

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
DEPARTAMENTO DE LETRAS CLÁSSICAS E VERNÁCULAS
MONOGRAFIA**

**FUNDAMENTOS PARA A ETIQUETAGEM E IDENTIFICAÇÃO DE
ETNOTEXTOS DO BANCO DE DADOS ALMA-H**

Gabriel Schmitt

**Professor Orientador:
Prof. Dr. Cléo Vilson Altenhofen**

**Porto Alegre
2019**

Gabriel Schmitt

**FUNDAMENTOS PARA A ETIQUETAGEM E IDENTIFICAÇÃO DE
ETNOTEXTOS DO BANCO DE DADOS ALMA-H**

Trabalho de Conclusão de Curso apresentado ao Instituto de Letras da Universidade Federal do Rio Grande do Sul como requisito parcial para a conclusão do curso de Licenciatura em Letras.

Professor Orientador:
Prof. Dr. Cléo Vilson Altenhofen

PORTO ALEGRE, 2019

Agradecimentos

Deixo os meus sinceros agradecimentos a todos aqueles que, com seu apoio, tornaram possível a realização do curso de Licenciatura em Letras, deste trabalho e das pesquisas que o subsidiam.

Agradeço aos meus pais, Solange Both Schmitt e Gilnei Schmitt, por terem me ensinado as coisas mais importantes da vida, e entre elas, o Hunsrückisch. Agradeço e agradeço meu pequeno irmãozão Uesley (*der is schon greeser als ich*) pelos anos que moramos no mesmo quarto.

Agradeço profundamente ao Prof. Dr. Cléo Vilson Altenhofen, que me mostrou que era possível encarar minha biografia linguística, principalmente em relação ao alemão que eu ouvi sendo falado durante toda minha vida, numa perspectiva diferente e positiva. Agradeço por compartilhar seu conhecimento, discutir e revisitar conceitos e noções, por ter sido favorável às saídas de campo dos estudantes da UFRGS. Parece redundante mas é importante: sem o Prof. Cléo, nada disso aqui existiria.

Agradeço ao meu primo Alex, que iniciou comigo a jornada da literatura, me apresentou Rammstein e trocou comigo minhas primeiras palavras em inglês.

Agradeço a minha prima Iasmin por estar na linha de frente do ensino em situação bilíngue.

Agradeço a minha vó Alícia Schmitt, por poder ser para ela o primeiro neto graduado.

Agradeço aos meus avós Romeu e Maria Nair Both, por terem desbravado Cunhataí, oeste de Santa Catarina.

Agradeço a meus padrinhos e madrinhas.

Agradeço à minha agora pequena afilhada Alice Schmitt;

Agradeço a todos os moradores e ex-moradores da Vila Olinda.

De maneira geral, gostaria de agradecer a todos os meus professores, especialmente àqueles que já me deram aula ao longo da vida. Fundamentalmente os professores da EMEF Luiz Loeser, em Nova Petrópolis, e aos professores do então chamado Instituto de Educação Ivoti. E, de maneira mais calorosa ainda, a todas as professoras e professores do Instituto de Letras e da Faculdade de Educação da Universidade Federal do Rio Grande do Sul.

Agradeço a todos os pesquisadores e professores do Anfiteatro Grego do Campus do Vale.

Agradeço a todos os funcionários terceirizados da UFRGS, em especial a dona Rita, aos trabalhadores dos RUs e os funcionários da Casa de Estudante Universitário.

Agradeço aos motoristas e cobradores de ônibus por me levar ao Campus do Vale por mais de 06 anos e meio.

Herzlichen Dank an Doris und Heinz Wittig, die mich mit so viel Aufmerksamkeit aufgenommen haben.

Danke an Jörg, der mich in einem kritischen Zeitpunkt beraten hat.

Agradeço também ao Lucas Löff Machado, à Viktorya Zalewski Pietsch dos Santos, à Paola Inhaquite Wollmann, à Iona Gessinger, à Bianca Barreto, por discutirem comigo sobre ensino de alemão.

Agradeço imensamente aos colegas e amigos Gerônimo Loss Bergmann, Angélica Prediger, Jussara Habel, Luana Cyntia dos Santos Souza, Sofia Froehlich Kohl, todas pessoas brilhantes. Obrigado de coração por tudo que fizemos no IHLBrI, foi um aprendizado inigualável. Fico muito feliz de ter tido a oportunidade de fazer pesquisa ao lado de vocês.

Agradeço à Universidade Federal do Rio Grande do Sul por me proporcionar a oportunidade de realizar minha graduação em Letras Licenciatura Alemão-Português com qualidade e dedicação.

Agradeço a minha amada companheira Ana Winckelmann pelo amor, pelos diálogos, pelos nossos projetos e pelas conquistas, pelo documentário e pela nossa amizade. Por compartilharmos a busca da compreensão dos nossos sentimentos. Amada, você foi fundamental para eu ver um sentido em tudo isso.

Resumo

O presente Trabalho de Conclusão de Curso (TCC) tem como tema a etiquetagem e identificação de etnotextos do banco de dados ALMA-H (*Atlas Linguístico-contatual das Minorias Alemãs na Bacia do Prata: Hunsrückisch*), coordenado por C.V. Altenhofen e H. Thun. *Texto* é definido por Pelen (1984) como estrutura decorrente da corrente de discurso e da consciência do enunciador e do ouvinte quanto a escolha não apenas de palavras, mas de registros específicos, relacionados a um código superior. Os estilos de uso da língua refletem práticas linguísticas dos falantes e aspectos da cultura local; esses “textos orais” são chamados de etnotextos, definidos por Pelen (1, p. 709) como “*discourse which a community makes with reference itself*”. Sua relevância reside no fato de propiciar uma gama de análises suplementares da língua em um estilo espontâneo, para além dos dados recolhidos por meio de questionário estruturado. Essas análises abordam desde a variação morfossintática e diafásica (entre estilos de fala), até aspectos do uso da língua em situação plurilíngue e plurivarietal, envolvendo por exemplo o *code switching*. Da correta identificação e etiquetagem dessas amostras de uso da língua depende, porém, o êxito e funcionalidade do banco de dados para atender aos diversos interesses de pesquisa. O registro dos dados orais no alfabeto tradicional de escrita do alemão é definido, no Projeto, como uma *transliteração*, em oposição à *transcrição fonética*, que se vale de um sistema fonético como o IPA (*International Phonetic Alphabet*). Devido à ausência de uma prática de escrita regular do Hunsrückisch, o ALMA-H, como também o IHLBrI (Inventário do Hunsrückisch como Língua Brasileira de Imigração), teve que adotar um sistema de escrita próprio, denominado ESCRITHU (cf. ALTENHOFEN *et al.*, 2007; ALTENHOFEN; HABEL; PREDIGER, 2018), que dialoga com a grafia da norma *standard* do alemão e que constitui a base da transliteração dos dados do Projeto. Conforme Altenhofen; Morello *et al.*, (2018, p. 99), “os etnotextos podem oferecer um quadro significativo de elementos da cultura local, visando tanto subsidiar a análise sociológica do estado da língua e do plurilinguismo na localidade, quanto [...] ações futuras de promoção e educação plurilíngue”. A pesquisa deste trabalho faz o seguinte questionamento: como organizar os metadados para que possam ser utilizados de maneira mais eficaz pelos pesquisadores? A relevância deste estudo está em discutir aspectos teóricos para dar sequência às pesquisas do Projeto ALMA-H, que tem por objetivo o estudo do *Hunsrückisch*, assim como oferecer materiais que possam ser usados futuramente para a elaboração do dicionário do *Hunsrückisch*.

Palavras-chave: Enotexto; Variedades Alemãs; *Hunsrückisch*; Transliteração; Línguas do Brasil.

Zusammenfassung

Die vorliegende Abschlussarbeit stellt eine Analyse der Etikettierung und Identifizierung von Ethnotexte im Rahmen des ALMA-H (*Sprachkontaktatlas der deutschen Minderheiten im La Plata-Becken – Hunsrückisch*), koordiniert von Professor C.V. Altenhofen (UFRGS) und H. Thun (CAU), dar. Pelen (1986) definiert *Text* als die Struktur des Strömung vom Diskurs sowie die Bewusstsein von wem spricht und hört, nicht nur um Wörter auszuwählen, ebenfalls von bestimmten Aufzeichnungen in Bezug auf höheren Codes. Die Arten des Sprachgebrauchs spiegeln die sprachlichen Praktiken der Sprecher und Aspekte der lokalen Kultur wider. Die „orale Texte“ werden *Ethnotexte* genannt, und von Pelen (1986, p. 709) "*discourse which a community makes with reference itself*" definiert. Seine Relevanz liegt in der Tatsache, dass es eine Reihe von zusätzlichen Analysen der Sprache in einem spontanen Stil bietet, jenseits die Daten aus dem festgelegten Fragebogen. Die Analyse fassen die morphologiesche Variation und die Diaphasie, und sogar die Sprache in plurilinguistische und plurivarietäten Situationen, bzw. *code switching*. Da der Erfolg und die Funktionalität der Datenbank jedoch von der korrekte Identifizierung und Kennzeichnung dieser Sprachgebrauchsproben abhängt, ist das Hauptziel dieser Arbeit, fundierten theoretische und methodische Grundlage für die Identifizierung und Kennzeichnung der Sprachgebrauchsproben zu schaffen und über die zukünftigen Benutzer der Datenbank nachzudenken. Die Aufzeichnung oralen Daten im traditionellen Alphabet wird im Rahmen des Projekts ALMA-H als *Transliteration* definiert, im Gegensatz zur phonetischen *Transkription*, die ein phonetisches System wie das IPA (*International Phonetic Alphabet*) verwendet. Aufgrund des Mangels an einer regelmäßigen Schreibpraxis auf Hunsrückisch, wurde ein eigenes Schriftsystem übergenommen, genannt ESCRITHU (vgl. ALTENHOFEN *et al.*, 2007; ALTENHOFEN; HABEL; PREDIGER, 2018), das mit dem schriftlichen standard Deutsch Beziehung setzt und die Transliterationen des ALMA-H basierten. Vgl. Altenhofen; Morello *et al.*, (2018, p. 99), "os etnotextos podem oferecer um quadro significativo de elementos da cultura local, visando tanto subsidiar a análise sociológica do estado da língua e do plurilinguismo na localidade, quanto [...] ações futuras de promoção e educação plurilíngue". Diese Arbeit stellt die folgende Forschungsfrage: Wie können die Metadaten so organisiert werden, dass sie von den Forschern effektiver genutzt werden können? Die Relevanz dieser folgenden Forschung des ALMA-H Projekts Arbeit liegt an theoretischen Aspekte diskutieren, um Materialien anzubieten, die für die Untersuchung der deutschen Sprachen in Süd-America und für die Entwicklung des hunsrückischen Wörterbuches verwendet werden können.

SCHLÜSSELWÖRTER: Ethnotext; Transliteration, Hunsrückisch; Deutsch; Deutsche Einwanderungssprache.

Lista de figuras, gráficos, quadros e tabelas

Fig. 01 – Transcrição fonética na cartografia da variação em <i>verlieren</i>	19
Fig. 02 – Dimensões de análise do ALMA-H.....	21
Fig. 03 – Estrutura do questionário do ALMA-H usada na etiquetagem.....	25
Fig. 04 – Exemplo de etnotexto do IHLBrI.....	27
Fig. 05 – Estrutura-base do cabeçalho de metadados no ALMA-H.....	28
Fig. 06 – Dados de anotações do Arquivo de Wossidlo transcritos	30
Fig. 07 – Ficha Técnica de uma entrevista do REPHO.....	31
Fig. 08 – Interface online do REPHO.....	32
Fig. 09 – Relação dos metadados identificados nos etnotextos analisados.....	38
Fig. 10 – Sugestão para cabeçalho de metadados do ALMA-H.....	39

Sumário

Introdução.....	9
Cap. 1 – Contextualização do Estudo.....	13
Cap. 2 – Fundamentos Teóricos.....	17
2.1 O conceito de <i>etnotexto</i> e sua relevância para a pesquisa dialetológica.....	17
2.2 Sistemas de escrita: transliteração e transcrição fonética.....	18
2.3 Pluridimensionalidade da análise da variação em dados de fala.....	20
2.4 Tipologização dos etnotextos: gêneros textuais e campos temáticos.....	22
2.5 Variação estilística: graus de espontaneidade e de formalidade da fala.....	23
Cap. 3 – Análise.....	25
3.1 Estrutura de base para a etiquetagem dos dados.....	25
3.2 Informações para o cabeçalho de metadados na transliteração de etnotextos.....	26
3.3 Identificação de dados de pesquisa nos projetos WossiDiA e REPHO.....	29
3.5 Como informar os participantes da entrevista?.....	35
3.6 Metadados relativos ao processo de transliteração dos etnotextos.....	36
3.7 Proposta de “cabeçalho de metadados”	38
Palavras finais.....	40
Referências bibliográficas.....	41
Anexo.....	43

Introdução

A palavra falada é a palavra viva. E há centenas de anos temos pessoas que se esforçam para registrar as manifestações culturais populares. Na Alemanha, as coletâneas de narrativas orais e descrição linguística dos irmãos Grimm, a iniciativa de Georg Wenker no DSA (*Deutsche Sprachatlas*), ou o legado do linguista e folclorista Richard Wossidlo e seu „Mecklenburgisches Wörterbuch“. No Brasil, os trabalhos por exemplo de Guimarães Rosa, Simões Lopes Neto e também de Mário de Andrade, ressaltaram a narrativa oral não apenas pelo conteúdo, mas também pela forma. Os documentários de Eduardo Coutinho deram voz e rosto para pessoas marginalizadas pelas dinâmicas sociais. A narrativa das pessoas *reais*, que relembrem *da* história por meio *de* histórias, despertam esse interesse, num processo que envolve presença e oralidade, articuladas em narração.

No centro deste trabalho está a etiquetagem e identificação de etnotextos do banco de dados ALMA-H (*Atlas Linguístico-contatual das Minorias Alemãs na Bacia do Prata: Hunsrückisch*), do qual participo como graduando e bolsista de iniciação científica desde 2015. Por *etnotextos* entende-se aqui textos representativos de uma cultura e, por extensão, da cultura linguística de um grupo ou comunidade (ver seção 3.1). Neste sentido, a representação da palavra falada se dará, neste trabalho, pela palavra escrita. Para a transliteração de um etnotexto, ou seja, para o registro de uma fala em texto, há um processo que chamaremos de *transliteração*. Entende-se por *transliteração* o registro de dados orais no alfabeto tradicional de escrita do alemão, em oposição à *transcrição fonética*, que se vale de um sistema fonético como o IPA (*International Phonetic Alphabet*).

Este trabalho apresenta uma discussão sobre aspectos técnicos e teóricos da transcrição de entrevistas realizadas pelo Projeto ALMA-H (*Atlas Linguístico-Contatual das Minorias Alemãs na Bacia do Prata – Hunsrückisch*), coordenado por C.V. Altenhofen (UFRGS) em parceria com H. Thun (CAU, Kiel), assim como também do IHLBrI (*Inventário do Hunsrückisch como Língua Brasileira de Imigração*), coordenado por R. Morello (IPOL) e C.V. Altenhofen (UFRGS/ALMA-H). Essa base de dados será apresentada no Cap. 1.

O interesse na análise de etnotextos do Projeto ALMA-H reside, especialmente, no fato de representarem um estilo de fala mais espontânea, de grande relevância para o

estudo da variação linguística, para além dos dados coletados por meio de questionário estruturado. Através dos etnotextos, pode-se analisar as falas entre os entrevistados, ou conversas de pessoas que não estão oficialmente sendo entrevistadas com os entrevistados, e levantar alguns dados sobre usos mais espontâneos. Essas análises abordam desde a variação morfosintática e diafásica (entre estilos de fala) até aspectos do uso da língua em situação plurilíngue e plurivarietal, envolvendo por exemplo o *code switching*. Além disso, as diferentes temáticas dos etnotextos espelham a cultura local na sua forma de abordar o mundo, na maneira de se fazer humor, ou nos ingredientes de uma receita. A diversidade de temáticas pode interessar diferentes públicos.

Da correta identificação e etiquetagem dessas amostras de uso da língua depende o êxito e funcionalidade do banco de dados, para atender aos diversos interesses de pesquisa. Daí o objetivo principal desse trabalho, que é analisar e fazer sugestões ao processo de etiquetagem e principalmente de identificação de etnotextos, contribuindo para a constituição de um *corpus* amplo da língua alemã falada no Brasil.

A priori, os etnotextos não são dados com a característica de poderem ser sistematicamente comparados entre si. É através da etiquetagem e catalogação precisa dos respectivos arquivos de dados que se propicia uma análise comparativa sistemática. A noção de *etnotexto* combina com a ideia de uma coletânea: cada etnotexto é um recorte de fala culturalmente significativa em uma língua. Uma das aplicações possíveis de um banco de dados de transliteração de entrevistas é o acesso por um mecanismo de busca semelhante a uma biblioteca virtual. Ou seja, tem-se em vista a organização dos metadados a fim de facilitar a busca de informações nas entrevistas de maneira eficaz por pesquisadores e demais usuários do banco de dados. Para que um eventual mecanismo de busca seja programado e possa fazer a leitura deste banco de dados, é preciso organizar os metadados dentro do que chamaremos de *Cabeçalhos* (ver seção 3.1).

A metodologia para a organização dos metadados e para a constituição do *corpus* do ALMA-H segue os princípios da Geolinguística Pluridimensional, proposta por Harald Thun. As dimensões de análise serão apresentadas na seção 2.2. Devido à ausência de uma prática de escrita regular do Hunsrückisch, o ALMA-H, como também o IHLBrI, teve que adaptar um sistema de escrita próprio, denominado ESCRITHU (cf. ALTENHOFEN *et al.*, 2007; ALTENHOFEN; HABEL; PREDIGER,

2018), que dialoga com a grafia da norma *standard* do alemão e que constitui a base da transliteração dos dados do Projeto.

Os objetivos específicos são responder as seguintes questões:

- a) Como e quais metadados devem compor o cabeçalho do etnotexto transliterado?
- b) Quais dados de identificação da situação de produção do etnotexto devem ser considerados (pense-se na localidade da pesquisa, local e situação de realização da entrevista, etc.)?
- c) Como definir os recortes de um etnotexto, em determinado segmento de fala gravado em arquivo de áudio/vídeo?
- d) Como identificar os participantes das entrevistas, considerando o princípio da pluridimensionalidade de análise da variação linguística?
- e) Qual a relevância de registrar os transcritores e revisores das transliterações?
- f) Quais critérios considerar para a escolha da língua a ser usada na etiquetagem e identificação do conteúdo dos etnotextos transliterados?

Para responder a essas questões, o trabalho se divide em três capítulos. O primeiro capítulo dedica-se à descrição do contexto da pesquisa e que motiva a presente análise. O cap. 2 apresenta os fundamentos teóricos considerados na análise e discussão da identificação e etiquetagem de etnotextos de fala espontânea e oral. O cap. 3 apresenta um etnotexto em Hunsrückisch do projeto ALMA-H. Também, são analisados exemplos de projetos afins, em especial do Arquivo Wossidlo (Univ. Rostock) e do projeto REPHO. Segue-se, nas palavras finais, uma proposta de “cabeçalho” (de identificação) e de critérios de etiquetagem, com base na discussão feita.

A compilação de etnotextos de uma língua de tradição essencialmente oral dá a possibilidade de visibilizar a língua e a cultura local. Seu potencial para uso em ações de conscientização e educação plurilinguística (BROCH, 2014) é grande. Não está em jogo, portanto, apenas a dimensão metodológica de organização do banco de dados, pensando em futuros usuários e possíveis objetivos de busca. Também se favorece a produção de materiais educativos sobre a língua. Assim, uma língua que, aos olhos da sociedade, é vista muitas vezes como um *atrapalho* na aprendizagem do português e do alemão *standard* teria sua escrita entrando numa circulação maior, podendo como tal ser discutida por quem a lê. Com isso, disponibiliza-se material para a qualificação da discussão sobre as muitas variedades de alemão faladas no Brasil, discussão que ingratamente às vezes se resume a

classificar e rotular ‘isso é dialeto’, ‘isso não é’. Qualquer discussão qualificada sobre o assunto solicita que as partes tenham acesso a materiais de qualidade, confiáveis, estáveis e sobretudo autênticos sobre e principalmente *na* língua.

Os desdobramentos e motivações deste TCC têm por pano de fundo contribuir para o estudo e conhecimento de uma língua minoritária do Brasil que corresponde a menos de 0,5% da população brasileira, na qual me incluo como autor deste trabalho. Diante da observação de que os jovens falam cada vez menos essa variedade em relação às gerações mais velhas, o registro do *Hunsrückisch* assume significado especialmente relevante, não apenas como acervo de dados que pode ser usado em um futuro museu: garante-se, deste modo, material de análise, consulta e LEITURA em *Hunsrückisch* para falantes ainda vivos hoje. O público de interesse pela língua de imigração vai muito além dos pesquisadores e estudantes das Letras, de história, arte, literatura oral, geografia, políticas públicas, publicidade, psicologia, gastronomia, direito, enfermagem, medicina, relações internacionais: há de interessar aos próprios falantes. É essa perspectiva que norteia também o presente estudo, ao buscar selecionar critérios estáveis para a organização dos metadados, a fim de facilitar uma possível biblioteca virtual de etnotextos em alemão falado na América do Sul. Neste sentido, este estudo serve de embasamento e preparação não apenas para a pesquisa do Mestrado no EMLex (*European Master in Lexicography*), que iniciarei a partir de outubro de 2019, na Universidade de Erlangen-Nürnberg, como também para um futuro dicionário do *Hunsrückisch*, um dos objetivos do Projeto ALMA-H (v. ANSCHAU, 2010).

Cap. 1 – Contextualização do Estudo

Conforme já se aludiu, o problema de pesquisa de que trata este trabalho surge em meio a um volume significativamente grande de dados orais que é preciso identificar e distinguir adequadamente. Trata-se de um problema de ordem mais metodológica, relativo ao tratamento de dados orais para a análise. Neste sentido, a transliteração, etiquetagem e tipologização desses dados visibiliza seu conteúdo (*visible speech*) e o torna acessível à busca a possíveis usuários. Cabe, inicialmente, delimitar de forma mais clara o contexto que gera essa base de dados, bem como as motivações e objetivos que subjazem a sua produção. Os dados que servem de base à análise feita aqui são, neste caso, de dois projetos já mencionados, o ALMA-H e o IHLBrI. Ambos os projetos apresentam motivação diferente, mas complementar e se ocupam com o Hunsrückisch, uma língua de imigração alemã que tem sua origem no médio-alemão centro-ocidental (*Westmitteldeutsch*), mais precisamente na região do Hunsrück e de seu entorno, na Renânia Central.

De uma maneira ou outra, cada ramo das Letras utiliza um *corpus* linguístico, bem definido ou não, que serve de parâmetro ou base empírica para uma análise consistente. Esses *corpora* podem ser configurados a partir de diferentes fontes: jornais, livros, revistas, ou, o que vamos tratar aqui, entrevistas, mais especificamente, entrevistas (semi-)conduzidas por um questionário. Conforme Manzini (2012, p. 155 e 156), há ao menos três tipos de questionário: não estruturados, semiestruturados e estruturados, e que são definidos por ele da seguinte maneira:

A entrevista **não-estruturada** é utilizada por um número reduzido de participantes. É uma entrevista que se inicia com uma pergunta geradora e o pesquisador não possui um roteiro previamente estabelecido. A entrevista é difícil de ser conduzida e recomendada para pesquisadores e entrevistadores experientes. Com o objetivo da pesquisa em mente, o entrevistador irá conduzindo a entrevista tentando relacionar o objetivo pretendido com as informações que estão sendo apresentadas pelo participante. Geralmente a entrevista é mais longa e, na maioria das vezes, o pesquisador retorna para realizar outras entrevistas com o participante para aprofundar o nível de informação. É indicada para estudar particularidades de forma profunda.

A entrevista **estruturada** tem como característica um roteiro prévio, com perguntas fechadas. O roteiro é usado pelo entrevistador (ou entrevistadores) na mesma sequência em que foi elaborado. É indicada para pesquisa com grandes populações, a análise é quantitativa e usa como procedimento de análise os conhecimentos da estatística. O roteiro precisa ser muito bem trabalhado e o pesquisador deve conhecer a fundo o tema da pesquisa, para inclusive elaborar as alternativas de resposta.

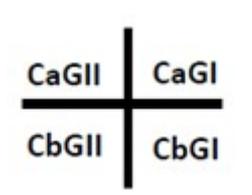
A entrevista **semiestruturada** tem como característica um roteiro com perguntas abertas e é indicada para estudar um fenômeno com uma população específica: grupo de professores; grupo de alunos; grupo de enfermeiras, etc. Deve existir flexibilidade na sequência da apresentação das perguntas ao entrevistado e o entrevistador pode realizar perguntas complementares para entender melhor o fenômeno em pauta.

Segundo essa descrição, as ‘conversas livres’, ou seja, conversas espontâneas são esperadas em uma entrevista não-estruturada. Em uma entrevista semiestruturada, subentende-se dados de conversa semidirigida, ou seja, há espontaneidade na fala, mas a demarcação clara entre entrevistador e entrevistado não permite depreender um segmento de fala totalmente espontâneo, por isso será referido nesse trabalho como ‘mais espontâneo’. Em entrevistas estruturadas, os dados esperados seriam totalmente condicionados às perguntas.

É possível que, em entrevistas estruturadas surjam etnotextos sobre temas que não são o do questionário. Um dos questionamentos desse trabalho é como identificar um etnotexto que foi gravado num contexto de pesquisa que não trata exatamente sobre aquele tema. Cabe ao pesquisador definir então um recorte para a transliteração, e o critério deve ser a temática. A sessão 2.4 continuará essa discussão sobre o corte da transliteração.

O ALMA-H é um projeto que tem por objetivo maior a confecção de mapas para a realização de macroanálises de tendências da variação do Hunsrückisch no espaço pluridimensional. O ALMA-H faz parte da chamada *trilogia rio-platense*, de Harald Thun, que engloba três atlas linguísticos que buscam abranger o conjunto da diversidade de contatos linguísticos da Bacia do Rio da Prata: o ADDU (*Atlas Lingüístico Diatópico y Diastrático del Uruguay*); o ALGR (*Atlas Lingüístico Guaraní-Románico*); além do ALMA-H (*Atlas Lingüístico-Contatual das Minorias Alemãs na Bacia do Prata: Hunsrückisch*), que tem por ênfase o contato de uma língua de imigração minoritária com as línguas de colonização, português e espanhol.

No âmbito do ALMA-H, foram feitas entrevistas em 41 localidades, com até quatro grupos sociais distintos por localidade, conforme a geração (GI = 18 a 36 anos; GII = acima de 55 anos) e a escolaridade (Ca = em nível superior; Cb = com escolaridade até o ensino médio). Com isso, totalizam-se 128 entrevistas estruturadas com questionário que podem envolver os seguintes quatro grupos de falantes:



As entrevistas do ALMA-H seguiram o princípio da pluridimensionalidade na análise da variação linguística (cf. THUN, 1998). Segundo esse princípio, correlacionam-se os dados da variação linguística com diferentes dimensões extralinguísticas de análise, englobando idade, escolaridade, situação de fala, entre outras (cf. ALTENHOFEN e THUN, 2016). Essas dimensões, que neste trabalho têm uma relação profunda e estreita com a maneira mais eficaz de organizar os metadados, serão apresentadas no cap. 2.3.

Ao lado do atlas linguístico, o projeto ALMA-H abriga uma série de subprojetos, entre os quais se inclui o Inventário do Hunsrückisch (ou IHLBrI), desenvolvido de 2017 a 2018 em parceria com o IPOL (*Instituto de Investigação e Desenvolvimento em Política Linguística*). Os levantamentos do IHLBrI atenderam a uma preocupação voltada ao reconhecimento, salvaguarda e promoção da língua como patrimônio cultural imaterial. O Inventário do Hunsrückisch foi concebido em consonância com a política do INDL (*Inventário Nacional da Diversidade Linguística*), instituída por meio do Decreto nº 7387/2010, para ser um instrumento de “identificação, documentação, reconhecimento e valorização das línguas”. O banco de dados de entrevistas do IHLBrI inclui dados em formato de vídeo, áudio e fotografia, além de reunir um amplo acervo de áudio-visuais e impressos sobre a língua inventariada, que complementam a base de dados do ALMA-H. Enquanto o ALMA-H prioriza a descrição da variação e mudança linguística do Hunsrückisch, o IHLBrI enfatiza a coleta de etnotextos em forma de relatos, depoimentos e explicações sobre fatos da língua, que também aparecem no ALMA-H, porém não exclusivamente.

As falas mais espontâneas podem ser encontradas justamente nos trechos de entrevista feita com a lista de temas para etnotexto. Essa lista foi desenvolvida a partir da *Parte E* do questionário do ALMA-H, e sugere temas para conversas ‘semidirigidas’, a partir de diferentes áreas de conhecimento e temas que poderiam ser abordados com diferentes entrevistados, conforme a situação.

Como em todo atlas linguístico, para garantir a devida identificação do ponto de pesquisa onde o dado foi coletado, cada ponto recebe um código. Assim, RS06 refere-se à localidade de Nova Petrópolis, que no IHLBrI é o centro de uma comunidade de referência que inclui outras localidades do entorno, neste caso, Picada Café e Linha Nova.

É importante salientar que, neste momento, as últimas transliterações em Hunsrückisch foram feitas no âmbito do IHLBrI, principalmente para o documentário “Viver no Brasil falando Hunsrückisch”. Portanto, foi durante a transliteração das cerca de 130 páginas, as quais compõem o banco de dados de transliterações do IHLBrI, que foram feitos os últimos avanços na identificação dos metadados de etnotextos em Hunsrückisch. Será apresentado um exemplo de transliteração desse projeto na seção 3.1.

Cap. 2 – Fundamentos Teóricos

2.1 O conceito de *etnotexto* e sua relevância para a pesquisa dialetológica

Primeiro, precisamos concordar que “a língua não existe separada da cultura, isto é, do conjunto socialmente herdado de práticas que determina a textura de nossas vidas” (SAPIR, 1921, p. 170-171). Isso significa que podemos encontrar elementos culturais também em trechos de discurso e em textos, em suma no uso da língua. A partir dos anos 70, a dialetologia passou a refletir o papel do discurso oral na pesquisa. O termo *ethnotext* surge em meio a esse debate, sobretudo na França. Segundo Pelen (1991), etnotextos são textos em que a própria comunidade fala sobre si, revelando deste modo consciência sobre sua cultura.

Semelhante à ideia de Platão de que existiria o mundo das ideias e o mundo real, concebe-se o *Etnotexto* (com *E* maiúsculo) como o ideal da língua legitimando a cultura da comunidade, naquilo que reflete costumes, hábitos, crenças, imaginário compartilhado em que a “comunidade se espelha, reproduz, se codifica e se decifra, se desenrola e principalmente, se garante e se legitima” (PELEN, 1986, tradução de SAMPAIO, 2001, p. 73). Em contrapartida, *etnotexto* (com *e* minúsculo) refere-se às diferentes manifestações de fato, imperfeitas porque nunca seriam completas ou totais, mantendo apenas traços pontuais do *Etnotexto* enquanto ideal, tendo em vista que este sempre será representado fora de seu contexto original de produção e enunciação. Essa característica dos etnotextos, de não serem completos e encerrados em si, fala a favor da consideração de uma coletânea de etnotextos, pois quando colocados lado a lado, abrem-se possibilidades de fazer comparações, de modos que as tendências da cultura podem ser observadas. O emprego deste conceito se justifica exatamente por esse motivo, tanto por abranger segmentos de fala espontânea maiores, quanto por incluir a relação entre língua e cultura, que no contexto da imigração e dos contatos linguísticos desempenha uma função importante.

De modo geral, diferentes tipos de texto podem, portanto, vir a ser tratados como etnotextos, tanto de origem oral quanto escrita. O presente trabalho enfoca especificamente a organização de etnotextos oriundos de entrevistas estruturadas ou semi-estruturadas.

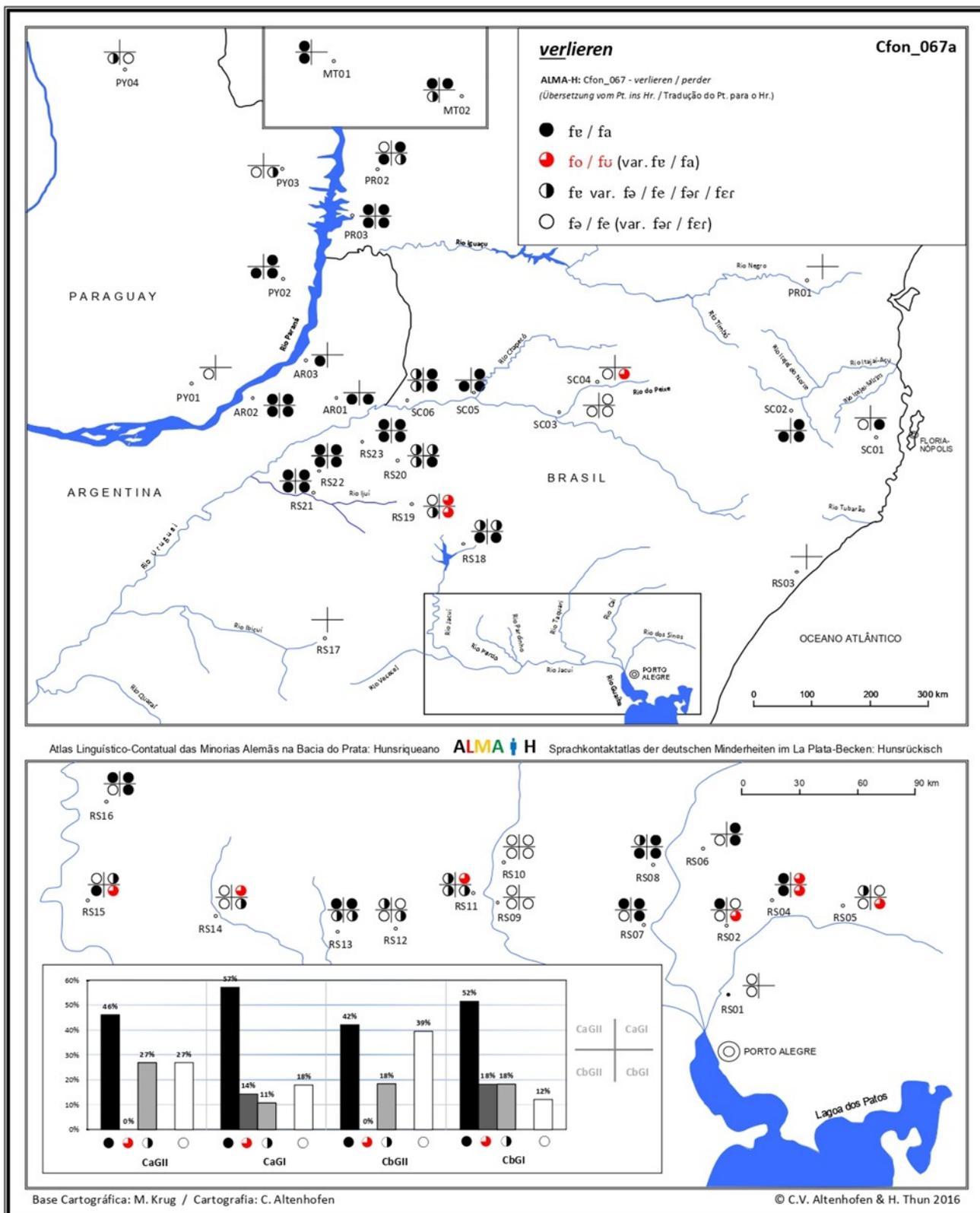
2.2 Sistemas de escrita: transliteração e transcrição fonética

Desde o estabelecimento dos *Fundamentos para uma escrita do Hunsrückisch falado no Brasil* (v. ALTENHOFEN et al., 2007), a transliteração de trechos de entrevista começou a se consolidar como um processo cada vez mais significativo para a visualização e apresentação de dados, a partir das orientações do ESCRITHU. Entre os principais méritos desse sistema de escrita estão o diálogo com a norma do alemão *standard*, permitindo um alcance e intercompreensão maior, e a possibilidade de respeitar as variações internas do Hunsrückisch (cf. ALTENHOFEN, PREDIGER e HABEL, 2018).

A definição do termo *transliteração* se dá em contraste com a *transcrição*, sendo ambas as escritas utilizadas no processo de cartografia do ALMA-H. A *transcrição fonética* é importante para análises principalmente do nível fonético/fonológico, pois possibilita uma visualização mais precisa dos dados. No arquivo-matrix da cartografia do ALMA-H, essa transcrição é feita utilizando-se o alfabeto IPA, pois quando se quer fazer mapas linguísticos, procura-se muitas vezes pares mínimos. Pode-se ver um exemplo de transcrição na legenda do mapa da fig. 01, no qual se analisa a variação fonética do prefixo *ver-* na pronúncia do verbo *verlieren* ‘perder’.

Pode ser prudente acentuar que o termo *transcrição* tem ao menos dois sentidos; em alguns casos, vem sendo usado como sinônimo de *transliteração*. Como já foi dito, a linguística utiliza-se da *transcrição fonética* para fazer a descrição detalhada dos sons de um determinado segmento de fala, usando alfabetos fonéticos. Mas também vem sendo utilizado nas ciências sociais e de linguagem o termo *transcrição de entrevistas*, para o processo de tornar visível em texto dados coletados em entrevista, utilizando uma língua-padrão, consolidada pela literatura e pela história. Também nesse sentido, Altenhofen (2017), numa resenha sobre os dois primeiros tomos do *Atlas Linguístico Guarani-Românico: Sociología* (ALGR-S, coordenação de H. Thun), observa que o *corpus* de guarani escrito foi constituído através de segmentos de fala do guarani: “[...] A **transcrição das falas** em guarani, por sua vez, segue a ortografia sugerida por Natalia Krivoshein de Canese e Feliciano Acosta Alcaraz (1997).” O processo de transcrever entrevistas, ou falas, seguindo um modelo de escrita, pode ser chamado de *transliteração*, pois o objetivo está no enquadramento linguístico escrito. Por hora, nos contentamos em perceber que *transcrição de entrevistas* é o que mais se aproxima do que definimos como *transliteração de etnotextos*.

Fig. 01 – Transcrição fonética na cartografia da variação em *verlieren*



O pesquisador pode se valer dos dois alfabetos num processo de escrita, a fim de facilitar a compreensão dos fenômenos da língua analisada, principalmente quanto aos comentários metalinguísticos. Também, muitas vezes, a transliteração pode deixar dúvida quanto a alguns aspectos variacionais, justamente porque muitas vezes é impossível manter tanto a relação filológica quanto a fidelidade sonora. Por exemplo, para registrar a variedade de *Kaste* em que o falante reforça o fone /ʃ/, poder-se-ia escrever *Kaschte*, mas com o risco de uma compreensão menor. Parece que a alternativa de uma transcrição fonética deixa mais clara a interpretação do dado, embora não seja tão universal (considerando usuários do banco de dados vindos de áreas distintas). Ela pode ocorrer de forma integral – por exemplo, [ˈkaʃtə] – ou parcial – por exemplo *Ka[ʃ]te*.

2.3 Pluridimensionalidade da análise da variação em dados de fala

Os primeiros grandes atlas linguísticos, como o *Deutscher Sprachatlas* (DAS, de Georg Wenker, e o *Atlas linguistique de la France* (ALF), de Jules Gilliéron, lançavam mão unicamente da dimensão diatópica, ou seja, a dimensão espacial. Um único falante representava um ponto de pesquisa. É preciso reconhecer que esses trabalhos possibilitaram avanços no campo da geolinguística, principalmente num período, no qual, se não era possível, era pouco prático fazer registros acústicos.

Porém, como escreve THUN (2005), estudar um dialeto a fundo não significa registrar apenas sua forma mais antiga, mais arcaica. Para justificar isso o autor cita os estudos de Rousselot, que já em 1879 afirmava que em sua própria casa se falavam variedades distintas da mesma língua, e uma investigação profunda deveria levar em conta variações entre as gerações (dimensão diageracional) e entre os sexos (diassexual), binariamente divididos entre homens e mulheres, velhos e jovens. Além dessas duas dimensões que foram estudadas dentro de sua casa, Rousselot apontou que tinha que ser levado em conta a distinção na fala de diferentes classes sociais (dimensão diastrática).

Essas dimensões de análise foram incorporadas à concepção de atlas pluridimensionais. O quadro a seguir apresenta um resumo das dimensões abordadas pelo ALMA-H.

Fig. 02 – Dimensões de análise do ALMA-H

Dimensão	Parâmetro	Critério
Diatópica	Topostático (informantes em um domicílio fixo)	41 pontos de inquérito.
Diatópica-cinética	topodinâmico (mudança de domicílio – mobilidade espacial)	Em grande parte, relação entre colônias velhas (matriz de partida) e colônias novas (matriz de chegada).
Diastrática	Ca = classe (socioculturalmente) alta Cb = classe (socioculturalmente) baixa	Ca (com formação universitária parcial ou completa). Cb (até ensino médio + profissão que não exija o uso da escrita).
Diageracional	GII (geração velha) GI (geração jovem)	= acima de 55 anos = 18 a 36 anos
Diassexual	Ho = homens Mu = mulheres	
Dialingual	hrs = hunsriqueano (Hunstrückisch) hdt = alemão-padrão (Hochdeutsch) pt = português sp = espanhol	Essa dimensão é complementada com dados dos atlas linguísticos do português (ALERS e ALiB) para o português.
Diafásica	Resp = respostas ao questionário Leit = leitura Tx = conversa livre (etnotextos)	Três estilos de uso da língua.
Diarreferencial	Lg = fala “objetiva” MLg = fala metalinguística	Esta dimensão é estimulada pela técnica de entrevista em três tempos: perguntar (resposta espontânea) – insistir – sugerir.
Diarreligiosa	Cat = católico Lut = evangélico-luterano	Tipo de localidade conforme as confissões religiosas presentes
Diamésica	Es cr = língua em meio escrito vs. Fal = meio falado	Coleta de dados em áudio e vídeo (oralidade) e em meio escrito (p. ex. impressos, cartas de imigrantes, inscrições em estabelecimentos comerciais, placas, topônimos e sepulturas)

Fonte: Altenhofen e Thun (2016, p. 375)

São especialmente relevantes para fins deste trabalho a dimensão diafásica e a dimensão diarreferencial. A diafasia está vinculada às situações e tipos de questionário utilizados na entrevista. Como pode ser observado acima, o ALMA-H utiliza-se de três estilos de uso da língua, por meio da forma do seu questionário: a) perguntas com resposta direta, a partir de um questionário pré-definido, o que caracteriza, conforme Manzini (2012, p. 155) uma entrevista estruturada; b) uma seção de leituras em diferentes línguas; c) temas para conversa semidirigida, o que, conforme Manzini (2012, p. 156) configura uma entrevista semi-estruturada.

A parte do questionário do ALMA-H de temas para conversa semidirigida foi ampliada e aprimorada no IHLBrI. A lista de temas utilizada apresenta flexibilidade e multiplicidade de áreas do conhecimento (história da localidade, biografia linguística, tempos escolares, fauna e flora da região, receitas etc.) e é adaptável à realidade dos

informantes da pesquisa. Foi possível, com ela, coletar dados metalinguísticos, sobre a percepção que os falantes têm de sua língua.

Os comentários metalinguísticos podem ser tão potentes que sintetizam os processos de nivelamento e contatos linguísticos de um determinado fenômeno, conforme Winckelmann (2018). Esses comentários são estimulados pela técnica em três tempos – pergunta, insistência e sugestão – que favorece a ocorrência de etnotextos e comentários sobre a sua língua e a língua dos outros.

Apesar da tendência de falas mais metalinguísticas na parte do questionário estruturado, é natural que numa entrevista longa também ocorram trechos de conversa mais espontânea, sobretudo quando o entrevistador sugerir temas previstos pela parte semi-estruturada do questionário. As dimensões de análise serão retomadas no cap. 3, ao se discutir a identificação dos metadados dos etnotextos.

2.4 Tipologização dos etnotextos: gêneros textuais e campos temáticos

Conforme Altenhofen (2004), o *corpus* de um atlas linguístico também pode fornecer material para análise e discussão de aspectos da língua, cuja análise dispense a consideração da diatopia (variação no espaço). A variação diatópica, como já dito anteriormente, esteve no princípio dos mapas monodimensionais, ou seja, está vinculada à ideia da representação básica dos mapas linguísticos, por exemplo, de uma variável fonético-fonológica. Contudo, o levantamento de dados para um *corpus* linguístico-contatual pode contemplar a análise de marcações culturais através da língua. Um dos recursos utilizados é o registro de conversa espontânea.

Ao se realizar análises de enunciados mais espontâneos, percebe-se tendências bem claras de tipos de narrativa em relação a sua estrutura. Isso não é nada mais do que dizer por exemplo que todas as piadas têm uma certa estrutura em comum, pois são construídas de uma maneira que um elemento no final da história deixe toda a narrativa engraçada; ou ainda que a descrição de uma receita culinária vai começar com um comentário geral sobre o prato, seguido pelos ingredientes e o modo de fazer.

Assim, fazer uma classificação relacionada a estrutura dos etnotexto pode facilitar pesquisas linguísticas. Por exemplo: para o estudo do uso ou perda de formas do pretérito no alemão falado na América do Sul, pode haver uma probabilidade maior de encontrar

ocorrências de pretérito em narrativas históricas, visto que narrativas históricas costumam utilizar verbos no passado.

Cada enunciado é sempre único, mas “cada campo de utilização da língua elabora seus *tipos relativamente estáveis* de enunciados, denominados *gêneros do discurso*”, conforme Bakhtin (2016). Esses gêneros são compostos por três elementos ‘básicos’: o conteúdo temático, o estilo e a forma composicional. Bakhtin também lembra que o contexto de enunciação é fundamental, ou seja, quem fala e onde, para quem fala de onde.

Ainda conforme Bakhtin (2016), parecem ser infinitas as possibilidades de classificação de gêneros textuais: por um lado, pelas inúmeras manifestações linguísticas possíveis no vasto emaranhado de culturas e línguas do nosso planeta, e por outro lado porque pouco se estudou sobre gêneros linguísticos que não fossem literários. No cap. 3.3, será discutido como organizar as características principais dos etnotextos do Projeto em relação aos gêneros orais, levando em conta que até este momento o Guia de Etiquetagem dos arquivos considera a classificação das transliterações conforme tipos de resposta aos diferentes questionários, conforme já foi dito no cap. 1.

2.5 Variação estilística: graus de espontaneidade e de formalidade da fala

A relevância de organizar um banco de dados de etnotextos no âmbito do Projeto ALMA-H está na característica dos graus de formalidade mais espontâneo de alguns trechos do questionário. Por isso, será feita uma reflexão sobre diferentes níveis de interação entre entrevistador e entrevistado.

Como foi apresentado no cap. 1, têm-se entrevistas com diferentes tipos de questionário. Mas toda entrevista pressupõe ao menos uma pessoa que realiza as perguntas, e outra(s) que as responde(m). Ao falar do processo de filmagem para seus documentários, Eduardo Coutinho (1997) aponta que há uma tendência de alguns tipos de documentários de apagar as provocações, os questionamentos dos entrevistadores, ou como ele chama, “dos questionadores”. Isso garante um tom de *verdade da filmagem*. É justamente nessa negociação entre os interesses do então questionador, que é responsável pelas filmagens, com as pessoas reais, que está a possibilidade de um clima de diálogo espontâneo. Ao ser apresentada a pergunta do questionário, ou seja, a fala que motivou aquela resposta, temos um tom de *filmagem da verdade*, da negociação linguística, nua e crua, com

questionamento e resposta, sem forçar que a situação pareça espontânea, e não um trecho de entrevista.

Conforme THUN (2005, trad. 2017), é impossível pensar no papel do entrevistador sem considerar o paradoxo do observador, ou seja, a ingrata posição de ter que observar o uso da língua, mas só poder fazer isso interferindo no diálogo. Thun questiona as estratégias de Labov para resolver esse conflito: um conjunto de técnicas que vão desde juntar pares de entrevistados até entrevistas curtas que não deem tempo de o entrevistado fazer um autocontrole da sua fala. Também apresenta a resposta de Gilliéron, que utiliza a técnica da fotografia instantânea, ou seja, o entrevistador pergunta e anota de uma vez a resposta, sem dar tempo para interferir ou fazer retoques no registro. Menos pessimistas quanto ao registro da forma mais espontânea são Jaberg e Jud, que, em seu atlas linguístico da Romênia consideram também em seus mapas trechos de fala sobre assuntos exteriores da conversa que foram produzidos entre os entrevistados. Ou seja, se considera falas não sistemáticas para auxiliar na análise dos fenômenos observados por outras ferramentas do questionário do atlas linguístico.

Cap. 3 – Análise

3.1 Estrutura de base para a etiquetagem dos dados

Como já foi dito no cap. 1, as entrevistas gravadas pelo ALMA-H foram etiquetadas seguindo a estrutura do questionário. A fim de viabilizar comparações futuras, seguiu-se no IHLBrI a mesma ordem do ALMA-H, que pressupõe a seguinte estrutura de organização do banco de dados:

Fig. 03: Estrutura do questionário do ALMA-H usada na etiquetagem

Fragebogen / Questionário	
Teil A - Informanten / Informantes	
AI	<i>Identifikation der Informanten / Identificação dos informantes</i>
AII	<i>Identifikation der Eltern / Identificação dos pais</i>
AIII	<i>(Meta)sprachliche Aspekte / Aspectos (meta)lingüísticos</i>
AIV	<i>Topodynamik der Informanten / Topodinâmica dos informantes</i>
Teil B - Informationen zum Ortspunkt / Informações sobre a localidade	
Teil C - Sprache / Língua	
Clex	<i>Lexik / Léxico</i>
Cfon	<i>Phonologie / Fonologia</i>
Cgram	<i>Grammatik / Gramática</i>
Teil Dleit - Lektüren / Leituras	
Teil Tx - Themenauswahl für freie Gespräche und ethnographische Aufnahmen / Seleção de temas para conversas livres e gravações etnográficas	
TEIL C - Lexik / PARTE C - Léxico	
□ Clex: Lexik / Léxico	
Clex01	<i>Der Körper: Gesundheit und Hygiene / O corpo humano: saúde e higiene</i>
Clex02	<i>Kleidung / Vestuário</i>
Clex03	<i>Ernährung / Alimentação</i>
Clex04	<i>Wohnhaus / Habitação</i>
Clex05	<i>Naturphänomene / Fenômenos naturais</i>
Clex06	<i>Landwirtschaft / Atividades agrícolas</i>
Clex07	<i>Pflanzen / Flora</i>
Clex08	<i>Tiere / Fauna</i>
Clex09	<i>Verwandtschaft und Familie / Parentesco e família</i>
Clex10	<i>Soziale Beziehungen: Freunde, Nachbarschaft und Arbeitsplatz / Relações sociais: amizade, vizinhança e local de trabalho</i>
Clex11	<i>Spiele und Feste / Jogos e diversões</i>
Clex12	<i>Quantitäten, Zahlen, Relationen / Quantidades, números, relações</i>
Clex13	<i>Verkehr und Kommunikation / Transporte e comunicação</i>
Clex14	<i>Arbeit, Handel / Trabalho, comércio</i>
Clex15	<i>Religion / Religião</i>

Fonte: Workshop de cartografia do ALMA-H (2017)

A mesma estrutura acima organiza a distribuição dos arquivos em áudio (e em parte também em vídeo) que compõem o banco de dados. Os etnotextos, como se pode observar nesse quadro, entram em uma pasta própria identificada com o prefixo Tx_. A sua ordenação interna, contudo, se dá por meio da identificação interna da **Tx_área + subárea + assunto + dados extralinguísticos**, incluindo principalmente o código do ponto de pesquisa e o perfil social do(s) respectivo(s) falante(s). Essa tarefa implica a oitiva e interpretação do conteúdo do respectivo arquivo de áudio ou vídeo. Para orientar essa tarefa, elaborou-se no âmbito do Projeto a tabela reproduzida no anexo deste TCC, que foi revista e atualizada para a presente discussão. Como se vê, o trabalho de etiquetagem tenta identificar a *área_subárea_assunto* que melhor traduz o conteúdo da entrevista. Não se trata, portanto, de uma opção totalmente livre, mas que exige uma interpretação precisa que atenda tanto ao conteúdo, quanto às buscas de possíveis usuários, mais ou menos na mesma linha das palavras-chave, na apresentação de projetos e resumos de artigos.

Em suma, tem-se assim o seguinte esquema de etiquetagem a ser preenchido conforme a tabela de áreas e temas pré-determinados no guia em anexo:

Tx_Área_(Subárea)_Assunto_in-RS01(_CbGII)

Algumas dificuldades quanto a essa identificação persistem. Uma questão em aberto é por exemplo em que língua escrever as informações. De maneira geral, se busca termos que podem ser compreendidos pela maior quantidade possível de pessoas, principalmente na *área e subárea*. O campo *assunto* nos parece mais suscetível ao uso de palavras também em Hunsrückisch, quando o assunto for em Hunsrückisch.

Outra reflexão pertinente é como identificar quando um arquivo tem dois ou mais etnotextos, ou seja, quando há duplicidade de áreas identificáveis. Podemos recortar o arquivo, ou seja, utilizar uma ferramenta computacional para dividir o arquivo em dois, *ou* fazer a identificação do segmento transliterado na marcação do tempo de transliteração em relação ao arquivo original.

3.2 Informações para o cabeçalho de metadados na transliteração de etnotextos

À medida que as transliterações de entrevistas foram sendo feitas, percebeu-se que nem todas as informações cabiam no nome do arquivo, devido à sua extensão limitada. Assim, surgiu a necessidade de desenvolver um cabeçalho de informações dentro do

documento da transliteração. Sua relevância está em fornecer subsídios que facilitem a análise e estudo de possíveis usuários e pesquisadores, além de abrir caminho para a criação de uma biblioteca virtual desses etnotextos que pode ser usada para inúmeras outras ações de promoção e pesquisa da língua.

É importante ressaltar ganho metodológico por trás do simples fato de mudar a disposição dos metadados do nome do arquivo para uma lista: as linguagens de programação são estruturadas de uma maneira muito semelhante de listas de objetos classificados, como uma tabela. Isso torna possível a elaboração de ferramentas de pesquisa remissivas a cada uma dessas linhas, idêntico à organização de uma biblioteca virtual por exemplo.

Outra das vantagens disso que chamamos de *Cabeçalho de Metadados* é a possibilidade de atualização sem prejuízo da localização do documento, por mudança do nome do arquivo. Para isso, precisamos fazer uma classificação primordial desses metadados: os metadados *estáveis* e os metadados *instáveis*. Considere o exemplo de etnotexto do IHLBrI na fig. 04:

Fig. 04 – Exemplo de etnotexto com o respectivo cabeçalho de metadados

Localidade (grupo social): RS04 Sta. Maria do Herval - Padre Eterno Ilges
Fonte: Acervo IHLBrI
Arquivo Transliterado: Autoriza_Lucia Zimmer & Maria Marlene Schuh & Armando Schuh_in-RS04_Padre Eterno Ilges_IHLBrI_27.09.17
Formato: () áudio (x) vídeo
Data do Arquivo: 27.09.17
E1: Gabriel Schmitt
I1: Armando Schuh
Transliteração: Gabriel Schmitt
Cobertura: parcial (02:00 – 02:30)
Revisão:
Revisão CVA:

[...]

F1: Kennt'ma soohn wie der andre, der hat vom, ich tun bissche woote gell.

E1: Nee, nee.

F1: Is vom Dokter zurick komm, do het die Frau gesooht: "Du woost schon siebzig Jahr alt." Do wer'er traurig, do het der Dokter gesooht: "Ja Mann, warum bist du so traurig?" Ja, het er gesooht, "der Dokter hot gesooht tet nur noch dreizich Tooch lewe". "-Och, du hast so lang gelebt, dann wet die dreizich ooch noch aushalle".

[...]

Por metadados *estáveis* entende-se aqui os dados que têm como fonte as informações extralinguísticas controladas pelo entrevistador/pesquisador logo após a entrevista. Isso inclui, no exemplo acima, a localidade onde a entrevista foi realizada, o projeto ao qual pertence o banco de dados, o nome do arquivo transliterado, seu formato e a data da entrevista. Não se deve esquecer, além disso, do perfil dos informantes, que no ALMA-H diz respeito a um dos quatro grupos entrevistados (CaGI, CaGII, CbGI ou CbGII).

Fig. 05 – Estrutura-base do cabeçalho de metadados no ALMA-H

Acervo de Enotextos ALMA-H	
Localidade: código - Ponto	
Grupo de Informantes:	<input type="checkbox"/> CaGII <input type="checkbox"/> CaGI <input type="checkbox"/> CbGII <input type="checkbox"/> CbGI
Arquivo Transliterado: (p.ex. Tx_Ling_Var_Dt_in-RS06_CbGII_AIII.wav)
Formato:	<input type="checkbox"/> áudio <input type="checkbox"/> vídeo
Data do Arquivo: DD.MM.AAAA	
E1: Entrevistador 1 (nome do entrevistador, se conseguir identificar)	
I1: Informante 1 (nome, se identificável)	
I2: Informante 2 (se houver; se não, apagar)	
I3: Informante 3 (se houver; se não, apagar)	
Transliteração: Nome	
Revisão: Nome	

Vale observar que todas as informações estáveis constam no nome do arquivo mesmo antes da etiquetagem, quando o entrevistador deposita os dados no projeto; o trabalho posterior consiste em acrescentar, após a oitiva dos arquivos, os dados de conteúdo, identificados aqui como instáveis, porque dependem da interpretação e podem ser readequados aos interesses da pesquisa.

Por metadados instáveis entende-se aqui os dados relacionados à transliteração, ou seja, dados feitos após a etiquetagem. Isso inclui tanto a identificação do gênero textual do etnotexto quanto os dados relativos aos participantes da entrevista, transliteradores e revisores, e uma possível marcação do tempo da transliteração em relação ao arquivo original.

Por outro lado, é preciso considerar que em ambos os exemplos temos, por ora, apenas esboços de cabeçalho que ainda não satisfazem plenamente os propósitos do projeto e tampouco otimizam todas as possibilidades de análise desse tipo de dado. Até mesmo a disposição dos dados – não apenas o que deve constar – desempenha um papel. Ela tem de

atender à praticidade e clareza da informação, e não se confundir com o texto transliterado. Neste sentido, vale consultar projetos afins, para verificar outras possibilidades e soluções.

3.3 Identificação de dados de pesquisa nos projetos WossiDiA e REPHO

A fim de comparar a análise e tratamento de dados afins, em projetos de forte base empírica, escolheu-se para este trabalho os projetos WossiDiA (*Wossidlo Digital Archive* – ver www.wossiDia.de), desenvolvido na Universidade de Rostock a partir do Arquivo de Richard Wossidlo, e o projeto REPHO (*Repositório de Entrevistas de História Oral/UFRGS*), desenvolvido com outro propósito, não propriamente linguístico. Esses dois projetos foram escolhidos, neste trabalho, sem grandes pretensões, apenas para testar parâmetros bem diferentes que possam ampliar o campo de visão.

O acervo de dados de Richard Wossidlo, nascido em 1859 em Rostock, no estado de *Mecklenburg-Vorpommern*, Alemanha, foi coletado durante toda a sua vida e engloba crenças, dizeres, adivinhas, receitas, superstições em baixo alemão, registradas em forma manuscrita em cerca de dois milhões de cartões (*Zettel*). Wossidlo foi auxiliado por centenas de colaboradores, que enviavam igualmente suas anotações pelo correio. Não se dispunha de aparelhos para gravação, então todos os registros tinham que ser feitos na hora ou *a posteriori*.

Parte desse material foi publicado em quatro volumes conhecidos como *Mecklenburgische Volksüberlieferungen* (1897–1931). Além disso, juntamente com os esforços de Hermann Teuchert, foi publicado entre 1937 e 1973 o *Mecklenburgisches Wörterbuch*, um dicionário bilíngue de *Mittelniederdeutsch* e *Hochdeutsch* que contém inúmeros exemplos das anotações.

Esse rico acervo é administrado hoje pela Universidade de Rostock, e disponibilizado num portal, denominado WossiDiA. O portal disponibiliza as anotações digitalizadas. Entre elas, há uma parte que foi transcrita à máquina por pesquisadores da antiga DDR. O “cabeçalho” dessas transcrições em fichas é especialmente interessante, tanto na sua forma de apresentação, quanto na seleção dos itens que constam nela. O exemplo a seguir ilustra um desses cartões originais transcritos e adicionados ao *corpus*.

Fig. 06 – Dados de anotações do Arquivo de Wossidlo transcritos

Wossidlo-Archiv Rostock		Archiv-Nr. <u>31</u>	A
Orig.-Nr. <u>B VII 12</u>	Ort der Aufnahme: <u>Gresenhorst</u>	Gattung: <u>Sage</u>	
Aufzeichner: <u>Wossidlo</u>	Kreis: <u>Rostock</u>	Gruppe: <u>Lokale Sage Spuk</u>	
Einsender:	Zeit der Aufnahme: <u>1. 8. 1920</u>	Motiv: <u>Kirche Tiet</u>	
veröffentlicht:	Erzähler: <u>Schäning, Gresenhorst</u>		
Quelle:	Art der Aufzeichnung: <u>Stichworte aufgefüllt</u>		
<p>As dat kirchenhollent inne gang wäst is, möt de köster nachts in de kirch (ubi nescit), hett licht anmaken müsst. as he rinkümmt, steiht up dat een chur 'n ganzen kridewitten hirsch. den köster ward grugen. he geht na 'n paster hen, dat is keenen goden hirsch. von de tiet af an sall dat kirchhollen vörbikamen sin.</p>			

Fonte: <<https://apps.wossidia.de/webapp/run>>, acesso em 26/06/19.

Como mostra o exemplo, a ficha de dados possui uma parte, digamos, estável que apresenta, no centro, a localidade (*Gresenhorst*), distrito (Rostock) e data do registro. O nome do entrevistado/informante (*Erzähler* = narrador) aparece em destaque em uma caixa de texto específica, bem ao centro da ficha. A parte à esquerda é reservada para o número de identificação do cartão original, no Arquivo (no caso, B VII 12) e o nome de quem fez o registro (neste exemplo, o próprio Wossidlo) e de quem enviou (*Einsender*) ou se foi publicado (*veröffentlicht*). Na parte superior à direita, aparece o número no Arquivo. Os dados abaixo dizem respeito a questões sobre o gênero de texto (*Gattung*, no caso uma lenda = *Sage*), ao grupo que se refere (no caso *Spuk*), além do tema (*Motiv*). São informações que remetem ao gênero oral, ao grupo e o tema. A linha inferior dedica-se à indicação da fonte (*Quelle*) e o modo como o texto foi registrado. Cada uma dessas partes aparece em caixas de moldura, a fim de facilitar uma visualização rápida.

Em relação ao segundo projeto analisado, o REPHO foi criado em 2017 pelo Núcleo de Pesquisa em História (NPH) da UFRGS. Conforme o manual de procedimentos do repositório (RODEGHERO, 2017), este se caracteriza como um espaço de captação, guarda e divulgação de entrevistas realizadas pelos docentes e discentes em História, da

UFRGS, a fim de que o acesso às entrevistas transcritas sirva de subsídio à formação dos próprios pesquisadores de História, na UFRGS e em outras instituições. Para isso, o manual “define os pré-requisitos mínimos que os materiais recebidos devem respeitar, de modo a produzir um *corpus* de entrevistas que seja apto à divulgação e ao uso por diferentes pesquisadores(as) ao longo do tempo” (RODEGHERO, 2017, p. 2).

É interessante destacar que o termo de autorização do REPHO inclui uma cláusula sobre a divulgação de dados que identifiquem o entrevistado. Especialmente por entrevistar pessoas que fazem denúncias graves, como tortura, torna-se relevante manter oculta a identidade dos participantes.

No portal do REPHO, ao selecionar uma entrevista, é necessário fazer *download* do documento para visualizá-lo. O documento é organizado da seguinte maneira: indicação do Departamento onde foi feita a entrevista e reunido o *corpus*; definição de uma etiqueta com o nome do entrevistado e do entrevistador, bem como indicação da circunstância da entrevista; uma ficha técnica (Fig. 07), um sumário, equivalente a um resumo da entrevista, e, por fim, a transcrição.

Tratando-se de pesquisa na área de história, é natural que a língua acabe sendo vista mais como um veículo de transmissão de ideias. Há alguma preocupação em marcar hesitação, ou emoções como raiva e riso. Isso leva muitas vezes à decisão de utilizar uma língua-padrão para a transliteração das entrevistas. No REPHO não há uma marcação específica do gênero textual, sendo todos os textos classificados como *entrevista*.

Fig. 07 – Ficha Técnica de uma entrevista do REPHO

FICHA TÉCNICA

Projeto: Ingresso e permanência de estudantes cotistas na UFRGS
Coordenador/a: Profª. Carla Simone Rodeghero (UFRGS)
Entrevistado/a (nome e telefone): Diogo Santos Gomes, [REDACTED]
Data de nascimento: 18/01/2000
Local e data da entrevista: Rua Ramiro Barcelos, 2705/Campus Saúde, Fabico – UFRGS -, Bairro Santana, Porto Alegre, RS. Data, 29/10/2018
Pesquisa e montagem do roteiro: Gabriela Machado Leindecker
Entrevistadora: Gabriela Machado Leindecker
Transcrição: Gabriela Machado Leindecker.
Revisão (nova conferência entre áudio e transcrição): Gabriela Machado Leindecker.
Tempo de gravação: 34 minutos e 1 segundo
Número de páginas digitadas: 13
Sumário: Apresentação de Diogo Gomes; infância, quando entrou em escola pública e sua vida escolar em Charqueadas; dificuldades no ensino fundamental; troca de escolas no ensino

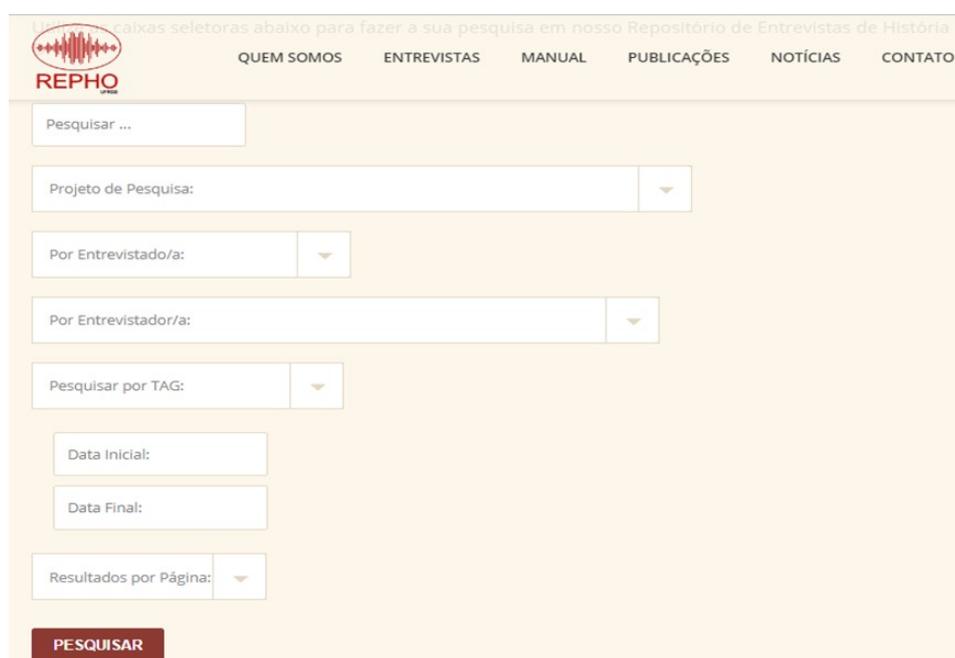
Fonte: REPHO –

ver <<https://www.ufrgs.br/repho>>. Acesso em: 03/07/2019.

Diferente de outros exemplos analisados até aqui, chama a atenção o grau de detalhamento dos dados dos participantes envolvidos. Evidentemente, tal só é possível com o consentimento livre e esclarecido dos entrevistados. O mesmo vale para a elaboração do roteiro da pesquisa, que inclui além disso um sumário informando os tópicos da entrevista.

Com a ficha técnica e a respectiva disposição dos dados, é abastecida a interface do repositório. Essa interface contém seis opções de pesquisa: um campo para escrever uma palavra e pesquisar; pesquisa por entrevistado e por entrevistador; uma pesquisa por TAG, equivalente a uma pesquisa por temas; delimitação de data.

Fig. 08 – Interface *online* do REPHO



Fonte: REPHO – <<https://www.ufrgs.br/repho/pesquisar-no-acervo/>>. Acesso em: 03/07/2019.

Vemos por meio da organização do REPHO a tentativa de disponibilizar formas de busca no repositório. Também na gênese das entrevistas, a ficha serve de base para a realimentação do banco de dados com novas entrevistas. Ambos os projetos, como se vê se diferenciam não apenas pelo tipo de dado e objetivo da pesquisa, mas também pela forma de captação do dado. Seguimos com a análise do “cabeçalho de metadados” do ALMA-H, para a transliteração de etnotextos, levando agora em consideração alguns *insights* novos.

3.4 A questão do tipo de etnotexto: entre o gênero textual e o campo temático

A tabela em anexo que serve de orientação para a identificação dos etnotextos do ALMA-H levanta a questão do tipo de texto. Na coluna referente à área, vemos ora como critério um campo temático (ou campo semântico) – por exemplo, Bot_ e Zoo_ - ora uma indicação que sugere determinado gênero textual – por exemplo, Rezept_, Humor_. Essa mescla de critérios nos parece compatível com os propósitos do Projeto, uma vez que combina léxico e cultura (induzidos pelo campo temático) com probabilidade de ocorrência de determinadas estruturas gramaticais (o que é motivado pelo tipo de texto ou gênero textual). Por exemplo, um etnotexto identificado como “Tx_Edu_Esc_Prof_antig_früher_in-” sinalizará para um léxico e aspectos culturais do funcionamento da escola no passado (p.ex. palavras como *Schul*, *Lehrin*, *Lehrer*, *lenne* são pressupostas) e, ao mesmo tempo, permitir estudos por exemplo do uso do pretérito e do *Perfekt*, já que a narrativa se dá no tempo verbal passado.

Como foi discutido na seção. 2.3, de acordo com Bakhtin é muito importante a busca pela definição dos gêneros textuais. As fronteiras entre os gêneros, mesmo que possam ser evidentes na teoria, muitas vezes se misturam na prática.

Cabe retomar, aqui, os elementos fundamentais de um gênero conforme Bakhtin: o estilo, a forma composicional e o conteúdo temático. Esses elementos sugerem, em outras palavras, que sejam adicionados ao cabeçalho de metadados do ALMA-H os seguintes campos: Tipo, Área e Assunto.

Os *tipos* de etnotextos estão previstos nos próprios questionários ALMA-H e IHLBrI, questão discutida no cap. 1. Para a determinação da *área* e dos *assuntos*, o Guia de Etiquetagem do ALMA-H (em anexo) serve, conforme já se aludiu, de orientação; em parte, os *tipos* de etnotexto integram o próprio nome do arquivo, pois “os dados estão etiquetados e classificados por temáticas (por exemplo, receitas, histórias, versos, descrições etc.)” (RAVNIKAR, 2014).

É importante ressaltar que também estão previstos dados metalinguísticos nas entrevistas. Inclusive pela tabela de dimensões de análise do ALMA-H, há uma distinção entre os segmentos de entrevista que fazem uso *da* língua e os que falam *sobre* a língua.

Uma solução possível seria identificar *metalinguística* no item *área*, do cabeçalho. Sugerem-se então os seguintes metadados:

Tipo de entrevista: (Referente à parte do Questionário do ALMA-H ou IHLBrI)
Área: (Mais amplo. Metalinguística ou áreas - Cf. Guia de Etiquetagem em anexo)
Assunto: (Mais restrito. Pode se basear no Guia de Etiquetagem)

Uma implicação do tipo de etnotexto e da estrutura que compõe um determinado gênero textual é a sua unidade temática e estrutural. As entrevistas foram muitas vezes feitas em um contínuo, ou seja, sem interrupção. Por este motivo, um arquivo de áudio comporta muitas vezes mais de um etnotexto enquanto unidade textual. Há, neste caso, a opção de recortar o texto, ou melhor, segmentar o arquivo de áudio em mais de uma unidade textual. De certo modo, o pesquisador faz isso em campo, ao fixar os chamados *tracks*, na função salvar, iniciar gravação de novo arquivo de áudio/vídeo. Definir recortes significa que é preciso tomar uma decisão: ou se recorta o arquivo original, de áudio ou vídeo, e assim se cria um arquivo novo; ou se faz a marcação do trecho transliterado diretamente no arquivo original. Esta última opção contextualiza o segmento de etnotexto na sua relação com o que é dito antes e depois. É uma solução mais complexa que envolve algumas decisões no cabeçalho.

Conforme apresentado no cap. 3.1, o cabeçalho do IHLBrI registra:

Segmento (transliterado): *total* ou *parcial* (02:00 – 02:30)
Duração do arquivo: 00:00:00

Parcial seria referente a um recorte da transliteração em relação ao arquivo como todo; a outra opção seria *total*, quando o recorte da transliteração é toda a extensão do arquivo. Fundamental também é marcar o tempo total do arquivo original da entrevista.

São dois apontamentos técnicos para auxiliar a percepção do pesquisador, transliterador e revisor que vai se debruçar sobre esses etnotextos, permitindo identificar segmentos de entrevista significativos.

3.5 Como identificar os participantes da entrevista?

Claramente, há dois tipos de participantes envolvidos nas entrevistas: de um lado, os entrevistados e, de outro, os entrevistadores. Diferentes são também as dificuldades em fazer a identificação de ambos, fundamentais em uma entrevista. Essa identificação é feita, no cabeçalho, com os nomes e o perfil social de cada participante.

Conforme já foi abordado neste trabalho, há diferentes formas de se referir a quem formula as perguntas e conduz a entrevista. Eduardo Coutinho (1997) fala em *questionador*, Georg Wenker em *explorador* (cf. THUN, 2000). Podemos pensar também no termo tradicional da dialetologia: *inquiridor*. Ou então sugerir uma forma ampla como *pesquisador*. Contudo, esses quatro termos são carregados de significados que os tornariam imprecisos. Uma opção ao mesmo tempo neutra e que não leva a outras interpretações parece ser *entrevistador*, que já vem sendo utilizado de forma recorrente no Projeto.

Sobre os entrevistadores parece não haver nenhuma imposição quanto à publicação de nomes. Parece aliás uma possibilidade de análise buscar tendências nas entrevistas de cada entrevistador. Essa pode ser uma solução para o paradoxo do observador: não esconder quem fez a entrevista, mas sim deixar bem claro os papéis de cada entrevistador.

A questão da identificação dos entrevistados é mais delicada, em função de questões éticas. O próprio termo *entrevistado* vem sendo evitado pelo ALMA-H, para distinguir a abreviatura de E em *entrevistador*. Tem-se optado, por isso, pelo termo *informante*, que porém tem sofrido algumas críticas, por sugerir um depoimento, uma delação, além de passar a ideia de que o entrevistado seria visto apenas quanto a sua capacidade de fornecer informação. Por esta razão, tem-se inclinado ao uso da designação *falante*, acentuando assim o papel do entrevistado como membro da comunidade de fala. Neste caso, teríamos as siglas E (entrevistador) e F (falante).

Em relação à publicação do nome dos entrevistados, ou falantes do projeto ALMA e do IHLBrI, todos os falantes gravados deram seu consentimento para uso dos dados para fins de pesquisa. Pode-se entender que, por grande parte dos entrevistados serem pessoas com pouco acesso ao ensino e ao mundo letrado, a publicação de seus nomes e de seus depoimentos por uma pesquisa científica de tamanha envergadura, a identidade dos entrevistados possa ser consolidada. Pode ser visto como positivo para a universidade, enfim, legitimar a língua minoritária, muitas vezes oculta por não possuir voz. Além disso, os nomes e sobrenomes somados aos dados das localidades podem auxiliar outras pesquisas futuras,

especialmente no campo da antroponímia e da genealogia.

Uma opção seria a criação de uma legenda para cada entrevistado, contendo informações relevantes sobre o perfil social do falante F: sexo ou gênero, idade e grau de escolaridade, além de um número para diferenciar mais de um participante, em caso de pluralidade de informantes. Uma sugestão seria *FIm*, referindo-se ao Sexo_geração_escolaridade_número de identificação.

Caso essa seja a opção escolhida, faz-se a seguinte sugestão:

F1: *FIm*

E1: (Nome do entrevistador)

A decisão entre anonimizar ou nomear os falantes não se restringe a uma simples questão de ética. Na perspectiva dos falantes, como seria a autoria de um conto ou relato histórico, a nomeação do autor da fala pode significar muito. Isso é perceptível em fórmulas narrativas do tipo “meu avô dizia”, ou “como dizia o Welter João”, entre outras asserções comuns. Também a recuperação futura de dados e falas na perspectiva familiar e de herança cultural não apenas confere ao banco de dados de uma língua de imigração como o Hunsrückisch uma função de acervo de pesquisa, mas também reflete seu papel como verdadeiro repositório da memória. Para fins de pesquisa, o falante representa um CbGII ou outro perfil social semelhante e se justifica seu anonimato; para os fins da memória cultural, parece sumamente importante nomear CbGII com seu nome próprio, com o qual participou da história da língua. Curiosamente, tanto Wossidlo em suas anotações, quanto o projeto REPHO nomearam seus informantes, ou aqueles de quem provêm os dados.

3.6 Metadados relativos ao processo de transliteração dos etnotextos

A oitiva de dados de fala requer habilidades especiais que precisam ser reconhecidas, tanto mais quando se trata de uma língua minoritária com marcas próprias. Felizmente, há no projeto participantes que têm o Hunsrückisch como língua materna, ou ao menos como segunda língua. Isso precisa ser levado em conta, mas não dispensa a relevância de diferentes ouvidos no grupo de pesquisa. Por esta razão, tal como já se vem

praticando no ALMA-H, cabe incluir no cabeçalho de metadados as informações sobre a coordenação, o trabalho de transliteração e revisão:

Transliterador:

Revisor:

Revisão Coord.:

Do ouvido e conhecimento do transliterador e dos diferentes revisores vai depender a correção e a exatidão do que foi transliterado. Por isso, a relevância desse controle. O repositório REPHO faz ainda uma identificação adicional: apresenta uma opção de revisão com reaudição do áudio original. Para os fins do projeto ALMA-H, parece redundante adicionar essa informação.

Justifica-se, portanto, o registro de quem participou do processo de transliteração dos etnotextos, para também ser possível a análise de uma eventual tendência entre os próprios transliteradores. De certo modo, poderíamos estender o paradoxo do observador para um *paradoxo do transliterador*: quando se escreve o que alguém falou, como não interferir no registro, se está se recriando uma fala? Como já referido acima, todos os transliteradores orientam-se pelo ESCRITHU, mas eventuais tendências estilísticas de cada transliterador não são possíveis de serem totalmente controladas.

Parece também ser muito importante para análises futuras a identificação mais precisa quanto à identificação da situação da gravação. Sendo a maioria dos arquivos do ALMA-H em áudio, gravados normalmente sentados ao redor do microfone, por vezes acompanhados de um chimarrão, essa situação pode não ser tão reveladora. Mas muitos dados do IHLBrI e uma parte dos dados do ALMA-H são em vídeo e englobam uma gama maior de situações de entrevista: por exemplo, na frente da igreja, pela janela do carro, numa plantação de fumo, etc.

Vimos que o etnotexto do WossiDiA não apresenta uma marcação específica sobre isso. O REPHO identifica o local da entrevista, mas não comenta fatores externos. Ficaria o questionamento de como fazer essa identificação das condições da própria entrevista no âmbito do ALMA-H.

3.7 Proposta de “cabeçalho de metadados”

Feita a análise dos diferentes aspectos envolvidos tanto na gravação de etnotextos quanto na sua transliteração para usos variados em pesquisas e ações educativas, retomamos sumariamente quais as informações que já estão no cabeçalho juntamente com as sugestões deste trabalho:

Fig 09 – Relação dos metadados identificados nos etnotextos analisados

Metadados identificados no ALMA-H/IHLBrI-	Metadados sugeridos adicionalmente
Localidade: código - Ponto Grupo de Informantes: <input type="checkbox"/> CaGII <input type="checkbox"/> CaGI <input type="checkbox"/> CbGII <input type="checkbox"/> CbGI Arquivo Transliterado: Formato: <input type="checkbox"/> áudio <input type="checkbox"/> vídeo Data do Arquivo: DD.MM.AAAA E1: Entrevistador 1 (nome) F1m/f: Falante 1 (nome) F2m/f: Falante 2 (nome) Transliteração: Segmento: Revisão: Revisão Coord.:	Tipo: Área: Assunto: Comprimento do Arquivo:

Em relação à pergunta sobre a forma de organização dos metadados para que possam ser utilizados de maneira mais eficaz por potenciais usuários do banco de dados, parece que a disposição e formatação do cabeçalho de metadados também desempenha um papel. As informações organizadas em modelo de lista representam um avanço metodológico em relação aos dados na etiqueta do arquivo, justamente pela facilidade que isso agrega. Porém, talvez seja possível ainda otimizar essa organização, a fim de ganhar em praticidade. A seguinte proposta tenta contemplar as diferentes opções tidas como mais pertinentes e eficazes para a organização e identificação dos metadados do etnotexto transliterado:

Fig. 10 – Sugestão para cabeçalho de metadados do ALMA-H



Arquivo Transliterado:			Fonte:
Formato: () áudio () vídeo Data do Arquivo: DD/MM/AAAA Duração do Arquivo: 00:00:00	Localidade: F1: F2:	Grupo social:	Tipo: Área: Assunto:
Segmento: (total ou parcial) 00:00:00 - 00:00:00 Transliteração: Revisão:	E1: E2:	Revisão Coord.: Publicação:	

Com esse modelo, são adicionadas novos metadados ao cabeçalho, conforme sugestões desse trabalho. O segmento transliterado e a duração do arquivo foram organizadas na parte inferior da esquerda. Foram adicionadas metadados referentes à identificação do tipo, área e assunto do etnotexto à direita. No canto inferior da esquerda foram adicionadas duas informações referentes à revisão da Coordenação e um espaço para informações sobre a publicação desse etnotexto.

Além de adicionar metadados novos, essa organização dos metadados em um modelo de tabela facilita a visualização e a leitura do etnotexto, além de ocupar menos espaço do arquivo de texto. Fechamos assim as sugestões desse trabalho.

Palavras finais

Esse trabalho se propõe a sugerir fundamentos para a organização de etnotextos, com aplicação no banco de dados do ALMA-H. Algumas sugestões foram feitas, mas ainda há questões que estão em aberto, principalmente em relação à padronização da língua da etiquetagem dos arquivos. Apesar desse ser um dos objetivos específicos, essa questão permanece em aberto.

Outros objetivos específicos foram bem explorados e foram feitas algumas sugestões pertinentes ao projeto. Em relação aos gêneros orais, decidiu-se priorizar as identificações de tipo de texto, sugeridos pelo Guia de Etiquetagem. O tema e o assunto dos etnotextos também receberam um lugar no cabeçalho.

E por que é relevante a produção e a organização de etnotextos em Hunsrückisch? Pois são narrativas orais de uma língua marginalizada da América do Sul, e podem servir para a construção da identidade para cerca de 1,25 milhão de sul-americanos que, conforme Altenhofen, Prediger e Habel (2018), falam Hunsrückisch hoje. É também de Altenhofen a metáfora dos etnotextos em Hunsrückisch como espelhos de papel, ou seja, o reconhecimento no escrito de que a língua, e portanto, a cultura que vem dessa língua tem a capacidade e mais, o direito de ser registrada em escrito. Não esquecemos o poder da palavra escrita num país de maioria analfabeta funcional.

Foi muito produtivo analisar a organização dos etnotextos de outros projetos. Com o REPHO, percebeu-se que mesmo uma interface simples resolve várias questões referentes a identificação e pesquisa dos metadados, oferecendo ao público em geral um acervo acessível e padronizado de entrevistas transcritas. Com o WossiDiA, percebeu-se a possibilidade de organização dos metadados não mais em lista, porém em tabela. Também a identificação dos gêneros orais desse projeto foi relevante.

Por fim, esse trabalho fecha meu ciclo de Graduação em Letras com um trabalho que parte da prática de transliterar e busca a teoria para tomar decisões técnicas. Desde 2015, quando entrei no Projeto ALMA-H como bolsista de iniciação científica, me envolvi, entre outras tarefas, com os etnotextos em Hunsrückisch. Vi a possibilidade de, com etnotextos de uma língua, poder-se colocar enfim na frente dos olhos dos críticos, estudiosos e dos próprios falantes, as variedades de alemão falado, que muitas vezes é vista como um *atrapalho* no aprendizado de português ou de alemão-standard. Não podemos

ignorar que no Brasil se falam muitas variedades de alemão, e às vezes é tarefa difícil e ingrata classificar e rotular ‘isso é Hunsrückisch’, ‘isso não é’.

De maneira geral, espera-se que esse trabalho possa, humildemente, contribuir para as pesquisas em Narrativas ou História Oral, principalmente em relação à identificação de elementos extralinguísticos, que podem interessar pesquisadores de diferentes áreas do conhecimento. Espero continuar o trabalho com etnotextos no meu mestrado.

Referências bibliográficas

ALTENHOFEN, Cléo V.; *A constituição do corpus para um “atlas linguístico-contatual das minorias alemãs na bacia do prata”*. In: Martius-Staden – Jahrbuch, São Paulo, n.51, p. 135-165, 2004.

ALTENHOFEN, Cléo V.; FREY, Jaqueline; KÄFER, Maria Lidiani; KLASSMANN, Mário; NEUMANN, Gerson R.; SPINASSE, Karen Pupp. *Fundamentos para uma escrita do Hunsrückisch falado no Brasil*. In: Revista Contingentia, v. 2 (nov.), p. 73-87, Porto Alegre, 2007.

Disponível em: <https://seer.ufrgs.br/contingentia/article/view/3867/2166>.

ALTENHOFEN, Cléo V.; THUN, Harald. *As migrações e os contatos linguísticos na geografia linguística do sul do Brasil e Bacia do Prata*. In: AGUILERA, Vanderci & ROMANO, Valter (eds.). *A geolinguística no Brasil: caminhos e perspectivas*. Londrina: Ed. UEL, 2016.

ALTENHOFEN, Cléo V.; *Plurilinguismo na escola e na sociedade em uma perspectiva macrolinguística*. Resenha de THUN, Harald et al. *Atlas Lingüístico Guaraní-Románico: Sociología (ALGR-S)*, tomo I: Comentários; tomo II: Mapas. Kiel: Westensee-Verlag, 2002. In: Organon, v. 32, n. 62, p. 311-317, jan./jul. [2017].

Disponível em: <http://seer.ufrgs.br/index.php/organon/article/view/74423>.

ALTENHOFEN, Cléo V.; *Workshop: etiquetagem de dados do IHLBrI*. Porto Alegre, 2017. (Não publicado)

ALTENHOFEN, Cléo V.; *Plurilinguismo na escola e na sociedade em uma perspectiva macrolinguística*. In: Organon, Porto Alegre, v. 32, n. 62, p. 311-317, jan/jul. 2017.

ALTENHOFEN, Cléo V.; HABEL, Jussara M.; PREDIGER, Angélica. *A escrita do Hunsrückisch*. In: ALTENHOFEN, Cléo V.; HABEL, Jussara M.; NEUMANN, Gerson R.; Prediger, Angélica (orgs.). *Hunsrückisch em prosa e verso*. Porto Alegre: Instituto de Letras, UFRGS, 2018. p. 23-34. Disponível em: <http://hdl.handle.net/10183/184118>.

ALTENHOFEN, Cléo V.; MORELLO, Rosângela; BERGMANN, Gerônimo L.; GODOI, Tamissa G.; HABEL, Jussara M.; KOHL, Sofia F.; PREDIGER, Angélica; SCHMITT, Gabriel; SEIFFERT, Ana Paula; SOUZA, Luana C.; WINCKELMANN, Ana C. *Hunsrückisch: inventário de uma língua do Brasil*. Florianópolis: Garapuvu, 2018. 248 p. Disponível em: <http://hdl.handle.net/10183/194384>.

ANSCHAU, Fábio. *Bases teórico-metodológicas para a elaboração de um dicionário do Hunsrückisch falado no Brasil*. Dissertação (Mestrado). Porto Alegre: Universidade

Federal do Rio Grande do Sul (UFRGS); Programa de Pós-Graduação em Letras, 2010. 124 p. Disponível em: <http://hdl.handle.net/10183/117599>.

BAKHTIN, Mikhail. *Os gêneros do discurso*. Trad. de Paulo Bezerra; São Paulo: Editora 34, 2016.

BROCH, Ingrid Kuchenbecker. *Ações de promoção da pluralidade linguística em contextos escolares*. Tese (Doutorado). Porto Alegre: Universidade Federal do Rio Grande do Sul (UFRGS); Programa de Pós-Graduação em Letras, 2014. 268 p. Disponível em: <http://hdl.handle.net/10183/102190>.

COUTINHO, Eduardo. *O cinema documentário e a escuta sensível da alteridade*. Proj. História, São Paulo, 15 de abril de 1997.

IPHAN - INSTITUTO DO PATRIMÔNIO HISTÓRICO E ARTÍSTICO NACIONAL. *Guia de Pesquisa e Documentação para o INDL - Inventário Nacional da Diversidade Linguística. Vol. 1: Patrimônio cultural e diversidade linguística*. Brasília: IPHAN, 2014a. Disponível em: http://issuu.com/designcasa8/docs/indl_guia_vol.1_21.

IPHAN - INSTITUTO DO PATRIMÔNIO HISTÓRICO E ARTÍSTICO NACIONAL. *Guia de Pesquisa e Documentação para o INDL - Inventário Nacional da Diversidade Linguística. Vol. 2: Formulário e roteiro de pesquisa*. Brasília: IPHAN, 2014b. Disponível em: http://issuu.com/designcasa8/docs/indl_guia_vol_2_28.

MANZINI, Eduardo José. *Uso da entrevista em dissertações e teses produzidas em um programa de pós-graduação em educação*. In: Revista Percurso - NEMO, v. 4, n. 2, p. 149-171, Maringá, 2012.

PELEN, Jean-Noël. *Mémoire de la littérature orale. La dynamique discursive de la littérature orale: réflexions sur la notion d'ethnotexte*. Recontres Internationales, Actes: Paris, 1991.

PELEN, Jean-Noël. *Memória da Literatura Oral – A dinâmica discursiva da Literatura Oral: reflexões sobre a noção de Etnotextos*. Tradução: Maria Thereza Sampaio. São Paulo: Proj. História. 22 de junho de 2001. [1986]

RAVNIKAR, Lilian Isabel Marie. *“Espelhos de papel”: produção de etnotextos em hunsriqueano como língua brasileira de imigração*. Porto Alegre, 2014. Disponível em: <<https://lume.ufrgs.br/handle/10183/114716>>

RODEGHERO, Carla Simone. *Manual de procedimentos*. Porto alegre, Laboratório de estudos sobre os usos políticos do passado – LUPPA, UFRGS, 2017.

SAPIR, Edward. *Language: an introduction to the Study of Speech*. New York: Hartcourt Brace, 1921

THUN, Harald. *La geolingüística como lingüística variacional general (com ejemplos del Atlas lingüístico Diatópico y Diastrático del Uruguay)*. In: International congress of romance linguistics and philology. Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza. Org. RUFFINO, Giovanni. Tübingen: Niemeyer, v. 5, p. 701-729, 1998. [1995]

THUN, Harald. *O velho e o novo na geolingüística*. Trad. Claudia Fernanda Pavan / Gabriel Schmitt / Eduardo Nunes / Viktorya Zalewski dos Santos. In: Cadernos de Tradução, Porto Alegre, n.40, p. 59-81, jan/jun 2017. [2000]
Disponível em: <https://seer.ufrgs.br/cadernosdetraducao/article/view/87208/50004>.

THUN, Harald. *Variação na interação entre informante e entrevistador*. Trad. Cléo V. Altenhofen / Filipe Neckel. In: Cadernos de Tradução, Porto Alegre, n.40, p. 82-107, jan/jun 2017. [2005]
Disponível em: <<https://seer.ufrgs.br/cadernosdetraducao/article/view/87180/50001>>.

WINCKELMANN, Ana C. *Percepção da norma oral e escrita na variação das consoantes /pf/, /f/ e /p/ no Hunsrückisch*. Porto Alegre: UFRGS, 2018.
Disponível em: <<https://lume.ufrgs.br/handle/10183/191142>>.

Outras Referências

VIVER no Brasil falando Hunsrückisch. Direção de Gabriel Schmitt e Ana C. Winckelmann. Porto Alegre: IHLBrI, 2018. DVD. (37 min.).

ALTENHOFEN, C.V. *O que é o Projeto ALMA?* Disponível em: <<https://www.ufrgs.br/projalma>>
Acesso em 15/07/2019.

REPOSITÓRIO DE ENTREVISTAS ORAIS/UFRGS. Núcleo de Pesquisa em História.
Disponível em: <https://www.ufrgs.br/repho/>

WOSSIDIA DIGITAL ARCHIV. Disponível em: <<https://www.WossiDiA.de>>.

Anexo

Guia para Etiquetagem de Etnotextos do ALMA-H

Coord. C.V. Altenhofen

Tx-hr_Área_(Subárea)_Assunto_in-RS01(_CbGII)

Tx-pt_Área_(Subárea)_Assunto_in-RS01(_CbGII)

	Área (Hauptgebiet) obligatorisch	Subárea (Unterteilung) fakultativ	Assunto (Thema) (cf. interpretação / oitiva) EXEMPLOS:	_in- Punkt_Ort_Inf (fonte=jornal etc.)
Tx_	Agri_	Kolon_ Plant_ Instr_ Fab_ Vieh_Anim_	Futter_Pasto_ Land_Terr_ Kolonie_ Plantoosch_ Pflanzmaschine_Ochsen_ Arado_Pflug_ Schlachten_Carn_ Zackern_Lavr_ Gewehr_Arma_ Schnaps_Cachaça_ Kees_Queijo_ Criolo_Palheiro_ Fum_Fumo_ Honig_Mel	z.B. _in- RS07_Harmonia _CbGII _in- Jornal_data_p.
Tx_	Alim_	Refeiç_Mahlzeit_ Utenc_ Coz_ Prod_	Siess_ Kaffee_ Frühstück_ Mittagessen_Almoç_ Geschirr_ Kochen_Coz_ Ofen_Forn_ Backofen_Forn_ Churrasco_ Galinhada_ Kerbessen_	
Tx_	Art_	Pint_ Escult_ Artesan_	Malen_ Theater_ Stricke_Croché_ Heckeln_Tricot_ Wandschoner_	
Tx_	Biogr_ (ver Memo)	(nome)		
Tx_	Biolg_	Part_Corp_ Krankh_Doenc_	Ros_Erisipela_	
Tx_	Bot_	Plant_Med_ Plant_Agri_	Bettleis_Picão_ Käsebaum_	

		Baum_Arv_ Palm_ Unkraut_Mat_ Blum_Flor_ Obst_Frut_ Holz_Lenha_ Wald_Mat_	Arrude_ Maniok_ Krankh_Doenc_	
Tx_	Cult_	Patrim_ Mat_ Imat_ Culin_ Costume_	Museum_ Saal_ Archiv_ Schild_Placa_ Denkmal_Monum_ Igrej_	
Tx_	Econ_	Com_Handel_ Geld_Moed_ Trab_Verdienen_	Einkauf_Compr_ Geld_Dinh_ Musterreiter_Caix_Viaj_ Vende_ Loj_Geschäft_	
Tx_	Edu_	Esc_Schul_ Prof_Lehrer_ Apred_Lern_ Livr_Did_Lehrbuch _ Schulweg_Esc_ Material_Esc_	früher_antig_ (palavra-chave)_	
Tx_	Escrit_ Escrit_Pt Escrit_Hr_ Escrit_Hdt_	Les_Leit_ Cart_Brief_ Imprens_ Escr_Schreib_ Gótica_ Lern_Aprend_	Alman_Kalender_ Brief_Cart_ Buch_Livr_ Zeitung_Jorn_ Jorn_Colun_Hr_ Lesen_Leit_	
Tx_	Etn_	Ale_ Deuschländer_ Br_ Gauch_ Ita_Gringo_ Polon_ Parag_ Ind_Buger_ Negro_	Ling_ Ling_Dt_ Trab_ Cult_ Fest_ Prat_Cult_ Valor_Werte_ Cas_Heirat_ Haus_Casa_ Pom_ Westf_ Boehm_ Relig_ Culin_Gastron_	
Tx_	Fam_	Herkunft_Orig_ Fam_Name_ Parent_ Trab_		

Tx_	Fest_		Kerb_ Oktoberfest_ Ostern_Pasc_ Weihnacht_Natal_ Geburtstag_Aniv_	
Tx_	Futur_	Plan_ Hipotese_ Traum_Sonh_ Wunsch_Desej_		
Tx_	Geo_	Toponymik_ Fluss_Rio_ Berg_Morro_ Terr_Boden_ Reg_ Front_Grenz_	(Ortsname)_ (Region)_ Placa_Schild_	
Tx_	Hab_	Haus_Casa_ Constr_	Miljehitt_Paiol_ Wohnhaus_Morad_ Küche_Coz_ Sitzmöbel_Moveel_ Backofen_Forn_ Dach_Telh_ Área_ Schloss_Fechad_ Fenster_Jan_ Churrasqueira_	
Tx_	Hist_ (fatos históricos oficiais, coletivos, de outros) (fatos na 3ª pessoa)	Migr_ Emigr_ Loc_ Kolon_ Fam_ Pers_ Biogr_ Narrativ_	Anfang_Inic_ Hunger_Fom_ Luís Buger_ Mucker_ PY_Krieg_Guerr_ Rev_ Seg_Guerr_Weltkrieg_	
Tx_	Humor_	Ling_ Relig_ Fam_ Esc_ Sex_ Polit_	dat-wat_ tudo pom Frida_ (palavra-chave)_	
Tx_	Imprens_	Jorn_ Livr_Buch_ Alman_Kalender_ Revista_ Leit_Les_	Bleetch_ St. Paulus-Blatt_ Familiekalender_ Gesangbuch_Livr_Cant_	
Tx_	Ling_	Biogr_Aprend_ Mater_Mutterspr_ Edu_Schule_ Verbot-Dt_Proib_ Vorurteil_Discrim_ Num_Falantes_ Cont_Var_	(palavra-chave)_ hinnerst vedderst_ Plattdeutsch_ über Guarani_ Westf_ Pom_ Pt_Br (Region)_	

		Fraseolog_ Express_ Lern_Aprend_ Fam_ Var_Dt_ Var_Pt_ Schule_Esc_ Spruch_Dit		
Tx_	Lit_	Ditado_ Vogel_Ave_Cant_ Poem_ Vers_ Hist_ Conto_ Cont_Fad_Märchen_ Livr_Buch_ Narrativa_	Mund_Maul_Schniss_ Gibt Reen_	
Tx_	Med_	Hospital_ Dokter_ Hebamme_Parteira_ Krankh_Doenc_ Medik_Remed_ Curat_ Pop_Volksmed_ Braucher_Simp_ Zahnarzt_Dent_ Teedokter_Benz_ Knochenflicker_	Olina_ Mainzer Tropfen_ Grippe_ Warze_Verruga_ ...	
Tx_	Memo_ (memória do passado vivido pelo informante) (Memórias em 1ª. pessoa)	Migr_ Biogr_ Fam_ Edu_ Milit_	1. Radio_ Erziehung_Edu_ Herkunft_Orig_ Heirat_Casam_ Hunger_Fom_ Serv_Quartel_ Kind_Inf_ Namor_ Lern_Ling_Aprend_	
Tx_	Mus_	Instr_ Pop_Volkslied_ Dt_Tradic_ Dt_ Singen_Cant_ Chor_ Kirch_Igrej_ Tanz_Danç_ Ball_ Bandinha_ Gauch_	früher_antig_ Mundharm_ Gaita_ Violão_ Marie unn die Greth_	
Tx_	Natur_	Klima_ Wetter_Temp_ Land_Terr_ Aqua_	Regenzeichen_Sin_Chuv_ Trockenheit_Seca_ Blitz_Raio_ Winter_Inv_	

		Wald_Mato_	Tip_Solo_	
Tx_	Polit_	Loc_ Ling_ Kandidat_ Pref_ Vereador_ Gov_ Wahl_Eleic_	Comício_	
Tx_	Relig_	Cat_ Evang_ Priester_Padre_ Kirche_Igrej_ Gebet_Rez_ Lied_Cant_ Missa_Gottesdienst_ Cult_Gottesdienst_ Taufe_Batism_ Kommun_ Fest_ Tod_Mort_ Beerdig_Enterr_ Friedhf_Cem_ Buch_Livr_ Gott_Deus_ Sant_ Weihnacht_Natal_ Ostern_Pasc_	Bibel_ Glocke_läuten_Sin_Toc_ Bau_Constr_ Predigt_Serm_ Vaterunser_ Seminar_ (nome_santo)_ Confiss_ Cas_Heirat_ Fest_ Kerb_ Rosekranz_ Grab_Sepult_	
Tx_	Rezept_ Prod_	Culin_Dt_ Culin_Gauch_	Tee_ Chimarrão_ Churrasco_ Kässhmier_ Kuchen_Cuca_ Schmier_ Weihnachtsdoss_Natal_ Criole_Palh_ Fum_ Honig_Mel_ Stärke_Polvilh_	
Tx_	Soz_	Rel_ Fam_ Vis_Besuch_ Fest_ Assist_ Rel_ Ponto_Enc_ Div_Freizeit_	Erziehung_Edu_ Fam_Treff_Enc_Fam_ Cas_Heirat_ Jugend_Juv_ Namor_ Verein_Club_ Amiz_Freund_ Ball_	
Tx_	Techn_	Instr_ Instr_Trab_ Gerät_Maq_ Prod_ Transp_ Comun_ Haus_Cas_	Luftschiff_Avião_ Ochs_Wagen_Carr_Boi_ Säge_Serra_ Zuckrer_Press_Moenda_ Energ_Eletr_Strom_ TV_Tele_ Comput_	

			Tel_ Cel_Handy_	
Tx_	Trab_	Beruf_Prof_ Agri_ Com_ Domest_ Tag_Diar_ Wert_Valor_	Schneider_Alfaiate_ Näherin_Costur_ Schmidt_Ferr_ Fab_Arbeit_ Motorist_Schofer_ Tagelohn_Diar_ Diar_Tag_ fleissich_	
Tx_	Tur_	Info_ Mapa_ Portal_ Pred_Hist_ Atraç_ Fest_	Stadt_Cid_ Museu_ Rota_	
Tx_	Viaj_	Exterior_ Br_ Ale_ Besuch_Vis_	(destino_ (motivo)_ Parent_	
Tx_	Zoo_	Schlange_Cobr_ Ochse_Boi_ Biene_Abelh_ Ameise_Formig_ Vogel_Ave_ Fisch_Peix_ Insekt_ Wild_Anim_Selv_ Anim_Domest_ Anim_Trab_ Zucht_Criaç_Anim -	Dreckbauer_ Hohrwickler_ Scharack_ Gift_Ven_ (nome)_	