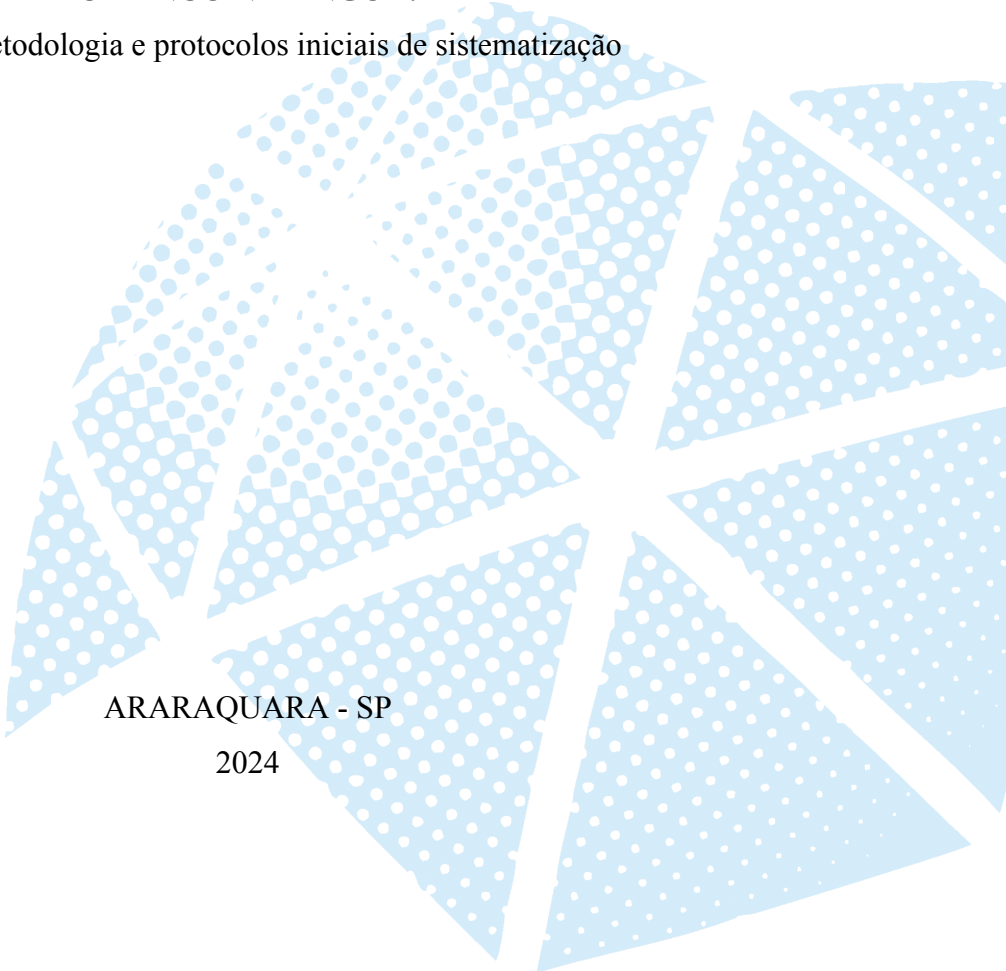


UNIVERSIDADE ESTADUAL PAULISTA - UNESP
Faculdade de Ciências e Letras - Campus de Araraquara

ANA LAURA GIANESELLA-FERREIRA

O BANCO NALÍNGUA:
dados, metodologia e protocolos iniciais de sistematização

ARARAQUARA - SP
2024



ANA LAURA GIANESELLA-FERREIRA

O BANCO NALÍNGUA:

dados, metodologia e protocolos iniciais de sistematização

Monografia de Conclusão de Curso (MCC)
apresentada à Universidade Estadual Paulista
(UNESP), Faculdade de Ciências e Letras,
Araraquara - SP, para obtenção do título de
Bacharela em Letras.

Área de Concentração: Aquisição da
Linguagem

Orientador(a): Prof^a. Dr^a. Alessandra Del Ré

Coorientador(a): Prof^a. M^a. Fernanda Martins
Moreira

ARARAQUARA - SP

2024

G433b Ganesella-Ferreira, Ana Laura
O BANCO NALÍNGUA: : dados, metodologia e protocolos iniciais
de sistematização / Ana Laura Ganesella-Ferreira. -- Araraquara,
2024
48 p.

Trabalho de conclusão de curso (Bacharelado - Letras) -
Universidade Estadual Paulista (UNESP), Faculdade de Ciências e
Letras, Araraquara

Orientadora: Alessandra Del Ré

Coorientadora: Fernanda Martins Moreira

1. Grupo NALíngua. 2. Banco de Dados Infantis. 3. Metodologia. 4.
Processo de constituição. 5. Aquisição da Linguagem. I. Título.
Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

ANA LAURA GIANESELLA-FERREIRA

O BANCO NALÍNGUA:

dados, metodologia e protocolos iniciais de sistematização

Monografia de Conclusão de Curso (MCC) apresentada à Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Araraquara - SP, para obtenção do título de Bacharela em Letras.

Área de Concentração: Aquisição da Linguagem

Orientador(a): Prof^ª. Dr^ª. Alessandra Del Ré

Coorientador(a): Prof^ª. M^a. Fernanda Martins Moreira

Data da defesa: 05/12/2024

Banca Examinadora:

Prof^ª. Dr^ª. Alessandra Del Ré

UNESP - Faculdade de Ciências e Letras - Campus de Araraquara

Prof^ª. Dr^ª. Marianne Carvalho Bezerra Cavalcante

UFPA - Universidade Federal da Paraíba

Prof. Me. Francisco Ebson Gomes Sousa

UFERSA - Universidade Federal Rural do Semi-Árido

AGRADECIMENTOS

Agradeço aos meus pais, Rafael e Fernanda, e aos meus avós Maria Cecília, Hilda e Nelson, por todo o suporte, carinho e incentivo ao longo da trajetória percorrida nestes seis anos (Graduação e Reingresso), como também pela inspiração para superar os desafios nela enfrentados.

Ao meu parceiro, Thiago, pela paciência, compreensão, carinho e incentivo, nos momentos prazerosos e árduos, condições distintas e inseparáveis da vida acadêmica. Sobretudo, pelas palavras de encorajamento e pelo acolhimento nas situações difíceis durante essa jornada.

Aos amigos e amigas que, comigo, trilharam o caminho, se fazendo presentes e amáveis nos momentos bons e ruins. À vocês, além de minha gratidão, desejo toda a felicidade e sucesso em suas vidas.

Aos docentes que me deram as bases para ingressar na Universidade e, também, aos que participaram da minha formação superior. Em especial, à professora Ana Maria Pereira Sbragia, uma grande incentivadora da minha opção e paixão pela Letras.

À minha orientadora, Alessandra Del Ré e à minha coorientadora Fernanda Martins Moreira, pelo acolhimento e disposição em me direcionar. Aos demais membros do Grupo GEALin por todas as trocas e contribuições.

RESUMO

Esta Monografia de Conclusão de Curso foi redigida em formato de revisão teórica e, igualmente, de relato de experiência sobre o fazer metodológico da constituição de um banco de dados de fala de crianças, registradas em diferentes contextos (escolar, familiar, com desenvolvimento típico e atípico, *etc.*) e em processos distintos de Aquisição da Linguagem (oral ou escrita). Trata-se da descrição dos procedimentos metodológicos empreendidos na construção do banco de dados do grupo NALíngua-CNPq (Del Ré *et al.*, 2016; Hilário *et al.*, 2024), grupo este que reúne pesquisadores de diferentes universidades, brasileiras e francesas, tendo como objetivo comum estudar a linguagem da criança, desde as primeiras vocalizações até a entrada na escrita, por diferentes olhares teóricos que elegem a interação como um elemento-chave dentro desse percurso aquisicional. Logo, tomando como ponto de partida os trabalhos de Perroni (1996) e Possenti (1996), esta pesquisa procura retomar a discussão da relação entre o dado e o fazer metodológico em pesquisas aquisicionais como fatores que incidem na decisão e na produção de protocolos, assim como nas características físicas (como modo de armazenamento, compartilhamento, transcrição, dentre outros) da elaboração de um banco de dados infantis inaugurado no Brasil. Diante disso, apresentamos aqui um levantamento das pesquisas e perspectivas adotadas pelos pesquisadores do grupo até o momento, fornecendo um panorama geral das informações que compõem um banco de dados e indicamos protocolos gerais para o armazenamento, coleta, manutenção de dados e análise, de modo a compatibilizá-los com as naturezas investigativas que integram o grupo NALíngua (CNPq).

Palavras-chave: Grupo NALíngua; Banco de Dados Infantis; Metodologia; Processo de constituição; Aquisição da Linguagem.

ABSTRACT

This Monograph was written as a theoretical review and also as an experience report on the methodological process of setting up a database of children's speech, recorded in different contexts (school, family, with typical and atypical development, etc.) and in different processes of Language Acquisition (oral or written). This is a description of the methodological procedures undertaken to build the database of the NALíngua-CNPq group (Del Ré et al., 2016; Hilário *et al.*, 2024), a group that brings together researchers from different universities, both Brazilian and French, with the common goal of studying children's language, from the first vocalizations to the entry into writing, from different theoretical perspectives that choose interaction as a key element in this acquisition process. Therefore, taking the works of Perroni (1996) and Possenti (1996) as a starting point, this research seeks to take up the discussion of the relationship between data and methodological practice in acquisitional research as factors that affect the decision and production of protocols, as well as the physical characteristics (such as storage, sharing, transcription, among others) of the development of a children's database inaugurated in Brazil. In the light of this, we present here a survey of the research and perspectives adopted by the group's researchers to date, providing an overview of the information that makes up a database and indicating general protocols for storing, collecting, maintaining data and analyzing it, in order to make them compatible with the investigative nature of the NALíngua group (CNPq).

Keywords: NALíngua Group; Children's Database; Methodology; Process of constitution; Language Acquisition.

LISTA DE FIGURAS

Figura 1 – Armazenamento em *bytes*

35

LISTA DE TABELAS

Tabela 1 – *Corpus* NALíngua

29

LISTA DE ABREVIATURAS E SIGLAS

CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CHAT	<i>Codes for the Human Analysis of Transcripts</i>
CHILDES	<i>Child Language Data Exchange System</i>
CLAN	<i>Computerized Language Analysis</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
ELAN	<i>Eudico Linguistic Annotator</i>
FCLAr	Faculdade de Ciências e Letras de Araraquara
FDC	Fala Dirigida à Criança
FFLCH	Faculdade de Filosofia, Letras e Ciências Humanas
FI	Fala Infantil
GEALin	Grupo de Estudos em Aquisição da Linguagem
IES	Instituição de Ensino Superior
LEAL	Laboratório de Estudos em Aquisição da Linguagem
NALíngua	Núcleo de Estudos em Aquisição da Linguagem
PUC	Pontifícia Universidade Católica
UFPB	Universidade Federal da Paraíba
UNICAMP	Universidade Estadual de Campinas
USP	Universidade de São Paulo

SUMÁRIO

1 INTRODUÇÃO	12
2 METODOLOGIA	13
3 O DADO NA LINGUÍSTICA: A DELIMITAÇÃO DO OBJETO DE PESQUISA E SUA RELAÇÃO COM A POSTURA METODOLÓGICA	14
3.1 O Dado em Aquisição da Linguagem	15
4 O FAZER CIENTÍFICO EM AQUISIÇÃO: PRINCIPAIS TEORIAS E METODOLOGIAS	18
4.1 Estudos longitudinais versus transversais	22
4.2 Estudos qualitativos versus quantitativos	23
5 A CONSTITUIÇÃO DE BANCOS DE DADOS DE FALA	25
5.1 A relação entre dado, autoria e a formação de Banco de Dados	26
6 BANCO NALÍNGUA: A FORMAÇÃO DE UM CORPUS COMUM NACIONAL	28
6.1 As diferentes abordagens e ferramentas que permeiam o fazer-científico do NALíngua	30
6.1.1 CLAN	32
6.1.2 ELAN	32
6.1.3 GRILLE	32
6.1.4 PRAAT	33
6.1.5 OUTRAS FERRAMENTAS COMPUTACIONAIS	33
6.2 Uma constante (re)elaboração: procedimentos de organização	34
7 CONSIDERAÇÕES FINAIS	38
REFERÊNCIAS	40
APÊNDICE A: QUESTIONÁRIO (GOOGLE FORMS)	43
APÊNDICE B: ÍNDICE DE COLETA	45
APÊNDICE C: ÍNDICE DE TRANSCRIÇÕES E MÍDIAS	46
APÊNDICE D: AUTORIZAÇÃO DE USO DE DADOS	47
APÊNDICE E: TERMO DE COMPROMISSO DE UTILIZAÇÃO DE DADOS	48

1 INTRODUÇÃO

O interesse pelo desenvolvimento da linguagem infantil é compartilhado por diversas áreas do conhecimento, como a linguística, a psicologia e as ciências cognitivas. A fala da criança não só nos oferece informações valiosas sobre os aspectos linguísticos desta fase, mas, igualmente, nos diz muitas coisas sobre a linguagem humana, de modo geral (François, 2006). Assim sendo, com o advento das tecnologias, bem como do desejo de compartilhamento e reunião de informações sobre esta temática dentro da comunidade científica, surgem diferentes iniciativas de criação daquilo que se denominou “banco de dados”.

Um banco de dados, dentro das especificidades da Aquisição da Linguagem, nada mais é do que um repositório estruturado, de amostras de fala e/ou escrita de crianças em diferentes faixas etárias e inseridas em contextos comunicativos heterogêneos. Esses bancos compartilham, em maior ou menor escala, materiais que podem ser utilizados em pesquisas sobre diferentes aspectos da entrada da criança no circuito da linguagem, propiciando, do mesmo modo, a análise de unidades menores de língua, como morfemas e fonemas, e também a observação de instâncias mais amplas, como a do discurso.

A formação desses tipos de repositórios não se restringe apenas à coleta de amostras, mas também exige que seus curadores desenvolvam critérios claros para a transcrição e o armazenamento dos dados, de forma a favorecer a clareza e legibilidade dessas informações, além de garantir sua segurança e utilização ética. Nesse sentido, os bancos de dados são ferramentas essenciais para que mais informações acerca da linguagem possam ser trabalhadas e estudadas sob diferentes perspectivas, visto sua maior aceitabilidade, nos dias atuais, como uma ferramenta sólida de compartilhamento e troca dentro da comunidade científica.

Isto posto, o objetivo desta monografia é discorrer sobre os princípios teóricos e metodológicos envolvidos na construção de um banco de dados de fala infantil, mais particularmente, naquele que está sendo desenvolvido pelo grupo NALíngua-CNPq. Para isso, são abordadas as questões que cercam a relação entre o dado e a metodologia nas pesquisas em Aquisição, os desafios enfrentados ao longo da estruturação/sistematização desses repositórios e os traços específicos que caracterizam o fazer científico do grupo NALíngua-CNPq, formado por pesquisadores do Brasil e da França, que se debruçam sob um mesmo conjunto de dados, a partir de óticas distintas, interdisciplinares e complementares.

2 METODOLOGIA

O presente trabalho é, simultaneamente, um resultado de uma revisão teórica sobre as questões que permeiam a relação dado-metodologia e um relato dos procedimentos que estão sendo realizados durante o processo de sistematização do Banco NALíngua (Del Ré; Hilário; Rodrigues, 2016; Hilário; Moreira; Vieira, 2024). Dito isso, este texto pode ser dividido em três grandes seções: a primeira recupera a ideia de que dado e metodologias são instâncias que não podem ser desvinculadas, sendo, especificamente na área de Aquisição da Linguagem, dois elementos que abarcam alguns impasses em termos de delimitação e escolha.

Já na segunda e terceira seções temos, respectivamente: o apontamento de sobre a formação de Banco de Dados e, igualmente, indicações sobre a correlação dado-autoria-constituição desses repositórios; por fim, temos a apresentação do Banco NALíngua, bem como a descrição de alguns protocolos de organização arquivística que estão sendo implementados na triagem dos dados e que, nas próximas etapas de trabalho, serão utilizados como base para a reestruturação de um ambiente virtual para o armazenamento e acesso a essas informações.

Por fim, é importante dizer que, para falar sobre as questões que envolvem os dados e as metodologias de pesquisa em Aquisição, nos embasamos nos trabalhos de Perroni (1996), Possenti (1996), Scliar-Cabral (2021) e Del Ré (2023); já no que diz respeito constituição de banco de dados, utilizamos as informações consultadas nas plataformas oficiais de alguns desses repositórios, como a do CHILDES (MacWhinney; Snow, 2000) e a do Corpora FI e FDC (Santos; Toni, 2020), como também o artigo de Pessoa e Ramos (2017). Finalmente, para a coleta e mapeamento da realidade prática do grupo NALíngua-CNPq, elaboramos um formulário via *Google Forms* que pode ser consultado em “Apêndices”.

3 O DADO NA LINGUÍSTICA: A DELIMITAÇÃO DO OBJETO DE PESQUISA E SUA RELAÇÃO COM A POSTURA METODOLÓGICA

Nesta seção discutimos a questão do dado enquanto um elemento central para a delimitação do fazer metodológico, não só das pesquisas em si, mas também como uma unidade definidora para a construção dos critérios-base de organização de um banco de dados em função da natureza das informações coletadas. Isto posto, comentaremos brevemente sobre os pontos que foram delimitadores para aquilo que se denomina “objeto linguístico” e como isso reflete na seleção daquilo que Possenti (1996) descreve como “dado **dado**” (p. 195, destaques do autor) pela parte de um pesquisador durante o ato investigativo de um determinado fenômeno.

Desde a consolidação da Linguística Moderna como uma ciência autônoma, por um movimento do filósofo e, posteriormente, linguista Ferdinand de Saussure (1857-1913), tal área do conhecimento se preocupa com a natureza da linguagem humana em seus mais diversos aspectos e dimensões: social-interacional, mental-biológico e textual-discursivo, portanto, elegendo como tópico de estudo diferentes facetas da língua(gem) de uma dada comunidade de fala. Ainda que se recorte objetos distintos, de uma maneira geral, todos eles se relacionam com o funcionamento do sistema linguístico e buscam capturar, no ato de delimitação, os princípios constitutivos e mais representativos do assunto específico a ser estudado.

Nesse sentido, a opção por um certo conjunto de dados é uma atitude que exige a consideração de muitos fatores que envolvem, principalmente: **I)** o entendimento que se tem sobre “Língua”; **II)** o que dela se deseja estudar; e **III)** as possibilidades de investigação e de resposta que cada aporte teórico nos oferece para a exploração do objeto-dado almejado. É justamente por estas razões que a própria definição do que seja a natureza do dado – mesmo o fornecimento de uma descrição genérica – não é exatamente fácil de se propor: desde o primeiro “grande recorte” feito por Saussure, muitas concepções acerca da língua(gem) se modificaram e, com elas, novas subáreas e perguntas para desvendar seus variados mistérios surgiram.

Como uma consequência da ampliação dos interesses da Linguística, na qualidade de um campo científico e, igualmente, da proposição de novos olhares para se abordar os elementos que permeiam o funcionamento da linguagem em sua totalidade, isto é, a consideração de seus elementos estáveis (sistema) e do uso (a língua em atividade), temos

uma multiplicidade do que se é tido por dado nas variadas pesquisas realizadas dentro desta grande área. Assim sendo, pensamos ser produtivo tomarmos a caracterização de Possenti (1996, p. 195, destaques do autor) sobre o *dado dado* que consiste na seleção daquilo que é “dado realmente dado, que serve, que resolve, que é dado bom, dado mesmo” dentro de uma dada perspectiva teórica.

Para os propósitos desta monografia, a conceituação supracitada se faz proveitosa não somente porque situa o leitor da necessidade de termos à vista e de modo claro a natureza de nosso fazer científico enquanto linguistas, mas também porque abre margens para pensarmos as especificidades exigidas pelo “grande assunto” de cada um dos desdobramentos dos estudos linguísticos que vêm sendo formados desde a fundação de Saussure até os dias atuais. Ademais, todo dado, necessariamente, evoca uma atitude metodológica do pesquisador, ação esta que, ao mesmo tempo, decide os caminhos de investigação a serem percorridos e define o objeto observado, como já nos alertava o filósofo e linguista no *Curso de Linguística Geral* (1916)¹.

Tendo essas questões em consideração, ao pensarmos a relação entre a natureza do dado e a constituição de um banco de dados, fica claro, a depender dos interesses de observação, a necessidade de estabelecer critérios específicos, que incidam, sobretudo, sob o modo de armazenamento, a disponibilidade e uma padronização adequada com o tipo de informações coletadas pelo pesquisador ou grupo de pesquisadores. Dito isto, passemos à descrição do dado na área de Aquisição da Linguagem e os desafios que cercam o tratamento de tal elemento.

3.1 O Dado em Aquisição da Linguagem

Como dissemos, o dado está intimamente ligado ao fazer metodológico; nesse sentido, é preciso comentar sobre a grande variabilidade da prática científica dentro do campo de Aquisição da Linguagem. Para discorrermos acerca de tais desafios e a maneira como incidem no tratamento dos dados, nos pautamos, particularmente, nas discussões feitas por Maria Cecília Perroni no capítulo *O que é o dado em aquisição da linguagem?*, da obra *O método e o dado no estudo da linguagem*, organizada por Maria Fausta Pereira de Castro no ano de 1996, fruto dos Ciclos de Seminários em Aquisição da Linguagem realizados no Instituto de Estudos da Linguagem (IEL/UNICAMP).

¹ Fazemos menção à primeira data de publicação da obra.

Primeiramente, é preciso ter em mente que, embora a Aquisição da Linguagem seja autônoma e mais consolidada dentro da grande área de Linguagens e da Ciência, ela ainda é um campo relativamente recente se comparado às demais vertentes dos estudos linguísticos e também, aos outros ramos com os quais esteve, em sua formação, atrelada. A isso, Perroni (1996) atribui principalmente à sua natureza de “criação bípede da psicologia e da linguística, [...] tentando equilibrar-se sobre as próprias pernas” (p. 15) o que, desde o germen de pesquisas dessa natureza, impõe ao pesquisador o dilema de não cair em incoerências metodológicas, decorrentes da interdisciplinaridade que fundou e ainda permeia a área.

Em um momento inicial, que Perroni (1996, p. 15) denomina de “primeira infância” da Aquisição, observamos uma grande oscilação entre metodologias dessas duas áreas progenitoras – a Linguística e a Psicologia – para o estudo da linguagem da criança. Sobre isso, no livro *The acquisition of language: the study of developmental psycholinguistics* (1970), David McNeill afirma ser muito difícil caracterizar qualquer atitude de investigação desse campo como uma “Metodologia com ‘M’ maiúsculo” (p. 6)² dada a necessidade de constante renovação dos métodos de pesquisa visto a rapidez do desenvolvimento linguístico infantil e da mudança no *status* gramatical dentro de um determinado sistema.

Uma vez enfrentando esses dilemas, como também discutido por Perroni (1996), é imprescindível comentarmos sobre os fatores determinantes do que a comunidade científica descreve como “fato” – ou, nos dizeres de Possenti (1996, p. 195, destaques do autor), como o *dado dado* – a partir das escolas teóricas existentes dentro da Aquisição da Linguagem. Isto posto, a grosso modo, o pesquisador pode agir com base em uma postura empírica, desenvolvendo então, generalizações para se explicar uma ocorrência elegida como um “dado bruto”; ou ainda, fundamentando-se em uma abstração, que costuma ser resultado de um conjunto de hipóteses e/ou pressupostos, originados a partir da interpretação que se faz a respeito de um “fenômeno” (Perroni, 1996, p. 16).

De qualquer modo, como procuramos apontar ao longo dos parágrafos anteriores, a complexidade em se definir o dado ou um recorte de dados na pesquisa, reside, dentre outras coisas, no fato de que não há como desenvolver um estudo em que a metodologia esteja desvinculada da teoria e, portanto, onde as evidências coletadas – posteriormente, tomadas por “dado” – sejam completamente e, *a priori*, empíricas ou descoladas da própria

² No original: “There is little in the study of language acquisition that can be called Methodology, with a capital ‘M’. The very speed of linguistic development constrains the methods used in studying it. Massive changes in the grammatical status of children take place between one-and-a-half and three years” (McNeill, 1970, p. 6).

contextualização histórico-formativa de um campo do saber. Isso, segundo Maria Cecília Perroni (1996, p. 17):

Contraria também a suposição da possibilidade de acesso a um mundo dito bruto, conhecido diretamente com conseqüente obscurecimento do fato de que os critérios chamados ‘objetivos’ de identificação de ‘comportamentos’, ‘eventos’, ‘entidades’ têm sido altamente circunscritos pela cultura, história ou contexto social. [...] Nos estudos recentes de aquisição da linguagem começa a crescer o reconhecimento de que qualquer metodologia é determinada pela teoria eleita pelo investigador, assim como é a natureza da unidade de análise, que varia ao sabor da pesquisa dominante (Peters, 1983).

Nesta mesma linha de raciocínio, a autora nos adverte sobre as etapas nas quais a metodologia necessariamente se desdobra, a saber: **I)** a da coleta, ou do trabalho em campo; e **II)** a da análise, alegando que diferentes “filtros” agem como instrumentos de seleção. O primeiro atuaria sobre as próprias decisões do sujeito-pesquisador, sobretudo no que tange à triagem de indivíduos para a pesquisa e na forma de se obter o comportamento pretendido em um dado estudo; já o segundo, agiria na própria interpretação, relacionada e possibilitada, ainda que indiretamente, pela primeira atitude de *escolha*. Perroni (1996, p. 25), ao final do capítulo, conclui que a metodologia é a responsável por gerar o que as diferentes ciências existentes consideram como “dado” ao dizer:

Se as ciências têm-se encantado com o fato de que uma metodologia sólida gerará fatos sólidos, é preciso também reconhecer que a própria opção por uma metodologia é ditada pela teoria abraçada, com todas as suas crenças e pressupostos a respeito da natureza de seu objeto de estudo (1996, p. 25).

Após essas considerações iniciais sobre a natureza do dado e sua íntima relação com a postura metodológica de cada pesquisador, bem como com o campo geral onde determinado estudo se insere, passemos às especificidades teóricas e metodológicas da área de Aquisição e como esses aspectos incidem, ainda que indiretamente, na constituição de um banco de dados brasileiro, em especial no que é desenvolvido pelo grupo NALíngua-CNPq (Del Ré; Hilário; Rodrigues, 2016; Hilário; Moreira; Vieira, 2024).

4 O FAZER CIENTÍFICO EM AQUISIÇÃO: PRINCIPAIS TEORIAS E METODOLOGIAS

Neste tópico procuramos situar o leitor sobre as diferentes teorias que a Aquisição da Linguagem mobiliza na tentativa de descrever e explicar os variados fenômenos linguísticos presentes ao longo do processo de desenvolvimento da linguagem infantil e, igualmente, acerca das (possíveis) formas de tratamento do *corpus*. O faremos, principalmente, a partir das discussões de Del Ré (2023), no capítulo *Pluralismo teórico-metodológico em Aquisição da Linguagem: o lugar do sujeito e da linguagem nas pesquisas*, que integra o livro *Trabalhando com Linguística no Brasil*, organizado por Adair Vieira Gonçalves e Marcos Lúcio de Souza Góis (Gonçalves; Góis, 2023).

Como comentamos em outras partes desta monografia, o universo da Aquisição, de certa maneira, é envolto em controvérsias, essas que advém da interdisciplinaridade fundadora da área, mas, também, dos diálogos com novos e múltiplos campos – Neurociências, Fonoaudiologia, Educação, dentre outros – que, subsequentemente, com ela, foram estabelecidos. Logo, algumas dicotomias, originadas de tamanha diversidade de pontos de vista, obrigatoriamente rodeiam o fazer científico dos sujeitos-pesquisadores inseridos neste domínio quando pensamos na atitude de *escolha* metodológica, a exemplo: a de inato *versus* social, natureza *versus* cultura, coletas de dados qualitativos *versus* quantitativos, propostas longitudinais *versus* transversais, *etc.*

Tal variedade nada mais é do que uma consequência da realidade histórico-cultural em que viviam as Ciências, acima de tudo, a Psicologia e a Linguística em meados da década de 1950 no que tange, dentre outros aspectos, a duas posturas: a de fomento da interdisciplinaridade dentro da grande área das Ciências da Linguagem e o crescente interesse na relação linguagem-pensamento. Del Ré (2023, p. 428) relata que uma das origens atribuídas à Aquisição é a Psicolinguística, uma subárea da Linguística que tem seu início em um evento, sediado na Universidade de Indiana no ano de 1954, no qual um grupo de linguistas e psicólogos procurou unir as contribuições feitas até então pela Linguística Estrutural, Teoria da Aprendizagem e Teoria da Comunicação, na tentativa de descrever a produção (*output*) dos falantes.

Ainda segundo a autora (Del Ré, 2023, p. 430), um dos primeiros resultados dessa união, em uma postura empirista, foi o trabalho *Language behavior* (1957), de Skinner. O psicólogo se interessava pelo modo de aprendizagem das crianças, defendendo que um dado comportamento seria internalizado/adquirido a partir de seu reforço – positivo ou negativo. Se

positivo, a conduta se manteria nos indivíduos, o oposto ocorreria quando uma resposta negativa fosse obtida; é preciso salientar que, neste momento, o processo de aquisição não tinha um *status* claramente definido dentro da Linguística e, por isso, a fala infantil era concebida como um comportamento.

Del Ré (2023, p. 430) coloca a proposta de Chomsky, feita em *Syntactic Structures* (1957), como um marco para a área de Aquisição da Linguagem, uma vez que, a partir deste trabalho, temos o questionamento acerca da criatividade encontrada nas produções das crianças, que não poderia, a princípio, ser explicada pelo circuito estímulo-reforço (E-R) proposto pelo behaviorismo. Desse modo, Chomsky coloca em xeque o modelo associacionista de aprendizagem, através da proposição de estruturas gramaticais inatas, que seriam engatilhadas pelo Dispositivo de Aquisição da Linguagem (DAL ou LAD, *Language Acquisition Device*, em inglês) sendo que, a partir desse pressuposto, a língua não mais seria vista como um comportamento, mas um saber.

Para Noam Chomsky, a língua é um sistema biologicamente determinado e transmitido de forma genética, assim sendo, as crianças já teriam as competências-base (estruturas abstratas e complexas) de um sistema linguístico antes de adentrarem no idioma materno. No entanto, este conhecimento inato só seria ativado por meio do contato da criança com um outro-falante que a exporia, por meio de sentenças (*input*)³, às especificidades da gramática de sua língua materna, acionando, portanto, o DAL; sendo esta inserção em um ambiente (social, cultural, *etc.*) no qual existam usuários em *performance* o que garantiria o início do processo de aquisição da linguagem.

Nos anos consecutivos, a área passa por vários períodos teóricos que se propuseram a analisar a linguagem da criança por diferentes vias de entrada – sobre isto, em um primeiro momento, podemos pensar em dois grandes caminhos de análise: um mais mentalista, influenciado pelos trabalhos de Chomsky (1957, 1965), e outro de cunho interacional (Vygotsky, 1956; Bruner, 1975). Nos anos 1980, temos, então, a emergência da perspectiva interacionista, que se subdivide em cognitiva e social. Em um primeiro momento dessa perspectiva, considerava-se que as estruturas linguísticas seriam adquiridas juntamente com as propriedades semânticas e discursivas, sendo a aquisição de todas essas unidades da língua um resultado do próprio funcionamento dos princípios cognitivos.

³ O autor ainda descreve a existência de uma Gramática Universal (G.U.), ou seja, “regras gerais” (*princípios*) comuns a diferentes sistemas linguísticos que são atualizadas a partir do contato com o meio. Tais *princípios* seriam o fundamento para se estudar a organização das línguas e, igualmente, seus *parâmetros* (possibilidades de preenchimento em relação às “regras gerais”).

O interacionismo de caráter cognitivista tem como seu maior representante o psicólogo suíço Jean William Fritz Piaget (1896-1980) (Del Ré, 2023, p. 433); para ele, os sujeitos constroem estruturas a partir da experiência biológica, em um movimento de interação-reação, com o mundo físico. O desenvolvimento de tais estruturas, que ocorre entre a infância e a adolescência⁴, é essencial para que os indivíduos adquiram e coloquem em operação diferentes conhecimentos. Desse modo, Piaget se opõe aos posicionamentos puramente racionalistas e/ou empiristas, visto que sua proposta considera o desenvolvimento cognitivo como fruto da ação conjunta dos mecanismos biológicos e sociais (experiência física).

Já a corrente interacionista de natureza social ou, como mais conhecida, a perspectiva sociointeracionista, coloca como um fator principal para a aprendizagem, a mediação qualitativa durante as interações, tendo como um de seus grandes nomes Lev Semionovich Vygotsky (1896-1934), com a proposição da Psicologia histórico-cultural. Para o autor, os infantes são ativos em seu processo formativo já que o aprendizado se daria em interação, por intermédio da mediação de um outro, na experiência social e material com o mundo; essas vivências (conhecimentos) seriam, posteriormente, reelaboradas internamente de modo individual pelos sujeitos – o saber passaria, então, por dois “estágios”, na direção do inter-psíquico para o intra-psíquico.

Ainda na esteira dos trabalhos que se voltaram para a interação enquanto um elemento central para o desenvolvimento, temos os estudos de Jerome Seymour Bruner (1915-2016) (1975, 1983), que se centrou sobre a díade interlocutor-criança⁵. Para ele, o aprendizado é um processo ativo, sustentado no recuo entre as experiências passadas e atuais (formação de esquemas, ritualização de interações – a criação de *formatos*) que permitem a organização das situações comunicacionais, garantindo um engajamento mútuo na atividade compartilhada entre os interlocutores. Nesse sentido, tanto a natureza da atividade (ação), como a própria atenção ali gerida refletiria no desenvolvimento das estruturas linguísticas e, com o gradual domínio destas, dentro da interação com o adulto, a criança adquiriria a linguagem (Del Ré, 2023, p. 433).

Mais recentemente, no Brasil, para além dos estudos que adotam uma abordagem gerativista ou (socio)interacionista, temos a proposta de uma via de análise para as produções infantis de cunho enunciativo (Silva, 2009; Diedrich, 2015; Oliveira, 2023) e outra de caráter

⁴ Os estágios de desenvolvimento descritos por Piaget em sua Teoria Cognitiva, partem do nascimento dos bebês e se estendem até os 12 anos de vida.

⁵ Bruner investigou, sobretudo, a díade mãe-criança e o papel dos jogos e brincadeiras no desenvolvimento da linguagem.

dialógico-discursivo (Del Ré *et al.*, 2014a/b, 2020, 2021). A primeira, estabelece um diálogo com os trabalhos de Benveniste (1976, 1989) e postula que a criança se desenvolve na linguagem via ato de enunciação, na interação intersubjetiva, singular e discursiva com os locutores e o mundo (aspecto de referência), aprendendo no e pelo aparelho de enunciação a manejar os aspectos formais da língua de uma dada cultura.

Já a perspectiva dialógico-discursiva se fundamenta nos trabalhos do Círculo de Bakhtin (Bakhtin, 2016; Volóchinov, 2018) e em outros autores interacionistas (Bruner, 1975; Vygotsky, 2005), considerando que as crianças entram na língua(gem) no e pelo discurso, através da heterogeneidade dos encadeamentos de enunciados, que se concretizam nos gêneros discursivos durante as situações de interação. Por esse viés, conforme a criança entra na língua de sua comunidade, ela também adquire conhecimento de mundo; desse modo, é construída, gradativamente, a subjetividade do indivíduo, sempre elaborada a partir das vivências intersubjetivas e do contexto social – imediato e mais amplo. (Del Ré; Mendonça; De Paula, 2014a).

Assim, por meio desta breve contextualização histórico-teórica, observamos um amplo leque de possibilidades para abordar as informações consideradas como dado dentro do território da Aquisição. Essas alternativas, a partir de determinado olhar teórico, resultam na gênese – ou, ao menos, na delimitação inicial de procedimentos adequados para “X” finalidade – dos métodos investigativos.

Diante desse cenário, algumas perguntas podem surgir: como podemos gerir e nos aproximar dos dados? Existe o “melhor modo” para se obter respostas nos estudos em aquisição? Há, para além do arcabouço teórico, outras “posturas” de partida? Procuraremos responder a esses questionamentos apresentando quatro condutas “macro” utilizadas até o momento para se coletar e abordar as especificidades da fala infantil que, como já dissemos, se encaixam mais ou menos, a depender das “lentes” teóricas assumidas como fundamento em uma pesquisa.

Antes de discorrermos sobre tais procedimentos, é preciso ter em mente que, a princípio, não há um “melhor modo” ou, ainda, uma “maneira padrão” para nos aproximarmos dos dados. Há diferentes caminhos que, a depender da posição teórica do pesquisador, bem como de seus interesses em relação ao objeto eleito, se adequam melhor ou não às respostas que se almeja dar frente aos dados. Nesse sentido, Del Ré (2023, p. 436) comenta sobre o caráter paradoxal do fazer-aquisicionista:

Por essa razão, a questão metodológica é crucial, central em Aquisição da Linguagem, pois é a partir dela que a pesquisa se configurará. Eis um dos paradoxos da pesquisa em Aquisição: a impossibilidade de se iniciar uma pesquisa dessa natureza sem uma metodologia e, ao mesmo tempo, em virtude do caráter interdisciplinar dos estudos que envolvem tanto a Psicolinguística quanto a Aquisição de Linguagem e, no caso desta última, dependendo do ponto de vista do investigador com relação a essa fala infantil, a dificuldade de haver um consenso quanto ao estabelecimento de uma metodologia – definitiva – de investigação que dê conta de todos os estudos nesse âmbito (Del Ré, 2023, p. 436).

Passemos, então, à descrição das quatro abordagens “macro” em termos de coleta, investigação e tratamento dos dados infantis. Faremos um contraste entre elas comentando sobre as implicações de se adotar uma posição ou outra e, igualmente, a sua “aderência” (maior ou menor presença e coerência, a dependente do ponto de vista de partida) em relação às diferentes perspectivas teóricas em Aquisição da Linguagem.

4.1 Estudos longitudinais *versus* transversais

Os estudos longitudinais, geralmente, estão mais associados a uma proposta naturalística de observação, visto que os dados são gerados a partir do acompanhamento regular de um ou mais períodos do processo de aquisição de uma criança ou de um grupo de crianças dentro de um intervalo periódico pré-estabelecido. Assim sendo, os pesquisadores podem optar por fazer visitas a cada 15 dias ou 1 vez ao mês, realizando registros com extensão média entre 20 minutos e 1 hora, sendo que a duração mínima de coleta deve ser entre 10 e 12 meses⁶ (Del Ré, 2023, p. 439).

Esta conduta, justamente por ser configurada com base em um rastreamento longo e periódico, não se propõe a trabalhar com uma grande quantia de indivíduos. Objetiva-se, então, capturar o desenvolvimento gradual e singular das produções linguageiras eleitas; logo, não há um grande controle sobre as variáveis, ainda que algumas atividades planejadas ou que apresentem um maior grau de “gerência” (como a proposição de jogos e brincadeiras, leitura conjunta de histórias, *etc.*) possam ser desenvolvidas entre os participantes ao longo das sessões. Nesse sentido, não há uma preocupação em se criar situações herméticas de linguagem, mas sim de se testar a teoria a partir do comportamento espontâneo considerando a singularidade de cada indivíduo em seu processo.

⁶ Há casos em que o acompanhamento é feito durante diferentes fases do processo de aquisição e/ou aprendizagem, podendo estender-se por muitos anos. No caso do banco NALíngua-CNPq (Del Ré; Hilário; Rodrigues, 2016), algumas crianças foram gravadas desde seu nascimento até os sete anos de idade, por exemplo.

Por oposto, nos estudos transversais, almeja-se constatar padrões de comportamento ou de desenvolvimento linguístico por meio da aplicação de testes e/ou implementação de situações artificializadas de interação em um grande número de crianças. A maior parte dos estudos que optam por essa abordagem são experimentais, ou seja, são feitos a partir de tarefas onde as variáveis são altamente controladas pelo pesquisador-interventor, os dados são criados em um curto período de tempo, sendo que estas informações, normalmente, dão foco para a compreensão e processamento da linguagem pela criança.

Logo, os dados são relativamente controlados do ponto de vista que procura assegurar uma certa estabilidade ou, até mesmo, uma homogeneidade das amostragens, já que elas são configuradas de modo equilibrado: seleção de igual número de meninos e meninas, em diferentes idades (posteriormente, separadas em faixas etárias no momento de aplicação), sendo esses indivíduos inseridos em uma realidade socioeconômica símil. Esse tipo de abordagem visa utilizar a recorrência de um fenômeno em populações distintas como fator de robustez e, portanto, descarta-se todo e qualquer tipo de resposta não esperada pela proposta do estudo em questão.

4.2 Estudos qualitativos *versus* quantitativos

Em pesquisas de natureza qualitativa os sujeitos são observados em sua totalidade e singularidade, isto é, não se parte do pressuposto que, na base, todos os indivíduos partem de um mesmo ponto ou, ainda, que apenas um elemento é determinante ou incide de maneira mais ou menos ativa dentro do processo de aquisição da linguagem. Ao contrário, leva-se em consideração o contexto concreto no qual o sujeito-alvo se constitui e é constituído por; desse modo, assume-se que cada desenvolvimento está inserido em uma realidade dinâmica e por isso, o pesquisador está muito próximo dos dados, sendo orientado ao longo do processo de observação. Isso, portanto, caracteriza-se como um movimento exploratório, descritivo e indutivo frente aos dados (Del Ré, 2023, p. 441).

Já em estudos quantitativos, presume-se a existência de um ambiente e realidade estáveis, onde estabelecem-se tarefas/ações bem objetivas e controladas (Del Ré, 2023, p. 441). Por esta razão, qualquer informação que possa surgir ao longo da aplicação do experimento e esteja ligada com a particularidade de um contexto para além do que está sendo conduzido, é desconsiderada. Desse modo o pesquisador não se centra no processo, mas sim nos resultados verificando (ou não) suas hipóteses a partir de um olhar mais generalizante para os casos observados. Trata-se, dessa maneira, de uma postura hipotético-dedutiva, a

partir de etapas previamente estabelecidas e, por esta razão, o investigador não é orientado pelos acontecimentos espontâneos da interação.

5 A CONSTITUIÇÃO DE BANCOS DE DADOS DE FALA

A criação de Bancos de Dados, de modo geral, muito se deve a uma necessidade de compartilhamento e/ou agrupamento de informações – principalmente, para fins de comparação e de correlação entre diferentes grupos de indivíduos – dentro da comunidade científica. Atualmente, a centralização de *corpora* é feita de maneiras distintas, a depender da natureza da grande área, do tema e subtemas de pesquisa, e do grupo – ou grupos – responsável por fomentar e monitorar as informações reunidas em um determinado banco.

É válido destacar que a construção desses tipos de repositórios, em termos globais, já é uma prática bem difundida nas Ciências da Linguagem. No Brasil, essa tentativa de agrupamento, sobretudo de dados de fala, foi impulsionada pelos estudos em Sociolinguística, em especial, pelo projeto de estudo da Norma Linguística Urbana Culta, isto é, o Projeto NURC (Pessoa; Ramos, 2017). No entanto, quando pensamos, especificamente, na formação de bancos de fala infantil, arriscamo-nos dizer que há um jogo entre movimentos distintos no que tange à organização e difusão dos *corpora* coletados.

Na área de Aquisição da Linguagem podemos mencionar como uma construção mais específica de dados os *corpora* FDC (Fala Dirigida à Criança) e FI (Fala Infantil), que estão sob curadoria do Laboratório de Estudos em Aquisição da Linguagem (LEAL-USP/FFLCH)⁷, coordenado pelas Profas. Dras. Elaine Grolla e Raquel Santana Santos. Ambos os *corpora* são sistematizados no formato de lista de frequência de palavras que foram tratadas com o auxílio de *scripts* em Python, têm uma linguagem de anotação própria e são compartilhados mediante solicitação via formulário.

Um exemplo colaboração internacional é a plataforma CHILDES (*Child Language Data Exchange System*)⁸, um projeto idealizado por Brian MacWhinney (Carnegie Mellon University, Pittsburgh) e Catherine Snow (Harvard University, Cambridge). Esse repositório reúne mais de 30 línguas distintas e separadas, sempre que possível, em “famílias”; assim, o banco é organizado de acordo com os idiomas e subtemas que nele são alocados. Além de estabelecer normas de transcrição padronizadas, a saber, as CLAN/CHAT (MacWhinney, 2000), a plataforma também sugere o uso de um conjunto de ferramentas que

⁷ Acessível em: <https://leal.fflch.usp.br/Corpora-FI-FDC>. Outros corpora brasileiros no campo da Aquisição são o ALIP-Aquisição (Banco de Dados de Aquisição | ALIP - Amostra Linguística do Interior Paulista), o BATALE (BATALE | GEALE – Grupo de Estudos sobre Aquisição da Linguagem Escrita) e o Banco de Dados de Fala e Escrita (Banco de Dados de Fala e Escrita).

⁸ Disponível em: <https://chilides.talkbank.org/>.

parte desde a indicação de procedimentos/materiais de coleta até o índice de outros *softwares* úteis para a análise da fala infantil.

Assim, há, de um lado, grupos que propõe a coleta de informações mais nichadas/para fins específicos, sendo estas compartilhadas em maior ou menor escala e, portanto, sem a necessidade de uma revisão exaustiva do fazer-metodológico em relação ao tratamento e procedimentos de armazenamento dos dados. E, de outro, coletivos de pesquisa que centralizam diferentes tipos de dado, mas que devem, de tempos em tempos, rever sua prática de organização para que essas diferentes informações possam ser acessadas e utilizadas de maneira funcional pelos pesquisadores que delas queiram usufruir – principalmente no que diz respeito a *leitura* (legibilidade) dos dados disponibilizados.

5.1 A relação entre dado, autoria e a formação de Banco de Dados

Como indicamos anteriormente, dado e metodologia são dois aspectos indissociáveis, uma característica ainda mais determinante quando nos voltamos para a área de Aquisição. Mas, como esses elementos se relacionam e incidem, mesmo que indiretamente, na construção de um banco de dados? Em primeiro lugar, de uma forma mais geral, devemos ter clareza sobre *aquilo que se quer agrupar*: quais dados serão reunidos/coletados? Como essas informações, na base, se conectam ou estabelecem paralelos entre si? Há a pretensão de que um banco “X” também alimente outro “Y”? Qual é a forma mais coerente de apresentação das informações, principalmente quando temos informações de diferentes naturezas, mas que se centram dizem sobre um mesmo “objeto geral”?

É, portanto, a partir dessa identificação inicial que podemos mobilizar diferentes operações para a filtragem do “esqueleto” de um banco de dados em potencial. Em um momento posterior, é necessário refletir sobre o tema ou, melhor, acerca do “fio condutor” do repositório: o que encontraremos nesse espaço? Dados de fala, escrita ou ambos? Eles serão extraídos de uma região ou nacionalidade específica? O banco terá a possibilidade de receber (novas) informações de forma contínua? Quais critérios de padronização dos dados serão eleitos, considerando a natureza dos dados e a pretensão de alimentar outros conjuntos?

Por fim, mas não menos importante, devemos estar atentos às questões de autoria. Como bem salientado por Pessoa e Ramos (2017, p. 2179-2180) a construção, tratamento e disponibilização de um *corpus* é um trabalho grande e feito a muitas mãos. Por esse motivo, é de suma importância que o devido acabamento seja dado no detalhamento das informações: apresentar um título para o banco e instruir como referenciá-lo adequadamente; atribuir

crédito às diferentes etapas e colaborações – mencionar os nomes dos organizadores, transcritores, aplicadores e auxiliares de pesquisa; e, igualmente, descrever a dimensão do *corpus* e o perfil dos informantes-participantes, dessa forma:

Em resumo, dentre as várias etapas de formação de um *corpus* deve-se incluir a de divulgação do trabalho realizado. Outra etapa também importante é a de disponibilização irrestrita. Para a disponibilização, é importante ressaltar as exigências de exclusão de nome e de informações que poderão permitir a identificação do informante. Outras tarefas, que vão variar conforme o tipo de banco de dados pretendido, são a etiquetagem, os cabeçalhos, a descrição da situação de coleta, perfil do entrevistador, dentre outras. O reconhecimento da complexidade das tarefas envolvidas na organização de um *corpus*, justifica tratá-lo como um trabalho independente e não como uma etapa metodológica de um projeto (Pessoa; Ramos, 2017, p. 2180).

6 BANCO NALÍNGUA: A FORMAÇÃO DE UM *CORPUS* COMUM NACIONAL

O Banco NALíngua é fruto do conjunto de coletas realizadas pelos pesquisadores e grupos de pesquisa que fazem parte do Núcleo de Estudos em Aquisição da Linguagem (NALíngua-CNPq). Fundado em 2008, esse coletivo interdisciplinar é atualmente coordenado pelas professoras doutoras Alessandra Del Ré (UNESP-FCLAr) e Marianne Carvalho Bezerra Cavalcante (UFPB), reunindo linguistas, psicólogos, fonoaudiólogos e educadores do Brasil e da França⁹.

Isto posto, vale dizer que o elo que une esses profissionais-pesquisadores com práticas teórico-metodológicas tão heterogêneas, mas complementares em um mesmo coletivo de colaboração, é o interesse em apreender o processo de constituição da criança enquanto sujeito falante, desde as primeiras vocalizações, elementos prosódicos, gestos, até a entrada na escrita, por meio da investigação de recortes específicos dos mais variados fenômenos de linguagem (Del Ré; Hilário; Rodrigues, 2016; Hilário; Moreira; Vieira, 2024).

Ademais, o esforço desses pesquisadores em reunir tais dados é algo inédito e um pouco diferente do que fazem outros bancos que temos no Brasil (como os desenvolvidos em instituições como a UNICAMP, USP, UFPB, PUC, *etc.*). Isso se dá por algumas razões: primeiro, há poucos dados acessíveis aos pesquisadores da área de Aquisição da Linguagem, especialmente no que diz respeito ao Português do Brasil (PB) e às crianças bilíngues que o têm como uma de suas línguas (Del Ré; Hilário; Rodrigues, 2016).

Durante esta pesquisa, não foi encontrado nenhum outro banco de dados que acompanhasse uma mesma criança por um período tão extenso quanto o deste projeto, no qual a maioria das crianças é monitorada desde o nascimento até os sete anos de idade. Além disso, em geral, outros bancos de dados não estão organizados de acordo com padrões globais ou transcritos conforme as normas CLAN/CHAT (MacWhinney, 2000). Essa falta de padronização dificulta a realização de pesquisas colaborativas em nível internacional, como comparações com dados de fala infantil provenientes de outros países, especialmente os EUA e os países europeus (Del Ré; Hilário; Rodrigues, 2016).

Embora a coleta de dados seja feita a muitas mãos, há recomendações a serem seguidas para que a boa qualidade das informações seja garantida como: o uso de câmeras semiprofissionais e/ou *smartphones* equipados com lentes e microfones de alta resolução; a produção de registros que se mantenham na extensão média de 20 minutos a 1 hora; e a

⁹ Mais informações sobre os membros do grupo NALíngua-CNPq podem ser consultadas no Diretório de Grupos de Pesquisa no Brasil (Lattes/CNPq), que está disponível em: <http://dgp.cnpq.br/dgp/espelhogrupo/25793>.

utilização – quando necessário – microfones acopláveis multi-set. Essa padronização dos equipamentos e métodos de coleta, segue, então, as orientações da plataforma CHILDES (*Child Language Data Exchange System*) que, constantemente, são revisadas e atualizadas.

No momento, o Banco NALíngua é constituído pelos dados relacionados na tabela a seguir. No entanto, é válido pontuar que sua ampliação é contínua e que todas as informações a ele vinculadas seguem as prescrições, em termos éticos e legais, dos Documentos Internacionais e as Resoluções 466/12 e 510/16, assim como suas complementares, do Conselho Nacional de Saúde. Lê-se:

Tabela 1 – *Corpus* NALíngua

COLETAS DE MENOR DURAÇÃO			
MONOLINGUISMO	BILINGUISMO		DESENVOLVIMENTO ATÍPICO (TEA, SD) ¹⁰
- Me, 2;1 a 2;11 anos; - An., 1;1 a 2;9 anos.	- Mar., 2;5 a 3;2 anos (PB/FR); - C., 6;10 a 8;11 anos (PB/ES); - F., 5 a 6 anos (PB/ES); - L., 8 a 9 anos (PB/AL).		- G., 5 a 5;7 anos. (TEA) - E., 5 a 5;4 anos. (SD)
COLETAS DE MAIOR DURAÇÃO			
OUTRAS CRIANÇAS		GÊMEAS	
- Gus.; - So.; - M.; - Ma.; - Leo.; - O.; - E.. Todas registradas desde o nascimento até os 7 anos de idade.		DADOS PB	
		Lu. e Le., desde o nascimento até o presente.	
		DADOS ING	
		Lu. e Le., dos 4 anos até o presente.	
CONTEXTO ESCOLAR			
DESENVOLVIMENTO ATÍPICO (PARALISIA CEREBRAL)	DEBATE	BILINGUISMO	RECORTES DE FENÔMENOS ESPECÍFICOS

¹⁰ Nesta tabela, “TEA” (Transtorno do Espectro Austista) e “SD” (Síndrome de Down).

• S., 6;0 a 10;10 anos, produções escritas.	• Alunos do 9º E. F. ao longo de 1 ano (25 participantes).	• G., 2;10 a 7 anos (PB/ING).	• Alunos da 4ª e 5ª etapa do E. I., durante 6 meses totais, nos anos de 2023 e 2024, (20 participantes), para a observação de questões relativas à Argumentação e ao Humor.
---	--	---	---

Fonte: elaboração própria.

Apesar da proposta de criação do Banco NALíngua preconizar o agrupamento de dados de crianças monolíngues e bilíngues, geralmente pertencentes às classes média e média-alta, de modo longitudinal, preferencialmente com registros desde o nascimento até os 7 anos de idade e de cunho naturalístico, seja em contexto familiar e/ou escolar, é importante destacar que ele está aberto a receber dados que sejam produto de pesquisas feitas em outros contextos como por exemplo em clínicas terapêuticas, desde que esses registros estejam em consonância com os padrões de coleta¹¹ adotados pelo grupo.

Uma outra informação relevante sobre o tipo de *corpora* que o grupo vem reunindo é o fato de, mais recentemente, termos a oportunidade de obter dados do meio escolar, em escolas da rede pública e privada, em diferentes propostas de intervenção – seja para trabalhar um fenômeno específico de linguagem como a Argumentação e o Humor, por exemplo ou para acompanhar o desenvolvimento de crianças em língua estrangeira. Nesse sentido, se comparado à proposta de criação do Banco NALíngua, já vemos algum tipo de expansão e deslocamento da natureza de dados coletados e gerados.

6.1 As diferentes abordagens e ferramentas que permeiam o fazer-científico do NALíngua

Como já dissemos, a prática teórico-metodológica do grupo NALíngua-CNPq é muito diversa visto que ele reúne profissionais e pesquisadores que atuam em diferentes frentes interessadas no desenvolvimento da criança – psicólogos, educadores, fonoaudiólogos e linguistas. Entretanto, em linhas gerais, podemos dizer, a partir da descrição que fazemos

¹¹ Aqui, para além das questões éticas e legais, referimo-nos, principalmente, ao uso de equipamentos de gravação audiovisual adequados, que resultem em arquivos de boa qualidade.

abaixo, que boa parte deles filia-se às perspectivas dialógico-discursiva, multimodal e interacionista. Ademais, as outras óticas que também integram o grupo estabelecem uma relação harmônica com essas primeiras uma vez que, cada uma dentro de suas particularidades, considera a interação e o papel do *outro* como aspectos essenciais para o desenvolvimento – cognitivo, social, linguageiro e subjetivo – infantil.

Assim sendo, conforme as 25 respostas obtidas pelo formulário¹², temos:

1. 10 pesquisadores que utilizam a concepção dialógico-discursiva;
2. 7 membros que se embasam em uma abordagem multimodal;
3. 6 integrantes que fazem uso da perspectiva interacionista;
4. 5 que realizam seus estudos a partir da interface aquisição da linguagem e educação/aprendizagem;
5. 4 participantes que operam a partir do olhar enunciativo, configurando 16%;
6. 2 pesquisadores que propõe trabalhos a partir da interface aquisição da linguagem e cognição;
7. 2 pesquisadores que se filiam à teoria das operações enunciativas, de Antoine Culioli; e
8. 4 membros que se valem de “outras abordagens”, sendo um de cada e respectivamente: materialismo histórico dialético, perspectiva psicolinguística, perspectiva linguístico-discursiva e gramática construcional;

Ainda que tenhamos uma grande diversidade de pontos de partida e, portanto, resultados diferentes a termos de gerenciamento de dados/coleta, a maior parte destes profissionais trabalha com informações naturalísticas, longitudinais e qualitativas.

Quanto às ferramentas de análise, para além dos pontos de vista teóricos mobilizados por cada pesquisador, temos como mais frequentes os programas CLAN, ELAN, Excel (Grille) e PRAAT. Além disso, outros membros também fazem uso de *softwares* como o Face Reader®, Paint, editores de vídeo gratuitos ou, ainda, análise de conteúdo por meio da pura observação das filmagens ou demais tipos de registro coletados e anotação de glosas

¹² Em algumas caixas como nas de base teórica e ferramentas de análise utilizadas, mais de uma seleção pelos respondentes era possível visto que, a depender do tipo de pesquisa, esses olhares/estratégias de análise podem ser complementares entre si.

desses materiais, a partir de critérios da filiação teórica de base. Desse modo, listamos a seguir, uma descrição resumida de cada um desses métodos de análise:

6.1.1 CLAN

O CLAN (*Computerized Language Analysis*) faz parte do projeto CHILDES (*Child Language Data Exchange System*), iniciado nos anos 1980 por Brian MacWhinney e Catherine Snow. O projeto criou uma base de dados centralizada para o estudo da aquisição de linguagem por crianças em diversas línguas e contextos. Atualmente, a plataforma conta com cerca de 230 *corpora*, advindos de 30 idiomas do mundo todo; as transcrições desses dados seguem o padrão CHAT, o que facilita a análise e busca automatizada de elementos, além de garantir a uniformidade das anotações. O CLAN permite o alinhamento de áudio e vídeo às glosas feitas pelo transcritor, além de possibilitar a marcação de unidades linguísticas e contextuais. Essa ferramenta é utilizada por 44% dos membros do NALíngua.

6.1.2 ELAN

O ELAN, desenvolvido no início dos anos 2000 pelo Instituto Max Planck, é um *software* de anotação que organiza as anotações em camadas ou “trilhas”, permitindo a manipulação simultânea de transcrições e recursos audiovisuais. Diferentemente do CLAN, ele oferece uma estrutura multinível, facilitando seu uso em diversas áreas científicas como a Aquisição da Linguagem, Análise da Conversação e áreas da saúde. No campo da Aquisição, o ELAN tem sido utilizado para estudar a multimodalidade infantil, integrando gestos, movimentos corporais e prosódia. O programa também permite a interface com outras ferramentas de análise como o PHON e o PRAAT, centralizando diferentes dados e viabilizando a análise por diferentes “pontos de entrada”. Esse instrumento é usado por 44% dos integrantes do grupo NALíngua.

6.1.3 GRILLE

A *Grille* ou Grade de Categorias¹³, proposta por Salazar Orvig e colaboradores (Salazar Orvig; De Weck, 2013; Salazar Orvig *et al.*, 2021), é um método de análise baseado em uma tabela Excel ajustável conforme os interesses de uma pesquisa. Originalmente desenvolvida para estudos com base dialógico-discursiva, a Grade inclui classificações

¹³ Tradução nossa da palavra francesa *grille*.

relacionadas a conceitos como dialogia, gêneros e alteridade, podendo ser expandida para novos estudos, conforme as especificidades de cada fenômeno investigado. Esse método não só possibilita uma análise qualitativa detalhada, como também permite quantificar dados pelo cruzamento de colunas e linhas no Excel. Além disso, por estar ancorada em fundamentos teóricos dialógicos, promove uma interpretação mais rigorosa e alinhada a essa teoria. Esse método é utilizado por 24% do NALíngua.

6.1.4 PRAAT

O PRAAT foi desenvolvido por Paul Boersma e David Weenink, da Universidade de Amsterdã. Ele é amplamente utilizado em estudos de fonética, linguística e demais áreas relacionadas à fala e linguagem, sobretudo, por sua flexibilidade e capacidade de manipulação. Ele permite que os usuários realizem diversas tarefas de análise fonéticas como o estudo de frequência, duração, intensidade, *pitch* (altura tonal) e formantes (ressonâncias vocais). Para mais, neste *software* os pesquisadores são capazes de gerar e editar representações visuais de sons, como espectrogramas por exemplo, o que facilita a visualização de características acústicas de forma detalhada. Por fim, o PRAAT também oferece ferramentas para a segmentação de fala, de anotação e de criação de *scripts* para a automatização de análises mais complexas. Esse recurso é usado por 20% do NALíngua.

6.1.5 OUTRAS FERRAMENTAS COMPUTACIONAIS

Outros recursos usados pelos membros do grupo são o Face Reader®, Paint e editores de vídeo em geral, a soma desses correspondendo a 8%. O Face Reader é uma ferramenta que possibilita a análise emocional, usando algoritmos de reconhecimento facial para identificar expressões e emoções do rosto. Já o Paint é um programa atrelado ao sistema Microsoft Windows, muito utilizado para edições gráficas simples, sendo popular desde os anos 90 por permitir que os usuários criem e editem imagens de maneira intuitiva. Quanto às ferramentas de edição de vídeo, os pesquisadores utilizam aquelas disponibilizadas em seus próprios dispositivos (computadores/celulares) ou outros programas livres de *download* gratuito, com o objetivo de recortar trechos específicos e ou de legendagem.

6.2 Uma constante (re)elaboração: procedimentos de organização

Neste subitem, discorremos sobre os protocolos já adotados na organização do Banco NALíngua, os que estão em desenvolvimento e aqueles que serão implementados nas próximas etapas de ordenação. Logo, uma primeira atitude de padronização está ligada aos procedimentos de coleta e a durabilidade dos materiais recolhidos; nesse sentido, o grupo adota as recomendações internacionais¹⁴ que prescrevem o uso de câmeras de alta resolução (AVCHD) ou, então, de outros dispositivos portáteis de gravação áudio-vídeo, como iPhones ou iPads, com microfones e tripés acopláveis.

Considerando, igualmente, a preocupação do grupo NALíngua em, futuramente, estabelecer diálogo com outras bases, sobretudo as internacionais, os dados, tanto em sua nomeação, como em termos de transcrição, seguem as normas CHAT do programa CLAN (MacWhinney, 2000)¹⁵. Desse modo, a etiquetagem dos materiais acontece da seguinte forma: SIGLA-DA-CRIANÇAanos_meses_dias, exemplo: B2_6_25 (lê-se, B. dois anos, seis meses e vinte e cinco dias); assim sendo, tanto a folha de transcrição (arquivo .cha), como o vídeo (arquivo em formato MOV ou MP4) são nomeados similarmente, possibilitando o alinhamento do documento transcrito com o registro audiovisual.

Vale dizer que as diretrizes CHAT (*Codes for the Human Analysis of Transcripts*), funcionam como uma “língua franca”, no sentido de que elas são um conjunto de códigos que estão integrados à sintaxe do programa, permitindo tanto a leitura de transcrições por falantes de diferentes línguas (as regras de anotação são as mesmas, independentemente do idioma de origem), como a realização de buscas específicas dentro de um *corpus*, e a análise automática de alguns elementos de língua (frequências lexicais, morfemas, fonemas, entre outros). Além disso, o CHAT não se restringe à anotação grafêmica do enunciado (linha principal ou ortográfica), mas possibilita a transcrição de outras partes integrantes da cena enunciativa, nas chamadas linhas adicionais¹⁶.

No presente, os dados estão sendo gradualmente inseridos em um conjunto de *softwares* e *hardware* alocados em um ambiente Cliente-Servidor (Del Ré; Hilário; Rodrigues,

¹⁴ Disponível em: <https://talkbank.org/info/dv.html>.

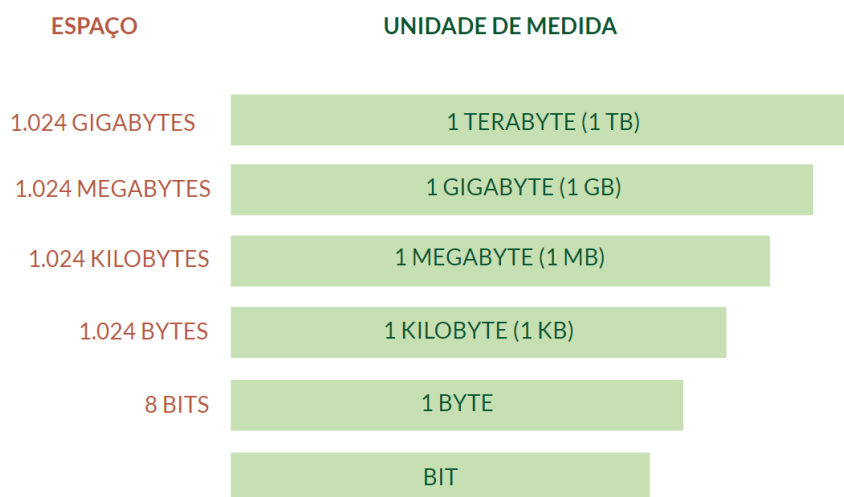
¹⁵ É importante registrar que alguns dos dados, devido às especificidades de pesquisa, também estão transcritos a partir de outros programas ou convenções, sendo que essas anotações estarão, igualmente, anexadas para consulta.

¹⁶ A saber: linhas **%pho**, de transcrição fonética; **%act**, descrição de ações, gestos e/ou atividade que acompanha o enunciado; **%sit**, descrição de uma situação ou atos gerais; **%com**, linha para comentários do transcrito (de cunho metalinguístico, explicações sobre o enunciado produzido, etc.); **%add**, linha para elucidar a quem o falante se dirige, nos casos de ambiguidade; e **%spa**, categorização dos elementos que constituem o enunciado.

2016), que está sendo reestruturado infraestruturalmente por um técnico de informática. Essa plataforma, no momento, já oferece as funções de gerenciamento (adição, remoção, alteração) de dados, embora a funcionalidade de compartilhamento ainda esteja operando de forma restrita. É importante destacar que esse rearranjo infraestrutural está sendo feito de acordo com as informações contidas nos arquivos-índice (descritos nos parágrafos a seguir) principalmente no que tange o desenvolvimento de seções e abas dentro do ambiente virtual.

Isto posto, atualmente, o conjunto de dados do Banco NALíngua, corresponde a, aproximadamente, 5TB (cinco *terabytes*), considerando, somente, as mídias digitais. Este Banco também conta com um acervo físico, correspondente a fitas Mini DVs, CDs, HDs externos e disquetes, sendo que, independentemente de serem mídias originalmente digitais ou físicas, todos os arquivos possuem uma cópia de segurança (*backups*), para que se evite ao máximo eventuais perdas. Para ilustrarmos de forma mais palpável a quantidade de dados sob curadoria do Banco NALíngua, apresentamos o esquema abaixo, que representa o volume e armazenamento, na unidade *bytes*:

Figura 1 – Armazenamento em bytes



Fonte: elaboração própria.

Além dessa primeira fase de triagem (etiquetagem de arquivos e realização de transcrições segundo as normas CLAN/CHAT, que é realizada progressivamente e de forma continuada a cada nova coleta), mais recentemente, desenvolvemos uma ficha-índice, a qual reúne informações essenciais sobre cada coleta realizada, como a identificação dos participantes, o número de sessões, objetivo de coleta, *etc.* – a ficha pode ser consultada na seção “Apêndices”. Para mais, é relevante dizer que este documento será utilizado para a

repaginação do servidor que abriga os *corpora* digitais do Banco NALíngua, pois a partir das respostas nele contidas, desenharemos as “caixinhas” que reunirão as informações ali alocadas em diferentes grupos, a depender das características de cada coleta.

Nesse sentido, tal ficha serve como um documento identificador e, ao mesmo tempo, como um molde. Conforme discutido ao longo das *seções 5 e 5.1*, diferentes questões e interesses estão imbricados na formação e manutenção de um banco de dados, incluindo a própria natureza e/ou finalidade dos dados, além do fator legibilidade. Por essa razão, ademais a uniformização das maneiras de nomeação e transcrição, é necessário que essa segunda atitude de triagem seja feita e que ela nos auxilie na visualização concreta das informações recolhidas pelos pesquisadores-interventores, igualmente, permitindo o esboço de “categorias-mães” de agrupamento.

O documento índice não só nos dá pistas quando pensamos no manejo e arranjo dos dados, mas também pode servir como um registro dos recursos – financeiros e humanos – mobilizados no ato de recolha e no pós-coleta. Isso porque a ficha-índice conta com um arquivo complementar, que poderá ser consultado pelos usuários de cada um dos dados coletados, contendo informações sobre as transcrições e mídias já disponíveis, bem como o local de armazenamento desses registros. Nesta ficha complementar¹⁷ também constam os recursos humanos envolvidos no recolhimento e tratamento dos dados, como os transcritores, revisores, interventores e auxiliares de coleta, por exemplo.

Um terceiro movimento de triagem, ainda em desenvolvimento e execução, é a criação de um sistema de catálogo, seguindo estratégias de arquivística e consulta a materiais, semelhante ao funcionamento de uma biblioteca, para os dados que se encontram nos dispositivos físicos que mencionamos anteriormente – os Mini DVs, CDs, HDs externos e disquetes. Como este tipo de material é de menor quantidade, se comparado com os dados digitais, estamos reunindo-os nas dependências do Núcleo de Pesquisa da Faculdade de Ciências e Letras de Araraquara, especificamente, na sala com uso destinado ao Grupo de Estudos em Aquisição da Linguagem (GEALin/UNESP-FCLAr), um grupo de pesquisa constituído por pesquisadores da UNESP, que também integram o grupo NALíngua-CNPq¹⁸.

Outro procedimento recentemente feito, foi a atualização dos termos de compartilhamento e utilização dos dados sob curadoria do Banco NALíngua. Como dissemos anteriormente, no momento, o servidor onde parte dos dados estão armazenados está passando

¹⁷ Também disponível em *Apêndices* – “Índice de Transcrições e Mídias”.

¹⁸ Para mais informações sobre os grupos GEALin (UNESP-FCLAr) e NALíngua (CNPq) consultar Hilário, Moreira e Vieira (2024) disponível em: <https://revista.abralin.org/index.php/abralin/article/view/2217>.

por uma reorganização estrutural a fim de que essas informações se tornem mais legíveis e possam ser compartilhadas em um grau mais amplo; por este motivo, o grupo optou por rever as condições de acesso e de ciência em relação ao uso e difusão dos dados na realização de pesquisas, artigos ou demais produtos científicos. Tanto a *Autorização de uso de dados* como o *Termo de compromisso de utilização de dados*, em suas novas versões, podem ser consultados em “Apêndices”.

Ademais, é válido dizer que, durante esta etapa de reformulação, o acesso aos dados está sendo feito por meio de uma chave-portal, com usuário e senha específicos, para os pesquisadores que desejam utilizar essas informações. A concessão de acesso, a partir de agora, só é feita após o preenchimento e assinatura do *Termo de compromisso* supracitado, assim como do recebimento da *Autorização de uso de dados*. Esses documentos reafirmam as normas prescritas pelos Documentos Internacionais e as Resoluções 466/12 e 510/16, assim como suas complementares do Conselho Nacional de Saúde, e devem ser assinados digitalmente por meio do *gov.br* ou através de outra assinatura eletrônica que possa ter sua autenticidade verificada.

7 CONSIDERAÇÕES FINAIS

Neste texto buscamos revisitar as questões que permeiam a relação entre dado-metodologia e a formação de bancos de dados dentro da área de Aquisição da Linguagem. Nosso intuito foi relacionar essas três instâncias comentando, principalmente, como essas questões teórico-metodológicas incidem na coleta de dados e em que medida a própria natureza dos dados, no caso de repositórios como o Banco NALíngua, nos quais há uma grande heterogeneidade de perspectivas, pode se tornar um critério para a criação de filtros e categorias de manutenção arquivística – seja física e/ou em ambiente virtual – desses conjuntos de informações.

Muito embora as práticas de constituição de bancos de dados nas áreas das Ciências da Linguagem não seja exatamente recente, pelas razões discutidas ao longo deste texto, assim como Pessoa e Ramos (2017), nos posicionamos a favor do reconhecimento da organização de um *corpora* não como mais uma etapa metodológica dentro das pesquisas, mas sim como um trabalho independente, contínuo, exaustivo e, sobretudo, produzido, desde as fases iniciais de coleta até a padronização deste material, por muitas mãos. Nesse sentido, o presente trabalho procurou oferecer um panorama sobre as especificidades da constituição de um banco de dados infantil, considerando as particularidades do grupo NALíngua-CNPq.

Dito isto, este trabalho procurou, da mesma forma, registrar as práticas adotadas por este conjunto de pesquisadores na coleta e tratamento dos dados, considerando uma realidade tão rica e diversa. Assim sendo, procuramos oferecer um esboço de práticas que podem ser endossadas ou adaptadas em *corpus* já formados e daqueles ainda em formação, em particular, dos que se centram sobre o agrupamento de informações acerca da fala infantil e os variados fenômenos encontrados dentro deste grande processo que é a entrada na linguagem, sejam elas de cunho naturalístico, experimental, longitudinal, transversal, *etc.*

Mesmo que esse seja um assunto para além do escopo desta monografia, é imprescindível reconhecermos a importância da tecnologia, seja no tratamento ou no compartilhamento de dados e os ganhos que ela nos dá dentro da realidade prática da comunidade científica. É justamente pelo desenvolvimento e aprimoramento de recursos de filmagem, áudio, *softwares* de análise, de sistemas em nuvem dentre outros que, hoje, podemos contar com uma sistematização e partilha de informações cada vez mais precisas, seguras e ágeis. Nessa perspectiva, devemos, ao mesmo tempo, estar abertos para as trocas que a tecnologia nos possibilita e atentos para o façamos de maneira ética e ordenada.

Na realidade de um banco de dados, além do cuidado com o tratamento, disposição e compartilhamento das informações nele contidas, visto que lidamos com materiais relacionados a pessoas – e, no caso específico do Banco NALíngua, a crianças –, é essencial divulgar os esforços de manutenção e constituição dessas informações. Portanto, é necessário valorizar a autoria, mencionando, sempre que possível e de acordo com as diretrizes éticas e de privacidade aplicáveis, todos os recursos, sejam eles humanos ou financeiros, envolvidos nas diferentes etapas de elaboração de um *corpus* ou *corpora*.

Finalmente, reconhecemos as limitações deste trabalho, especialmente quanto à necessidade de um maior detalhamento de cada uma das etapas já realizadas, em relação à sistematização do Banco NALíngua. Aqui não o fazemos, devido ao espaço reduzido de discussão e tendo em conta que algumas dessas fases ainda estão em momento de concepção e/ou aplicação inicial. Por esta razão, nos centramos em contextualizar a configuração mais atual do NALíngua, incluindo informações acerca de sua prática teórico-metodológica, bem como comentando sobre os procedimentos que, aos poucos, vêm sendo implementados na organização de seus *corpora*.

Assim sendo, esperamos ter contribuído de alguma forma para a sugestão de parâmetros iniciais para a reunião e arranjo de dados da linguagem infantil, ao apresentar os documentos que elaboramos durante o processo de triagem das informações e que têm finalidade arquivística. Em breve, esperamos poder, em outras produções, fornecer maiores detalhes sobre cada uma das etapas de manutenção e gerência de dados, no contexto do Banco NALíngua, almejando que estes registros possam auxiliar outros pesquisadores – do Brasil ou de outras localidades – a (re)pensarem a organização e compartilhamento de seus dados.

REFERÊNCIAS

- BENVENISTE, É. **Problemas de Linguística Geral**. Trad. Maria da Glória Novak e Luiza Neri. São Paulo: Editora Nacional/Editora da Universidade de São Paulo, 1976.
- BENVENISTE, É. **Problemas de linguística geral II**. Trad. de Eduardo Guimarães [et al]. Campinas: Pontes, 1989.
- BEZERRA, J. T G. de M.; SILVA, P. M. S. da.; CAVALCANTE, M. C. B. Softwares de transcrição como auxílio para as pesquisas com enfoque multimodal no processo de aquisição da linguagem. **Texto Livre: Linguagem e Tecnologia**, Belo Horizonte, MG, v. 9, n. 1, p. 77-93, 2016. Disponível em: <https://periodicos.ufmg.br/index.php/textolivre/article/view/16715>. Acesso em: 29 mai. 2024.
- BRUNER, J. S. From communication to language: a psychological perspective. **Cognition**, [S. l.], v. 3, n. 3, p. 255-287, 1975. Disponível em: <https://psycnet.apa.org/record/1976-27626-001>. Acesso em: 05 mar. 2024.
- BRUNER, J. S. **Como as crianças aprendem a falar**. Trad. de Joana Chaves. Lisboa: Horizontes Pedagógicos, 1983.
- CHOMSKY, N. **Syntactic Structures**. Nova Iorque, Berlim: Mouton & Co., 1957.
- CHOMSKY, N. **Aspects of the Theory of Syntax**. Cambridge: MIT Press, 1965. Disponível em: <https://www.jstor.org/stable/j.ctt17kk81z>. Acesso em: 05 mar. 2024.
- DIEDRICH, M. S. **Aquisição da Linguagem: o aspecto vocal da enunciação na experiência da criança na linguagem**. Tese (Doutorado em Letras) – Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, p. 147. 2015.
- DEL RÉ, A.; DE PAULA, L.; MENDONÇA, M. C. **A linguagem da criança: um olhar bakhtiniano**. São Paulo: Editora Contexto, 2014a.
- DEL RÉ, A.; DE PAULA, L.; MENDONÇA, M. C. **Explorando o Discurso da Criança**. São Paulo: Editora Contexto, 2014b.
- DEL RÉ, A. Pluralismo teórico-metodológico em aquisição da linguagem: o lugar do sujeito e da linguagem nas pesquisas. *In*: _____. **GONÇALVES, A. V.; GÓIS, M. L. de S. (Orgs.). Trabalhando com Linguística no Brasil**. Campinas: Pontes Editores, 2023, p. 427-456.
- DEL RÉ, A.; HILÁRIO, R. N.; RODRIGUES, R. A. O Corpus NALíngua e as tecnologias de apoio: a constituição de um banco de dados de fala de crianças no Brasil. **Revista Artefactum – Revista de Estudos em Linguagem e Tecnologia**. ANO VIII – Nº 02/2016. Disponível em: <http://artefactum.rafrom.com.br/index.php/artefactum/article/view/1363/667>. Acesso em: 19 dez. 2023.
- DEL RÉ, A.; HILÁRIO, R. N.; VIEIRA, A. J. A linguagem da criança na concepção dialógico-discursiva: retrospectiva e desafios teórico-metodológicos para o campo de Aquisição da Linguagem. **Bakhtiniana: Revista de Estudos do Discurso**, [S. l.], v. 16, n. 1,

jan./mar., p. 12-38, 2021. Disponível em:
<https://www.scielo.br/j/bak/a/dRS98pVJT4mJdmcc7JvkjyB/>. Acesso em 06 mai. 2022.

DOMINGUES, S. K. **As tecnologias e a Aquisição da Linguagem: metodologias de pesquisa e produção de dados em estudos em Aquisição da Linguagem**. Monografia de Conclusão de Curso (Bacharel em Letras) – Faculdade de Ciências e Letras, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Araraquara, SP, p. 49. 2022.

FRANÇOIS, F. **Morale et mise en mots**. Paris: L’Harmattan, 1994.

FRANÇOIS, F. O que nos indica a “linguagem da criança”: algumas considerações sobre a “linguagem”. Tradução de Guacira Marcondes Machado Leite. *In: _____*. **DEL RÉ, A. Aquisição da linguagem: uma abordagem psicolinguística**. São Paulo: Contexto, 2006.

FLORES, V. N.; CARDOSO, J. L. Transcrição e Enunciação. **Caderno de Estudos Linguísticos**, Campinas, SP, v. 64, [S. d.], p. 1-9, 2022. Disponível em:
<https://lume.ufrgs.br/bitstream/handle/10183/264457/001175301.pdf?sequence=1&isAllowed=y>. Acesso em: 05 mar. 2024.

HILÁRIO, R. N.; MOREIRA, F. M.; VIEIRA, A. J. Os grupos NALíngua (CNPq) e GEALin (UNESP/FCL-AR) e as pesquisas em Aquisição da Linguagem no Sul Global: um relato de experiências. **Revista da ABRALIN**, [S. l.], v. 23, n. 2, p. 599–623, 2024. Disponível em:
<https://revista.abralin.org/index.php/abralin/article/view/2217>. Acesso em: 16 set. 2024.

MACWHINNEY, B. **THE CHILDES PROJECT: Tools for Analyzing talk**. Volume II: The Database. Mahwah: Laurence Erlbaum Associates, 2000.

OLIVEIRA, G. F. **Do homo loquens ao homo loquens scriptor: por uma perspectiva semiológico-enunciativa da aquisição da escrita**. Tese (Doutorado em Letras) – Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, p. 428. 2023.

PERRONI, M. C. O que é o dado em aquisição da linguagem?. *In: _____*. **CASTRO, M. F. P. de. (Org.). O método e o dado no estudo da linguagem**. Campinas: Editora da UNICAMP, 1996, p. 15-29.

PESSOA, F. da C. R.; RAMOS, J. M. A formação de um banco de dados de fala de Teresina (PI): um estudo de caso. **Fórum linguístico**, Florianópolis, SC, v. 14, n. 2, abr./jul., p. 2173-2183, 2017. Disponível em:
<https://periodicos.ufsc.br/index.php/forum/article/view/1984-8412.2017v14n2p2173>. 05 mar. 2024.

POSSENTI, S. O *dado* dado e o dado **dado**: o dado em análise do discurso. *In: _____*. **CASTRO, M. F. P. de. (Org.). O método e o dado no estudo da linguagem**. Campinas: Editora da UNICAMP, 1996, p. 195-207.

SALAZAR ORVIG, A. *et al.* Dialogical factors in toddlers’ use of clitic pronouns. **First language**, [S. l.], v. 30, n. 3, p. 375-402, 2010a. Disponível em:
https://www.researchgate.net/publication/254092428_Dialogical_factors_in_toddlers'_use_of_clitic_pronouns. Acesso em: 19 dez. 2023.

SALAZAR ORVIG, A. *et al.* Dialogical beginnings of anaphora: the use of third person pronouns before the age of 3. **Journal of Pragmatics**, [S. l.], v. 42, n. 7, p. 1842-1865, 2010b. Disponível em:

https://www.researchgate.net/publication/248452230_Dialogical_beginnings_of_anaphora_The_use_of_third_person_pronouns_before_the_age_of_3. Acesso em: 19 dez. 2023.

SALAZAR ORVIG, A.; DE WECK, G. P. Profils de mères et implications des enfants dans la co-construction de récits. **ANAE**, [S. l.], 124(5), p. 269-278, 2013.

SALAZAR ORVIG, A.; DE WECK, G.; HASSAN, R.; RIALLAND, A. (Orgs.). **The Acquisition of Referring Expressions: a Dialogical Approach**. Amsterdã: John Benjamins, 2021.

SANTOS, R. S.; TONI, A. Corpus FI: banco de dados. Disponível em:

<https://leal.fflch.usp.br/Corpora-FI-FDC>. Acesso em 11 ago. 2024.

SANTOS, R. S.; TONI, A. Corpus FDC: banco de dados. Disponível em:

<https://leal.fflch.usp.br/Corpora-FI-FDC>. Acesso em 11 ago. 2024.

SILVA, C. L. da C. **A criança na linguagem: enunciação e aquisição**. Campinas: Pontes, 2009.

SCLIAR-CABRAL, L. Desafios metodológicos à Linguística de Corpus em aquisição da linguagem no português brasileiro. **Revista da Anpoll**, [S. l.], v. 52, n. 1, p. 205–219, 2021. Disponível em: <https://revistadaanpoll.emnuvens.com.br/revista/article/view/1499>. Acesso em: 05 mar. 2024.

SKINNER, B. F. **Verbal behavior**. Nova Iorque: Appleton Century Crofts, 1957.

WITTENBURG, P. *et al.* ELAN: a Professional Framework for Multimodality Research. **In: _____**. **Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation**, 2006.

APÊNDICE A: QUESTIONÁRIO (*GOOGLE FORMS*)

MAPEAMENTO: GRUPO NALíngua-CNPq
Olá! Muito obrigada por dedicar um pouco do seu tempo para responder às perguntas deste formulário. Elas vão ajudar na escrita da minha monografia que, em linhas gerais, aborda as metodologias em aquisição, dando foco para a criação de bancos de dados (com destaque para o NALíngua). OBS: apenas membros ainda atuantes/cadastrados no Grupo NALíngua-CNPq devem responder ao formulário.
1 IDENTIFICAÇÃO DO PESQUISADOR(A)
A) Nome completo. B) E-mail principal. C) Titulação e função do membro/a (mais de uma caixa de seleção possível. De preferência, selecione a titulação mais alta já obtida e a função mais atual). ☐ Professor/a titular ☐ Professor/a adjunto/a ☐ Pós-doutorando/a ☐ Doutor/a ☐ Doutorando/a ☐ Mestre/a ☐ Mestrando/a ☐ Outros: ____
2 IDENTIFICAÇÃO DA PRÁTICA TEÓRICO-METODOLÓGICA
A) Perspectiva teórica assumida (mais de uma seleção possível). ☐ Perspectiva dialógico-discursiva ☐ Perspectiva enunciativa ☐ Perspectiva interacionista ☐ Perspectiva semiológica ☐ Perspectiva multimodal ☐ Interface aquisição da linguagem e cognição ☐ Interface aquisição da linguagem e educação/aprendizagem ☐ Outros: ____ B) Natureza da análise em suas pesquisas (mais de uma seleção possível). ☐ Qualitativa ☐ Quantitativa ☐ Qualiquantitativa ☐ Transversal ☐ Longitudinal ☐ Naturalística ☐ Experimental ☐ Outros: ____ C) Ferramentas de análise que utiliza em suas pesquisas (mais de uma seleção possível). ☐ CLAN ☐ ELAN ☐ Grille ☐ PRAAT ☐ PHON ☐ Outros: ____

D) Caso tenha selecionado a opção <u>“outro”</u>, por favor, escreva o nome da ferramenta e dê uma breve descrição de seu funcionamento.
3 MAPEAMENTO DE DADOS
A) Dê uma breve descrição dos dados utilizados em seu(s) estudo(s). Você realizou coletas? Em que contextos? Esses dados pretendem constituir o Banco NALíngua? B) Em resposta afirmativa à pergunta anterior, especificamente no caso em que haja a pretensão de se disponibilizar os dados via banco NALíngua, preencha o anexo abaixo. Índice de Coleta - NALíngua
4 COMENTÁRIOS
A) Outras informações que você gostaria de destacar e que tenham a ver com a estruturação do Banco de Dados NALíngua (a termos metodológicos, de constituição, etc.).

APÊNDICE B: ÍNDICE DE COLETA

DESCRIÇÃO DA COLETA			
Nome(s) da(s) criança(s):	NOME(S) COMPLETO(S) DA(S) CRIANÇA(S) ALVO		
Data(s) de nascimento:	DIA/MÊS/ANO	Período de coleta:	De DIA/MÊS/ANO a DIA/MÊS/ANO
Foco(s)/objetivo(s) da coleta:	Descrição da proposta de coleta (qual fenômeno focalizado, em que contexto, em quantas crianças, que respostas buscou obter com as gravações, etc.).		
Inscrição no Comitê de Ética (ou outro órgão de avaliação):	☐ Sim ☐ Não N° de inscrição: XX/XXXX		
A coleta foi realizada com a verba de algum edital?	☐ Sim ☐ Não N° do edital: XX/XXXX Órgão de fomento: CAPES, IES, CNPq, etc.		
Natureza da coleta:	☐ Experimental ☐ Naturalística ☐ Quantitativa ☐ Qualitativa ☐ Transversal ☐ Longitudinal		
Contexto	☐ Contexto escolar/ de aprendizagem formal ☐ Familiar ☐ Outro contexto: _____ Descrição adicional do contexto.		
DESCRIÇÃO DAS SESSÕES			
Número total de sessões:	X sessões	Extensão média das sessões:	X minutos / horas
Datas das coletas e participantes:	DIA/MÊS/ANO - SIGLA: NOME DO PARTICIPANTE CONFORME NORMAS CLAN/CHAT EXEMPLO: 31/07/2024 - CHI: Maria Clara OBS: Ana Laura MOT: Karen FAT: Guilherme SIS: Luísa		
Observações:	Descrição de quaisquer outras informações relevantes sobre o andamento/realização da coleta.		

APÊNDICE C: ÍNDICE DE TRANSCRIÇÕES E MÍDIAS

DESCRIÇÃO			
Nome(s) completo(s) e sigla(s) da(s) criança(s)	José da Silva Santos Jos0_5_12 a Jos6_9_25		
Sessão	Transcrição	Mídia	
Jos0_5_12	☐ Sim ☐ Não ☐ CHAT/CLAN ☐ Outra: _____	☐ Digital ☐ Física: _____	
Jos ...	☐ Sim ☐ Não ☐ CHAT/CLAN ☐ Outra: _____	☐ Digital ☐ Física: _____	
LOCAL			
TRANSCRIÇÕES	MÍDIAS		
Jos0_5_12: (local de armazenamento)	DIGITAL		
	Jos0_5_12: (local de armazenamento)		
	FÍSICA		
	Jos0_5_12: (local de armazenamento e código de catalogação)		
TRANSCRITORES E REVISORES			
Jos0_5_12: nomes dos transcritores e revisores			
INTERVENTORES E AUXILIARES			
Jos0_5_12: nomes dos interventores e auxiliares			
STATUS			
Sessão	Jos0_5_12	Revisão	☐ Sim ☐ Não Data da revisão mais recente: _____

APÊNDICE D: AUTORIZAÇÃO DE USO DE DADOS



AUTORIZAÇÃO DE USO DE DADOS

Autorizamos, **FULANO DE TAL**, portador(a) das células de identidade **RG: XXX** e **CPF: XXX**, a fazer uso do *corpora* do NALíngua, composto, majoritariamente, por crianças filmadas longitudinalmente desde o nascimento até os 07 anos de idade, em situações naturalísticas. Informamos tratar-se de um *corpora* público e que estes dados são regidos pela resolução nº 510 do Conselho Nacional de Saúde, de 07/04/2016 (parágrafo único, inciso V). A referência ao *corpus* NALíngua-CNPq deve ser feita por meio da citação dos artigos:

- DEL RÉ, A.; HILÁRIO, R. N.; RODRIGUES, R. A. O Corpus NALíngua e as tecnologias de apoio: a constituição de um banco de dados de fala de crianças no Brasil. **Revista Artefactum: Revista de Estudos em Linguagem e Tecnologia**. ANO VIII – Nº 02/2016.
- HILÁRIO, R. N.; MOREIRA, F. M.; VIEIRA, A. J. Os grupos NALíngua (CNPq) e GEALin (UNESP/FCL-AR) e as pesquisas em Aquisição da Linguagem no Sul Global: um relato de experiências. **Revista da ABRALIN**, [S. l.], v. 23, n. 2, p. 599–623, 2024.

Local, dia de mês de ano.

Profª. Drª. Alessandra Del Ré
Coordenadora do Grupo NALíngua (CNPq)
Docente da UNESP-FCLAr

Profª. Drª. Marianne Cavalcante
Vice-Coordenadora do Grupo NALíngua (CNPq)
Docente da UFPB

APÊNDICE E: TERMO DE COMPROMISSO DE UTILIZAÇÃO DE DADOS



TERMO DE COMPROMISSO DE UTILIZAÇÃO DE DADOS

Eu, **FULANO DE TAL**, aluno(a)/pesquisador(a)/professor(a) da **INSTITUIÇÃO TAL**, do curso de **XXX**, do departamento de **XXX**, portador(a) das células de identidade **RG: XXX** e **CPF: XXX**, no âmbito do projeto de pesquisa intitulado “_____”, comprometo-me com a utilização dos dados contidos no Banco NALíngua, a fim da obtenção dos objetivos previstos no projeto supracitado, mantendo a confidencialidade destes dados, bem como a privacidade de seus conteúdos, como preconizam os Documentos Internacionais, as Resoluções 466/12 e 510/16, e suas complementares do Conselho Nacional de Saúde.

Esclareço, ainda, que os dados a serem utilizados se referem aos de **SIGLA DA CRIANÇA**, do período de **X** a **Y** anos.

Declaro entender que concessão de acesso/utilização é restrita somente aos dados firmados neste termo, sendo minha responsabilidade zelar pela integridade destas informações e, igualmente:

- I)** Manter a confidencialidade dos dados utilizados, bem como a privacidade de seus conteúdos e dos indivíduos que terão suas informações acessadas;
- II)** Não repassar os dados acessados ou o banco de dados em sua íntegra e/ou parcialmente, às pessoas não envolvidas na equipe da pesquisa anteriormente mencionada;
- III)** Declaro que os dados acessados serão usados exclusivamente para a(s) finalidade(s) previstas no protocolo do projeto vinculado (citado nas linhas iniciais deste termo);
- IV)** Asseguro o compromisso com a privacidade e a confidencialidade dos dados utilizados, preservando plenamente o anonimato e a imagem do participante de pesquisa.

Local, dia de mês de ano.

Nome e Assinatura gov.br OU outra assinatura eletrônica equivalente do solicitante/Aluno

Nome e assinatura gov.br OU outra assinatura eletrônica equivalente do Orientador