/10.1075/tsl.45>.

CHANG, H. The Acquisition of Chinese Syntax. **Language Processing in Chinese**. In: H.C. Chen and O.J.L. Tzeng (eds.), p. 277-311, 1992. DOI [https://doi.org/10.1016/S0166-4115\(08\)61895-6](https://doi.org/10.1016/S0166-4115(08)61895-6).

CHANG, L. **Wh-in-situ phenomena in French**. Dissertação de Mestrado - University of British Columbia, Vancouver, 1997. DOI <https://doi.org/10.14288/1.0087736>.

CHOMSKY, N. Principles and parameters in syntactic theory. In: HORNSTEIN, N.; LIGHTFOOT, D. (Eds.). **Explanation in linguistics: The logical problem of language acquisition**. London: Longman, p. 32-75, 1981.

CHOMSKY, N. **The Minimalist Program**. Massachusetts: MIT Press, 1995. Disponível em: <https://mitpress.mit.edu/9780262531283/the-minimalist-program/>.

CHOMSKY, N. Derivation by phase. **MIT Occasional Papers in Linguistics**, n. 18, p. 482-496, 1999.

DEROMA, C. L. Z. **Divide et Impera: Separating Operators from their Variables**. Tese de Doutorado - University of Connecticut, Connecticut, 2011.

DIESEL, H.; TOMASELLO, M. The development of relative clauses in spontaneous child speech. **Cognitive Linguistics**, v.11, n.1-2, p. 131-151, 2000. DOI <https://doi.org/10.1515/cogl.2001.006>.

FIGUEIREDO SILVA, M. C.; GROLLA, E. Some syntactic and pragmatic aspects of WH-in-situ in Brazilian Portuguese. In: KATO, Mary A.; ORDÓÑEZ, Francisco (Eds.). **The Morphosyntax of Portuguese and Spanish in Latin America**, p. 259-285, 2015.

GROLLA, E. **Aquisição da periferia esquerda da sentença em português brasileiro**. Dissertação de mestrado, Universidade Estadual de Campinas, São Paulo, 2000.

GROLLA, E. Speculations about the Acquisition of Wh-Questions in Brazilian Portuguese. In: PIRES, A.; ROTHMAN, J. (Eds.). **Minimalist Inquiries into Child and Adult Language Acquisition: Case Studies across Portuguese**, Mouton deGruyter, p. 85-104, 2009. DOI <https://doi.org/10.1515/9783110215359>.

GUASTI, M. T. An excursion into interrogatives in early English and Italian. In: Friedemann, M. A., & Rizzi, L. (Eds.). **The acquisition of syntax: Studies in comparative developmental linguistics**, Routledge, p. 105-128, 2014. DOI <https://doi.org/10.4324/9781315839899>.

HAMLAOUI, F. On the role of phonology and discourse in Francilian French wh-questions. **Journal of linguistics**, p. 129-162, 2011. DOI <https://doi.org/10.1017/S0022226710000198>.

HORNSTEIN, N.; NUNES, J.; GROHMANN, K. K. **Introduction to Minimalist Program**. Cambridge: Cambridge University Press, 2005.

HUANG, C. T. J. Move wh in a language without wh-movement. **The linguistic review**, vol. 1, n. 4, p. 369-416, 1982. DOI <https://doi.org/10.1515/tlir.1982.1.4.369>.

HUDDLESTON, R.; PULLUM, G. **The Cambridge Grammar of the English Language**. Cambridge: Cambridge University Press, 2002. DOI <https://doi.org/10.1515/zaa-2005-0209>.

HUTTENLOCHER, J.; VASILYEVA, M.; CYMERMAN, E.; LEVINE, S. Language input and child syntax. **Cognitive Psychology**, n.45, n. 3, p. 337–374, 2002. DOI [https://doi.org/10.1016/S0010-0285\(02\)00500-5](https://doi.org/10.1016/S0010-0285(02)00500-5).

JIMÉNEZ, M. L. **Semantics and pragmatics conditions on word order in Spanish**. Tese de Doutorado - Georgetown University, Washington D.C., 1997.

KRIFKA, M. Questions. In: MAIENBORN, C.; HEUSINGER, K. von; PORTNER, P. (Eds.). **Semantics: An international handbook of natural language and meaning**, Berlin: Mouton de Gruyter, v. 2, p. 1742-1784, 2011.

LACERDA, R. **Middle-field syntax and Information Structure in Brazilian Portuguese**. Tese de doutorado - University of Connecticut, Connecticut, 2020.

LESSA-DE-OLIVEIRA, A. S. C. **Aquisição de constituintes-QU em dois dialetos do português brasileiro**. Dissertação de Mestrado - Universidade Estadual de Campinas, São Paulo, 2003.

LEVINSON, S. C. **Pragmática**. Tradução de Luís Carlos Borges e Aníbal Mari. São Paulo: Martins Fontes, 2007.

LIDZ, J.; GAGLIARDI, A. How nature meets nurture: Universal grammar and statistical learning. **Annual Reviews of Linguistics**, vol. 1, n. 1, p. 333-353, 2015. DOI <https://doi.org/10.1146/annurev-linguist-030514-125236>.

LOPES-ROSSI, M. A. G. **A sintaxe diacrônica das interrogativas-Q do Português**. Tese de Doutorado - Universidade Estadual de Campinas, São Paulo, 1996.

MIOTO, C. As interrogações no português brasileiro e o critério-WH. **Letras de Hoje**, vol. 96, p. 19-33, 1994.

MUNARO, N.; POLETO, C.; POLLOCK, J. Eppur si muove! On comparing French and Bellunese Wh-movement. **Yearbook of Comparative Syntax**, p. 147-180, 2001. DOI <https://doi.org/10.1075/livy.1.07mun>.

NGUYEN, A. D.; LEGENDRE, G. The acquisition of wh-questions: Beyond structural economy and input frequency. **Language Acquisition**, v. 29, n. 1, p. 79-104, 2021. DOI <https://doi.org/10.1080/10489223.2021.1968867>.

OUSHIRO, L. Condicionamento discursivo-pragmático no uso variável de Interrogativas-Q. **Estudos Linguísticos**, vol. 39, n. 2, p. 628-639, 2010. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/453>.

OUSHIRO, Livia. Análise sociolinguística da posição variável do constituinte interrogativo. **Revista da ABRALIN**, v.11, n.2, p. 43-87, 2012. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/1118>.

PERKINS, L.; FELDMAN, N.; LIDZ, J. Learning an input filter for argument structure acquisition. **Proceedings of the 7th workshop on cognitive modeling and computational linguistics**, p. 11–19, 2017. DOI <https://doi.org/10.18653/v1/W17-0702>.

PIRES, A.; TAYLOR, H. The syntax of wh-in-situ and common ground. **Proceedings from the Annual Meeting of the Chicago Linguistic Society**, Chicago, vol. 43, n. 2, p. 201-215, 2007.

POLETO, C.; POLLOCK, J. Arguing for remnant movement in Romance. In: GREWENDORF, G. (Ed.). **Remnant Movement**. Mouton De Gruyter, p. 135-178, 2015. DOI <https://doi.org/10.1515/9781614516330-006>.

QUIRK, R.; GREENBAUM, S.; LEECH, G.; SVARTVIK, J. **A Comprehensive Grammar of the English Language**. New York: Longman, 1985.

REGLERO, L. **A'-Dependencies in Spanish and Basque**. Tese de doutorado - University of Connecticut, Connecticut, 2004.

REINHART, T. Wh-in-Situ in the Framework of the Minimalist Program. **Natural Language Semantics**, vol. 6, n. 1, p. 29-56, 1998. Disponível em: <https://link.springer.com/article/10.1023/A:1008240014550>.

RIZZI, L. The fine structure of the left periphery. In: HAEGEMAN, L. (Ed.). *Elements of Grammar: Handbook of Generative Syntax*, Kluwer, p. 281-337, 1997. Disponível em: [https://link.springer.com/chapter/10.1007/978-94-011-5420-8\\_7](https://link.springer.com/chapter/10.1007/978-94-011-5420-8_7).

SANTOS, R. S. **Projeto de Aquisição do Ritmo em Português Brasileiro**. FAPESP 2003/13565-4, 2005.

SCHNEIDER, J; PERKINS, L; FELDMAN, N. A noisy channel model for systematizing unpredictable input variation. **Proceedings of the 44th Annual Boston University conference on Language Development**, MA, Cascadilla Press, p. 533-547, 2020. Disponível em: <https://www.lingref.com/buclid/44/BUCLD44-43.pdf>.

SIKANSI, N. As interrogativas-Q na gramática infantil do PB. **Caderno de Estudos Linguísticos**, São Paulo, vol. 36, p. 85-103, 1999. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8637106>.

SILVEIRA, V. L. **A Emergência de Estruturas A-Barra no Contexto da Aquisição do Português Brasileiro como Língua Materna**. Tese de Doutorado - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

STALNAKER, R. Assertion. In: COLE, P. (Ed.). *Syntax and Semantics, Vol. 9: Pragmatics*. vol. 9, New York, Academic Press, p. 315-332, 1978.

URIBE-ETXEBARRIA, M. In situ questions and masked movement. **Linguistic Variation Yearbook**, vol. 2, p. 259–303, 2002. DOI <https://doi.org/10.1075/livy.2.09uri>.

VIEIRA, C.; GROLLA, E. The pragmatics of wh-in-situ questions in Brazilian Portuguese: data from child and adult language. **Proceedings of the 44th Boston University Conference on Language Development**, MA, Cascadilla Press, p. 677-690, 2020. Disponível em: <http://www.lingref.com/buclid/44/BUCLD44-54.pdf>.

VIEIRA, C. **Aquisição do QU-in situ no português brasileiro e no francês**. Tese de Doutorado – Universidade de São Paulo, São Paulo, 2023. Disponível em: <https://www.teses.usp.br/teses/disponiveis/8/8139/tde-21112023-140101/>.

VLACHOS, C.; CHIOU, M. The syntax, semantics and pragmatics of ‘optional’ wh-in situ in Greek. **Journal of Greek Linguistics**, vol. 20, n. 1, p. 102-131, 2021.

YANG, C. For and against frequencies. **Journal of Child Language**. Cambridge University Press 42, p. 287-293, 2015. DOI <https://doi.org/10.1017/S0305000914000683>.

ZUCKERMAN, S. **The Acquisition of “Optional” Movement**. Tese de Doutorado. Boston, Massachusetts, 2001. Disponível em <https://pure.rug.nl/ws/portalfiles/portal/3089667/c2.pdf>.