

Textos didatizados e não didatizados de alemão como língua adicional: uma análise quantitativa e qualitativa para o projeto *Zeitgeist*

[Didactic and non-didactic texts in the context of German as an additional language: a linguistic-quantitative analysis for the *Zeitgeist* project]

<http://dx.doi.org/10.11606/1982-88372444421>

Marceli Aquino¹
Adriano Cherchiglia²

Abstract: In this work we analyze the frequency of occurrence of reflexive pronouns extracted from texts of four textbooks of German as an additional language (*Blaue Blume*, 2011; *DaF kompakt neu A2*, 2016; *Menschen A2*, 2013; *Studio d A2*, 2006), aiming to discuss ways to identify didactic and non-didactic texts. This investigation is justified in the context of designing an empiric basis to help classify texts for a project whose main goal is to develop a didactic material designed to the Brazilian university context. For the proper treatment of the samples, a quantitative and qualitative analysis was performed with the help of the softwares *WordSmith Tools* (2016) and *QuAX-DaF* (2019). These tools allow us to assess the frequency of occurrence of our target subject (reflexive pronouns), as well as the degree of authenticity of the samples based in a German *corpus* (*Leipzig Corpora Collection*). The results suggest that the knowledge of the frequency of occurrence can be helpful to assist teachers and researchers in choosing texts to be used in specific teaching contexts, regarding the lexical and grammatical progression.

Keywords: Frequency of occurrence; Authenticity in the context of textbooks; German as an additional language.

Resumo: A partir de dados sobre a frequência de ocorrência de pronomes reflexivos, retirados de textos de quatro livros didáticos de língua alemã (*Blaue Blume*, 2011; *DaF kompakt neu A2*, 2016; *Menschen A2*, 2013; *Studio d A2*, 2006), o presente artigo discute meios para a identificação de insumos didatizados e não didatizados. Essa proposta é justificada pelo interesse em desenvolver uma base empírica que auxilie a seleção de textos para um projeto de criação de um material didático destinado ao ensino de alemão em contexto universitário no Brasil. Para o tratamento dos textos utilizamos os programas *WordSmith Tools* (2016) e *QuAX-DaF* (2019). Tais ferramentas apresentam dados de frequência de ocorrência nos textos investigados, além de

¹ Universidade de São Paulo, Avenida Professor Luciano Gualberto, 403, Cidade Universitária, Butantã, São Paulo, SP, 05508-010, Brasil. E-mail: marceli.c.aquino@usp.br. ORCID: 0000-0003-0518-7639

² Universidade Federal do ABC, Avenida dos Estados, 5001, Santa Terezinha, Santo André, SP, 09210-580, Brasil. E-mail: adriano.cherchiglia@ufabc.edu.br. ORCID: 0000-0002-2457-1671



estabelecer uma relação com um *corpus* autêntico de língua alemã (*Leipzig Corpora Collection*). Os resultados indicam que a investigação da frequência de ocorrência pode estabelecer uma base quantitativa e qualitativa para a escolha textual de acordo com a progressão lexical e gramatical desejada, podendo ser um critério para a inclusão ou de exclusão de textos para o ensino de línguas.

Palavras-chave: Frequência de ocorrência; Autenticidade de textos didáticos; Alemão como língua adicional.

1 Introdução

A abordagem comunicativa (AC), embora tenha surgido nos 70, ainda hoje é muito difundida no ensino e aprendizagem de línguas adicionais (LA) e na produção de livros didáticos (LD). Uma de suas premissas é que a língua não é encarada como um sistema de formas, mas um instrumento de ação humana (NEUNER; HUNFELD 1993: 31). Assim, nessa abordagem, a gramática passa a ocupar um papel secundário, valorizando a intenção comunicativa, a partir de contextos próximos à realidade. Esta mudança de foco, que se afasta de um caráter estruturalista, também é influenciada pelas novas descobertas na área de estudos da linguagem, como a linguística pragmática, sob cuja perspectiva a língua é um instrumento de comunicação e interação humana.

No final do século 20, houve uma nova alteração de paradigma com o advento da abordagem intercultural, para a qual não há ensino de línguas sem cultura (KRAMSCH 1998: 3). Nesse momento, sobretudo devido ao desenvolvimento da AC e, a transformação social proveniente do processo de globalização, discussões sobre temas interculturais ganharam destaque, fazendo com que se tornassem um dos objetivos essenciais do ensino e aprendizagem de LA. Nesta abordagem a gramática deveria ser apresentada a partir de textos, orais e escritos, que contêm as estruturas a serem aprendidas. Deposita-se especial atenção aos aspectos de pesquisa em recepção e produção de textos, como o uso de textos literários em sala de aula, assim como a estrutura textual e seu contexto.

Atualmente, novas abordagens, como é o caso da Abordagem Intercultural e da Consciência Linguística (*Sprachbewusstsein*), além de levar em conta a relação entre língua e cultura, incluem em sua perspectiva de investigação a reflexão metalinguística (como a inferência) no aprendizado de línguas. Essas abordagens buscam uma sensibilização linguística interdisciplinar por meio da compreensão holística do processo de aprendizagem (LUCHTENBERG 2008: 112), na qual os fenômenos linguísticos

(estrutura, regras, modelos) devem ser tratados de forma reflexiva e consciente. Com estudos acerca do papel facilitador do processo de aprendizagem, a perspectiva da Consciência Linguística evidencia a importância da reflexão crítica para o ensino de LAs. O objetivo é, portanto, tornar o aprendiz consciente das normas linguísticas e socioculturais da língua alvo, assim como da sua própria.

Desde a AC, até as discussões atuais, lidar com textos de vários níveis é um componente central do ensino e aprendizagem de línguas. Conseqüentemente, a compreensão e uso de textos está vinculada à competência discursiva, isto é, os significados das sequências linguísticas estão ligados a elementos externos, como por exemplo, quem fala, para quem se fala, com que intenção, o que o texto está respondendo a favor ou contra, quais são os papéis sociais, entre outros (KRAMSCH 1998: 5).

Nesse sentido, o ensino de LAs se desenvolve e manifesta no uso receptivo e produtivo de textos, constituindo objetivos primordiais da competência comunicativa e da capacidade de discurso (HALLET 2010: 173). Por esse motivo, o desenvolvimento sistemático da competência textual pode ser considerado como o núcleo da aprendizagem em LA, uma vez que textos, tanto no contexto social como didático, representam as unidades básicas discursivas de toda a comunicação (HALLET 2010: 174).

Para Rösler (2012: 37), os materiais didáticos e LDs atuais apresentam fenômenos de língua e cultura a serem introduzidos conforme a lógica do nivelamento dos conhecimentos e competências da(o)s estudantes. Eles também primam por certa variedade de gêneros textuais, buscando oferecer uma quantidade diversa de exercícios e tarefas, que servem como subsídio à autonomia na aprendizagem. Um dos elementos que se impõem do ponto de vista da prática reflexiva reside no questionamento sobre as formas de uso dos textos em LDs e, sua capacidade em atender às necessidades da(o)s aluna(o)s e das demandas específicas do contexto universitário. Assim, defendemos neste trabalho que se faz necessário pensar criticamente sobre a concepção e uso de textos nos materiais didáticos atuais.

A maioria dos LDs de Alemão como Língua Adicional (ALA) utilizados no Brasil são produzidos na Alemanha, com base em parâmetros do Quadro Europeu Comum de Referência para Línguas (CONSELHO DA EUROPA 2001). Tais materiais pouco consideram a comunidade para a qual são elaborados (ARANTES 2018: 6), negligenciando o contexto dos aprendentes, seus interesses e motivações. Sendo que os LDs e, conseqüentemente o gênero textual, representam componentes centrais no ensino e aprendizagem de línguas,

deve-se problematizar e investigar se sua abordagem é relevante em diferentes contextos de aprendizagem. Tal investigação pode ser feita de diversas formas, como levantamento de opinião da(o)s estudantes e professora(e)s acerca do material, teste de autenticidade, assim como, abordagens linguístico-quantitativas baseadas na identificação e análise da frequência de ocorrência, como é o intuito deste trabalho.

Tendo isso em vista a importância da discussão sobre a produção de materiais didáticos que enfatizem as necessidades locais e regionais para o ensino de ALA, apresentamos neste artigo os resultados de um estudo que se insere em um projeto de elaboração de um material didático voltado ao contexto acadêmico brasileiro. Levando em conta o debate sobre as especificidades do ensino de LA em contexto universitário, o projeto *Zeitgeist* busca desenvolver um livro didático que contemple variantes locais em um discurso acadêmico crítico alinhado às reflexões teóricas em ALA (OLIVEIRA; LADEL, no prelo). Nesse sentido, um dos objetivos principais do projeto é a seleção de conteúdo e temas que permitem uma progressão flexível, que concede liberdade à/ao professora/professor no planejamento de atividades para o seu público-alvo.

Parte da discussão essencial da elaboração de materiais didáticos se situa em torno da seleção e trabalho com textos autênticos. A busca por textos autênticos é uma tarefa desafiadora, mas que pode ser facilitada pelo uso de ferramentas que ajudem a identificar e selecionar materiais que condizem com diferentes contextos de aplicação. Assim, este estudo exploratório oriundo do projeto *Zeitgeist* tem a intenção de apresentar soluções para a sistematização de seleção textual que auxilie o trabalho da(o) professora/professor. Tal sugestão consiste em apresentar ferramentas de análise de dados que identifiquem a frequência de ocorrência, como por exemplo, o número de palavras, os temas mais recorrentes, a complexidade da estrutura e vocabulário, apresentando, portanto, alternativas para o ensino e aprendizagem de ALA, de acordo com a especificidade de cada contexto.

Para tanto apresentamos uma análise quantitativa e qualitativa – repetição de palavras e averiguação das estruturas gramaticais e lexicais – com textos didatizados e não didatizados de quatro LDs. Tal investigação permitirá apreciar como os softwares, *Word Smith Tools* e *QuAX-DaF*, podem ser utilizados para encontrar alternativas confiáveis para a seleção de textos a serem utilizados na elaboração de materiais adaptados a condições específicas, respeitando o princípio da autonomia da(o)s professora(e)s e aprendizes. Assim, ao analisar textos de LDs globais, introduzimos uma

explicação das possibilidades de tratamento textual, propondo soluções para o trabalho com texto que envolva interesses e necessidades específicas.

Para uma melhor compreensão da proposta deste trabalho, apresentamos uma revisão teórica que serviu de base para a análise de dados (Seção 2), como as teorias do trabalho com texto em LA (HALLET 2010), assim como a discussão sobre a autenticidade e, a sua relação com o ensino de línguas (RÖSLER 2012, ANDRADE E SILVA 2017). Com o intuito de levantar hipóteses sobre a didatização dos textos, discutimos as teorias de análise de dados quantitativa e qualitativa dentro da perspectiva da análise de *corpus* (BYBEE 2010). A metodologia de pesquisa (Seção 3) com textos para o ensino de pronomes reflexivos em quatro livros didáticos – *Blaue Blume, DaF kompakt neu A2, Menschen A2, Studio d A2* – foi desenvolvida por meio de dois programas de análise de *corpus*, *QuAX-DaF* e *WordSmith Tools*. Após o tratamento dos textos selecionados com os dois *softwares*, realizamos uma análise dos resultados (Seção 4), buscando encontrar evidências com relação à autenticidade dos textos. Discutimos na conclusão (Seção 5) como os resultados desta pesquisa podem auxiliar na escolha de textos para o desenvolvimento de materiais didáticos locais, tendo em vista as necessidades e interesses de contextos específicos. Assim, defendemos que as consultas em ferramentas de análise de *corpus* podem apoiar a busca de insumos textuais para o trabalho em sala de aula, ou para projetos editoriais, como é o caso do *Zeitgeist*.

2 Arcabouço teórico

A(O)s aprendizes de uma LA são constantemente contrastados com uma grande variedade de textos no decorrer do processo de aprendizagem e serão convidada(o)s a produzir textos como parte desse processo. Rösler (2012: 39) defende que o texto é mais do que uma coleção de orações, já que mantém uma relação de significados em determinado contexto temático com uma função específica para a comunicação. Portanto, além da distinção entre produção e recepção, é necessário pensar nas características do texto que são produzidas e recebidas. Nesse sentido, Hallet (2010: 174) divide os tipos de textos e sua função em termos de sua aplicação no ensino e aprendizagem de LA:

- Textos didatizados (*didaktisierte Texte*): aqueles que têm a função de apresentar certas estruturas linguísticas ou novo vocabulário, tendo em vista uma progressão linguístico-gramatical. Os textos didatizados foram bastante criticados na abordagem comunicativa,

pois o conteúdo desses textos era subordinado apenas a intenções didáticas específicas para o ensino e aprendizagem de LA, sem levar em conta os aspectos comunicativos e culturais.

- Textos relativos ao nível de aprendizagem (*lernstufenbezogene Texte*): insumos voltados para o ensino de LA que imitam textos originais (artigos de jornal, textos literários, entrevistas, e-mails etc.), tendo em vista o nível de aprendizagem de cada grupo. Esses textos são orientados para o conteúdo, podendo, portanto, desenvolver uma competência textual mais complexa.

- Textos não didatizados (*nicht didaktisierte Texte*): esses textos devem ser entendidos como manifestação autêntica da língua e cultura em LA e servir de uso para a sala, facilitando, portanto, o diálogo intercultural. Esses textos representam as formas convencionais da comunicação (telefonema, cartas, notícias de TV, entre outros) ou da literatura, funcionando como modelos genéricos de comunicação para o ensino e aprendizagem de línguas.

- Textos de aprendizes (*Lernertexte*): com o objetivo de desenvolver habilidades discursivas, esses textos desempenham um papel central no ensino de LA. As produções da(o)s aluna(o)s representam atos de comunicação individuais autênticos e constituem a interação comunicativa entre texto e destinatário, bem como entre a(o)s própria(o)s estudantes.

- Discurso em sala de aula (*Classroom discourse und Instruktion*): instruções e/ou discursos relacionados com a organização e interação durante do processo de aprendizagem em sala de aula.

As diferentes práticas didáticas com o uso de textos para o ensino de língua têm o propósito principal de preparar a(o) aluna(o) para a compreensão e produção textual em vários níveis. Para tanto, são utilizadas algumas estratégias como a compreensão geral do texto (compreensão global), a identificação e extração de informações individuais específicas (compreensão seletiva) ou com maiores detalhes (compreensão detalhada). Não obstante, ainda é possível o desenvolvimento de estratégias cognitivas com a interpretação do contexto para a compreensão de palavras desconhecidas, análise da derivação e composição morfológica das palavras, ativação de conhecimentos anteriores (linguísticos e/ou contextuais), entre outros (AQUINO 2021: 42). Logo, lidar com diferentes textos, como é a norma na vida cotidiana, oferece às/aos aprendizes uma correspondência intertextual e ajuda a entender e usar a língua-alvo (HALLET 2010: 54).

O trabalho com textos é, portanto, essencial para o ensino de línguas, pois fornece padrões para o processamento e produção de experiências, bem como estruturas para interação comunicativa e social (AQUINO 2021: 23). Segundo Bakhtin (1997), os gêneros constituem-se de linguagem concreta e viva, permitindo à/ao aluna/aluno a reflexão sobre os sentidos estabelecidos por meio do seu uso. Assim, o texto, tanto escrito como oral, desempenha um papel importante para o ensino e aprendizagem de LA, não apenas para a mediação de linguagem, mas como prática sociocultural, ampliando a possibilidade de participação cultural (HALLET 2010: 52).

Nesse sentido, desde a virada comunicativa, houve um aumento na demanda por textos autênticos. Porém, a maioria dos LDs disponíveis no mercado brasileiro para ensino e aprendizagem de ALA, além de não apresentar insumos autênticos, utiliza o texto apenas para a mediação de vocabulário e gramática, não respondendo ao propósito comunicativo e intercultural. O desafio continua, portanto, na possibilidade de encontrar e trabalhar com textos relevantes e autênticos, com a progressão lexical e gramatical adequada, que oferecem exemplos de linguagem em uso e práticas socioculturais.

A discussão sobre a autenticidade e, a sua relação com o ensino de línguas, continua bastante atual na área da Linguística Aplicada e ensino e aprendizagem de LA (RÖSLER 2012: 39), porém não existe consenso sobre o significado desse termo, que pode estar associado aos materiais, aos participantes, à situação sociocultural, às propostas do ato comunicativo, ou a uma combinação entre esses fatos. O termo autêntico continua presente no debate sobre os LDs em ALA, mas efetivamente poucos textos realmente autênticos são encontrados. Segundo Rösler (2012: 39), em sua acepção mais corrente, texto autêntico é entendido como o original, aquele sem alterações para fins didáticos.

A importância do uso de textos autênticos no ensino de línguas está atrelada ao trabalho com elementos que caracterizam o gênero textual e, na capacidade da(o) aluna(o) de reconhecer as convenções deste gênero em determinada língua e cultura. Muitas vezes os textos são descritos com base na estrutura, na gramática e na forma, mas pouca atenção é dada ao conteúdo vinculado pelo texto, à temática ou à questão cultural (ANDRADE E SILVA 2017: 19). O material autêntico pode, portanto, ser essencial para desenvolver a reflexão crítica e o contato da(o) aprendiz com os discursos de comunidades da língua-alvo. Nesse sentido, destacamos a importância da discussão sobre a escolha dos textos nos LDs, já que estes servem como suporte para outras atividades comunicativas.

Como mencionado anteriormente, a motivação para este artigo adveio da tarefa de encontrar alternativas para a seleção de textos durante o desenvolvimento de um livro didático local, isto é, para o contexto de aprendizagem de alemão em universidades brasileiras. Assim, a análise de textos dos livros didáticos globais auxilia a compreender sobre a didatização dos materiais utilizados atualmente em meio universitário no Brasil, propondo uma alternativa para o trabalho com texto que envolva interesses e necessidades específicas. No intuito de levantar hipóteses sobre a autenticidade dos textos, e assim, a sua função em sala de aula em diferentes contextos, apresentamos dados qualitativos e quantitativos da contagem de repetição de palavras, averiguação das estruturas (gramaticais e lexicais) empregadas nas orações e, por fim, uma análise comparativa entre a frequência de ocorrência no texto e em um *corpus* de língua alemã.

Segundo Bybee (2010), a repetição de estruturas linguísticas, sejam elas palavras ou expressões, tem efeitos cognitivos em nosso dicionário mental, isto é, palavras usadas com maior frequência são lembradas com mais facilidade, ao passo que as menos utilizadas são recuperadas com mais dificuldade pela(o)s aprendizes. Ainda segundo a autora, as palavras com baixa frequência de ocorrência dependem de conexões com outros itens lexicais e classes de palavras para serem ativadas no léxico mental. Conseqüentemente, a cada vez que esses itens são acessados, reforçam a coesão de sua rede. A frequência exerce, portanto, um papel importante para o ensino e aprendizagem de fenômenos linguísticos. Mesmo assim, ainda existe muito a ser investigado e discutido sobre esse tema, como, por exemplo, a definição de limites precisos para faixas de frequência e sua adequação para o ensino e aprendizagem de LA.

Logo, uma investigação sobre a frequência de ocorrência em textos didatizados e não didatizados pode auxiliar professora(e)s e aprendizes a compreender as diferentes funções exercidas por textos empregados no ensino de línguas, assim como para a seleção do material e LD que reflita os interesses de um determinado público-alvo (AQUINO 2019: 146). A investigação acerca da frequência de palavras em um texto apresenta, assim, algumas vantagens metodológicas, já que procura considerar como o gênero escrito, tão essencial para o ensino e aprendizagem de LA, é utilizado nos materiais didáticos comunicativos, e conseqüentemente, na prática em sala de aula. Finalmente, uma investigação linguístico-quantitativa parece ser favorável para estabelecer uma relação entre a frequência de ocorrência, e a didatização/não didatização de textos, dando indicações sobre a sua autenticidade e relevância para o ensino e aprendizagem de ALA.

3 Metodologia

No sentido de investigar textos didatizados e não didatizados para um projeto de elaboração de um livro didático brasileiro de ALA para o contexto acadêmico, selecionamos materiais que são adotados em cursos universitários no Brasil, tanto de graduação, como cursos de extensão. Segundo os dados coletados em Schmidt (2017: 42), os livros *DaF Kompakt* e *Studio d A2* são um dos mais adotados em curso de ensino superior. Ainda, segundo a autora, *Menschen* é um dos materiais mais escolhidos pelas instituições para cursos extracurriculares voltados para a comunidade acadêmica. O livro *Blaue Blume* é adotado atualmente em apenas uma instituição, a Unicamp, não configurando como um livro de grande circulação. Não obstante, a relevância deste livro para esse estudo se justifica por ele ser o único que contém textos autênticos, isto é, não didatizados. Além disso, *Blaue Blume* apresenta uma versão brasileira, com traduções e adaptações para o contexto local, fazendo com que essa variante se aproxime do objetivo do projeto *Zeitgeist*, isto é, de desenvolvimento de um material didático voltado a um contexto específico de ensino e aprendizagem de alemão. Portanto, como objetos de estudo, utilizaremos textos dos quatro livros didáticos elencados acima: *Blaue Blume*, *DaF kompakt neu A2*, *Menschen A2*, *Studio d A2*.

Uma vez tendo selecionado os livros, buscamos algum tema que fosse trabalhado em todos eles por meio de textos escritos, de modo que pudéssemos realizar uma comparação quantitativa. Essa foi a motivação para a escolha da área de interesse, ou seja, os pronomes reflexivos³ (associados a diferentes verbos), que foi trabalhado nos materiais didáticos nos seguintes trechos.⁴

Blaue Blume (2011: 96): *Haben Sie jemals Ihr Herz verloren?*

DaF kompakt neu A2 (2016: 30): *Notizen aus Köln.*

Menschen A2 (2013: 65): *Sind Sie restlos glücklich?*

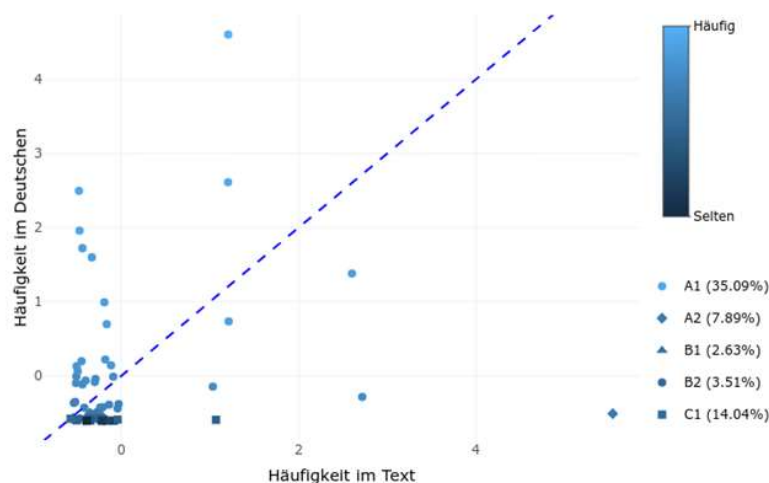
Studio d A2 (2006: 61): *Freizeitaktivitäten.*

³ Neste trabalho não temos a intenção de discutir acerca dos pronomes reflexivos como conteúdo gramatical. A seleção de tal tema se deu apenas pelo interesse em analisar materiais didáticos que apresentavam o ensino de tópicos gramaticais por meio de textos. Assim, a seleção dos pronomes reflexivos se justifica apenas pelo fato que os quatro livros didáticos apresentam textos para o ensino deste tema, configurando uma base comparativa para a análise de dados.

⁴ Os textos completos estão em anexo ao final do artigo.

De posse dos textos, nossa próxima tarefa foi a de escolher a(s) ferramenta(s) mais apropriada(s) para realizar a análise. Recentemente foi desenvolvida uma plataforma, disponível on-line e gratuita, denominada *QuAX-DaF*⁵. A intenção do autor desse programa, o Leitor do DAA na Zhengzhou University, Daniel Jach, é a de dar indicações com relação à autenticidade de textos em língua alemã. Para tanto, o usuário insere o texto de sua escolha na plataforma e obtém um gráfico no qual a frequência das palavras que o constituem é comparada com a frequência das mesmas na base Leipzig Corpora Collection (LCC) (GOLDHAHN; ECKART; QUASTHOFF 2012). Entende-se, portanto, autenticidade por quanto mais próxima a frequência das palavras do texto de entrada esteja da frequência da base LCC. Utilizando o texto citado acima do livro *Blau Blume*, como exemplo de análise, obtemos os seguintes dados do programa:

Gráfico 1 - Frequência de ocorrência no texto e em comparação com o *corpus*



Fonte: *QuAX-DaF*

Como podemos perceber no gráfico acima, o programa apresenta os lemas⁶ (pontos azuis), em termos da sua frequência no texto (eixo horizontal), e sua frequência no *Corpus LCC* (eixo vertical). O *software* apresenta como referência uma “reta ideal”, na qual os pontos que estejam sobre essa teriam a mesma frequência no texto alvo que no *Corpus LCC*. Dessa forma, pontos que se localizem abaixo da “reta ideal” teriam frequência mais elevada no texto de entrada, do que seria esperado tendo por base a análise do *corpus*, o que poderia indicar que o texto foi didatizado, se distanciando da autenticidade. Além disso, quão mais dispersos da reta ideal estejam os pontos, maior é a

⁵ <https://danieljach.shinyapps.io/quax-daf/>

⁶ Em linguística, forma canônica de uma palavra encontrada em dicionários e léxicos ou também a palavra de base (NATION; WARING 1997: 9).

indicação de não-autenticidade. Conforme discutiremos na seção dos resultados, essa análise é tão mais precisa quanto maior seja o texto de entrada. Gostaríamos, por fim, de mencionar que, para o presente trabalho, os números apresentados nos eixos (referentes ao logaritmo da frequência) não nos serão relevantes. Sendo assim, não descreveremos em maiores detalhes o gráfico, mas caso o leitor esteja interessado na metodologia utilizada, maiores informações podem ser encontradas no manual do programa.

Além da classificação de acordo com a frequência, o programa apresenta os diversos lemas do texto de entrada como pontos azuis, no qual a tonalidade do azul indica quão frequente é aquele lema no *Corpus LCC*. No caso, azul-claro é reservado para lemas frequentes e escuro para a situação oposta. Além da tonalidade do azul, o programa organiza a forma dos pontos (círculo, losango, triângulo, quadrado) em termos da classificação do lema de acordo com o nível do Quadro Comum de Referência de Línguas Europeias - GER (Gemeinsamen Europäischen Referenzrahmens) (CONSELHO DA EUROPA 2001). Essa divisão é baseada em estudos empíricos realizados por Tschirner (2019: 101) que possibilitam uma relação entre a frequência dos lemas no *Corpus LCC* e os diversos níveis GER. Ou seja, os 800 lemas mais frequentes no *Corpus LCC* seriam referentes ao nível A1, os 1600 mais frequentes ao nível A2 e assim sucessivamente, como podemos depreender da tabela a seguir:

Tabela 1 - Relação entre nível GER e extensão do vocabulário do falante

Nível GER	A1	A2	B1	B2	C1
Vocabulário	800	1600	3200	4000	5500

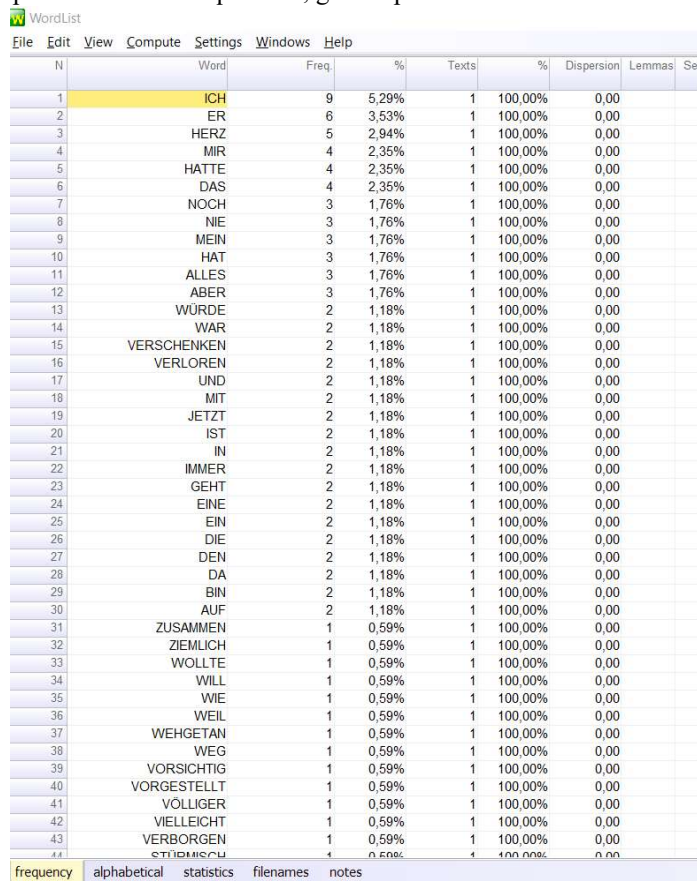
Fonte: Jach (2019: 5)

Além do gráfico 1, o *QuAX-DaF* pode ser utilizado para gerar materiais a serem usados em sala de aula. Como não teremos intenção nesse trabalho de utilizar esses materiais, deixaremos a cargo do leitor consultar maiores informações no manual da ferramenta.

Uma vez tendo a análise acerca da autenticidade indicada pela ferramenta *QuAX-DaF*, necessitamos quantificar a frequência de ocorrência das palavras do texto em estudo, tarefa que será realizada através da ferramenta *WordSmith Tools* versão 7.0. Esse software gratuito deve ser instalado no computador do usuário e tem uma série de funcionalidades para análise textual. O texto a ser analisado deve ser salvo em formato *txt* em uma pasta do computador, a qual será acessada pelo *software*. Para nossos interesses de pesquisa

utilizamos duas das ferramentas, a primeira relativa à contagem das palavras do texto inserido (*Word list*), e a segunda (*Concord*) responsável por identificar trechos do texto original. Como exemplo, mostramos a seguir a frequência das palavras do primeiro texto (*Blaue Blume*) conforme gerado pelo *WordSmith Tools*:

Figura 1 - Lista de palavras e suas frequências, gerada pela ferramenta *Word list* do *WordSmith Tools*



N	Word	Freq.	%	Texts	%	Dispersion	Lemmas	Set
1	ICH	9	5,29%	1	100,00%	0,00		
2	ER	6	3,53%	1	100,00%	0,00		
3	HERZ	5	2,94%	1	100,00%	0,00		
4	MIR	4	2,35%	1	100,00%	0,00		
5	HATTE	4	2,35%	1	100,00%	0,00		
6	DAS	4	2,35%	1	100,00%	0,00		
7	NOCH	3	1,76%	1	100,00%	0,00		
8	NIE	3	1,76%	1	100,00%	0,00		
9	MEIN	3	1,76%	1	100,00%	0,00		
10	HAT	3	1,76%	1	100,00%	0,00		
11	ALLES	3	1,76%	1	100,00%	0,00		
12	ABER	3	1,76%	1	100,00%	0,00		
13	WÜRDE	2	1,18%	1	100,00%	0,00		
14	WAR	2	1,18%	1	100,00%	0,00		
15	VERSCHENKEN	2	1,18%	1	100,00%	0,00		
16	VERLOREN	2	1,18%	1	100,00%	0,00		
17	UND	2	1,18%	1	100,00%	0,00		
18	MIT	2	1,18%	1	100,00%	0,00		
19	JETZT	2	1,18%	1	100,00%	0,00		
20	IST	2	1,18%	1	100,00%	0,00		
21	IN	2	1,18%	1	100,00%	0,00		
22	IMMER	2	1,18%	1	100,00%	0,00		
23	GEHT	2	1,18%	1	100,00%	0,00		
24	EINE	2	1,18%	1	100,00%	0,00		
25	EIN	2	1,18%	1	100,00%	0,00		
26	DIE	2	1,18%	1	100,00%	0,00		
27	DEN	2	1,18%	1	100,00%	0,00		
28	DA	2	1,18%	1	100,00%	0,00		
29	BIN	2	1,18%	1	100,00%	0,00		
30	AUF	2	1,18%	1	100,00%	0,00		
31	ZUSAMMEN	1	0,59%	1	100,00%	0,00		
32	ZIEMLICH	1	0,59%	1	100,00%	0,00		
33	WOLLTE	1	0,59%	1	100,00%	0,00		
34	WILL	1	0,59%	1	100,00%	0,00		
35	WIE	1	0,59%	1	100,00%	0,00		
36	WEIL	1	0,59%	1	100,00%	0,00		
37	WEHGETAN	1	0,59%	1	100,00%	0,00		
38	WEG	1	0,59%	1	100,00%	0,00		
39	VORSICHTIG	1	0,59%	1	100,00%	0,00		
40	VORGESTELLT	1	0,59%	1	100,00%	0,00		
41	VÖLLIGER	1	0,59%	1	100,00%	0,00		
42	VIELLEICHT	1	0,59%	1	100,00%	0,00		
43	VERBORGEN	1	0,59%	1	100,00%	0,00		
44	STREBISCH	1	0,59%	1	100,00%	0,00		

Fonte: *WordSmith Tools*

Como podemos perceber, na primeira coluna são mostradas as palavras que constituem o texto, na segunda coluna a frequência e na terceira a frequência relativa (razão entre a frequência da palavra estudada e o total de palavras do texto). As demais colunas são relevantes apenas quando se realiza uma análise comparada entre textos distintos.

A seguir apresentamos uma descrição dos resultados fornecidos por meio da ferramenta *Concord*. Para utilizá-la, deve-se selecionar alguma palavra dentre as mostradas na figura 4. Marcando, por exemplo, a palavra “*mir*”, a ferramenta *Concord* apresenta a seguinte informação:

Figura 2 - Trechos de frases com palavra “*mir*”, gerada pela ferramenta *Concord*

The screenshot shows the Concord software interface. At the top, there is a menu bar with options: File, Edit, View, Compute, Settings, Windows, and Help. Below the menu bar is a table titled 'Concordance'. The table has two columns: 'N' (line number) and 'Concordance' (text snippet). The first row (N=1) is highlighted in yellow and contains the text: 'mit Mahmut jetzt ein Jahr zusammen. An dem gefällt mir einfach alles. Aber heiraten will ich nicht. Nie! Das'. The second row (N=2) contains: '. Ich hab mein Herz noch nie verloren. Er hat mir in der Fleischerei immer mit den schweren Kisten'. The third row (N=3) contains: '. Jetzt geht ihm das Schreien auf die Nerven. Ich hatte mir das alles anders vorgestellt! Ich würde mein Herz'. The fourth row (N=4) contains: 'jeder hatte seinen Dickkopf. Als er gegangen ist, hat mir das sehr wehgetan. Seither bin ich vorsichtig'.

N	Concordance
1	mit Mahmut jetzt ein Jahr zusammen. An dem gefällt mir einfach alles. Aber heiraten will ich nicht. Nie! Das
2	. Ich hab mein Herz noch nie verloren. Er hat mir in der Fleischerei immer mit den schweren Kisten
3	. Jetzt geht ihm das Schreien auf die Nerven. Ich hatte mir das alles anders vorgestellt! Ich würde mein Herz
4	jeder hatte seinen Dickkopf. Als er gegangen ist, hat mir das sehr wehgetan. Seither bin ich vorsichtig

Fonte: *WordSmith Tools*

Em negrito está indicado a palavra selecionada (*mir*) e, todos os trechos em que a mesma aparece no texto inserido. Observando a figura 4, havia 4 ocorrências da palavra “*mir*”, fato esse também indicado pela ferramenta *Concord*. Um fato a se mencionar é que, ao mostrar as ocorrências das palavras no texto, podemos estudar as estruturas gramaticais e lexicais nas quais essa palavra aparece. No nosso exemplo, percebemos que “*mir*” terá função de pronome reflexivo em apenas uma das ocorrências, “*mir vorgestellt*”. Para as demais ocorrências teremos apenas função de pronome pessoal no caso dativo.

Nesse sentido, a aplicação das diversas ferramentas expostas nesta seção será essencial para analisarmos de maneira mais completa e global o tema gramatical dos pronomes reflexivos. Entretanto, essa prática não é exclusiva para o nosso estudo, mas deve sempre ser utilizada ao se optar por descrições numéricas, de modo a se evitar conclusões enviesadas. De modo a auxiliar a(o) professora/professor e pesquisador(a) na escolha das ferramentas mais eficazes para seus próprios objetivos, formulamos a seguir algumas possibilidades de estudo, indicando como tais ferramentas podem ser utilizadas na investigação de frequência de ocorrência em textos de LDs.

Objetivo I: indicação para escolha textual com base nas palavras e temas relevantes para as lições (progressão lexical/temática);

Abordagem I: a contagem de frequência disponibilizada pelo *WordSmith Tools* para percepção inicial da frequência do tema estudado no texto alvo. Utilização do *Concord* para averiguação das estruturas (gramaticais e lexicais) empregadas. Gráfico comparativo de frequência do *QuAX-DaF* para uma primeira percepção acerca das palavras do texto e sua classificação de acordo com o nível GER.

Objetivo II: averiguar a autenticidade de um determinado texto;

Abordagem II: gerar um gráfico comparativo entre frequências no *Corpus LCC* e no texto alvo disponibilizado pelo *QuAX-DaF*. Análise com relação à dispersão em torno da reta ideal e quantidade de pontos abaixo da mesma.

Por fim, tendo apresentado as ferramentas de análise dos textos, descreveremos na próxima seção os resultados que obtivemos ao aplicar a metodologia descrita nessa seção para os quatro textos de LDs de ALA.

4 Análise e resultados

Nessa seção apresentaremos uma descrição quantitativa e qualitativa que permitirá apreciar como *softwares* podem ser utilizados na escolha de textos para o ensino e aprendizagem de ALA. Conforme já apresentado na seção anterior, utilizaremos quatro textos de LDs distintos. No presente trabalho, exemplificaremos como as ferramentas *Word Smith Tools* e *QuAX-DaF* podem ser utilizadas para se correlacionar a frequência de ocorrência⁷ de pronomes reflexivos e a autenticidade dos textos empregados.

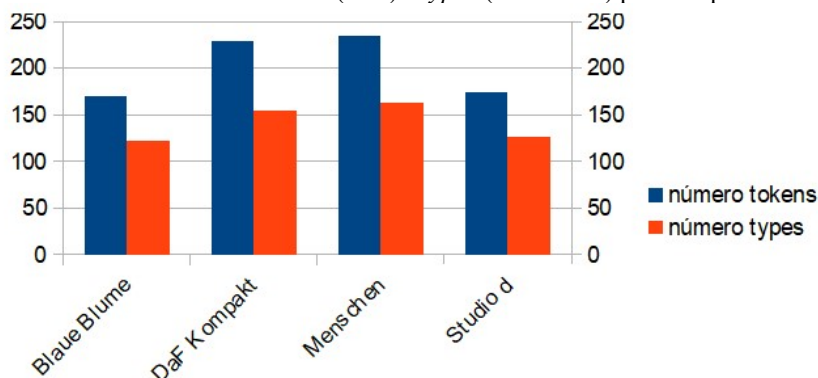
Nesse sentido, nosso primeiro objetivo será apresentar uma descrição generalista dos textos, com o intuito de traçar correlações entre os diversos materiais didáticos. Para tanto, faremos uso do software *WordSmith Tools*, empregando duas de suas ferramentas: 1) *Word list*, que analisa as frequências de ocorrência do tema analisado (pronomes reflexivos) nos diversos livros; 2) *Concord*, que possibilitará observar diferenças com relação às estruturas gramaticais empregadas nos materiais. Uma vez que tenhamos essa descrição generalista, passaremos a formular hipóteses com relação à autenticidade dos textos e sua relação com os resultados obtidos. Isso nos levará a empregar o software *QuAX-DaF*, que gera dados acerca da autenticidade dos diversos textos.

Dado o panorama delineado acima, iniciamos a apresentação e discussão dos resultados. Utilizando o software *WordSmith Tools*, observamos que os quatro textos continham um número de palavras (*tokens*) entre 170 e 235, ao passo que o número de palavras distintas (*types*) variou entre 123 e 163, conforme pode ser verificado pelo

⁷ Para analisar a frequência de ocorrência, realizamos uma adaptação do conceito proposto por Bybee (2002). Na perspectiva deste artigo, a frequência de ocorrência será sempre um conceito relativo, ou seja, ela representa, por exemplo, a razão entre o número de pronomes reflexivos presentes no texto e o número total de palavras.

gráfico abaixo onde em azul representamos o número de *tokens* e em vermelho o número de *types*.⁸

Gráfico 2 - Indicando o número de *tokens* (azul) e *types* (vermelhos) para os quatros LDs analisados



Fonte: autoria própria

Foquemos inicialmente nos *tokens*. Como podemos verificar, os textos de *DaF kompakt neu A2* e *Menschen* apresentavam insumos maiores (aproximadamente 25% maiores), que os contidos em *Blaue Blume* e *Studio d*. No caso dos *types*, também encontramos números superiores nos textos empregados por *DaF kompakt* e *Menschen* com relação aos contidos em *Blaue Blume* e *Studio d*, sendo a proporção semelhante à dos *tokens*, ou seja, os textos dos dois primeiros livros possuem aproximadamente 25% mais *types* que os textos dos dois últimos. Novamente, essa informação pode ser útil à professora/ ao professor e pesquisador(a), pois um grande número de *types* implica que a(o)s aluna(o)s necessitam um léxico mais abrangente para compreender satisfatoriamente o texto. Para quantificar essas percepções é preciso empregar a razão entre *types* e *tokens* de um dado texto e pode-se, inclusive, associar o número obtido com o nível GER. Por exemplo, a Iniciativa PASCH⁹ adota a seguinte relação entre as razões *types-tokens* e níveis GER: 0,40 (B1), 0,45 (B2), 0,49 (B2C1), 0,55 (C1).

Para os textos analisados neste artigo, obtemos razões semelhantes entre os quatro LDs, da ordem de 0,70. Esses números (sozinhos) indicariam que os textos seriam adequados para nível C1, o que obviamente não se aplica, pois os materiais analisados estão em lições para o nível A2. Esse fato indica apenas que as análises quantitativas devem sempre se apoiar em diferentes dados, de modo a se evitar resultados enviesados. Um exemplo (extremo) para se entender esse comportamento pode ser obtido ao se

⁸ Boojj (2005) define *types* e *tokens* como: *types*, número de itens lexicais diferentes encontrados no texto e/ou corpus; *tokens*, total de itens lexicais encontrados no texto e/ou corpus.

⁹ <https://www.pasch-net.de/de/pas/cls/leh/unt/dst.html>

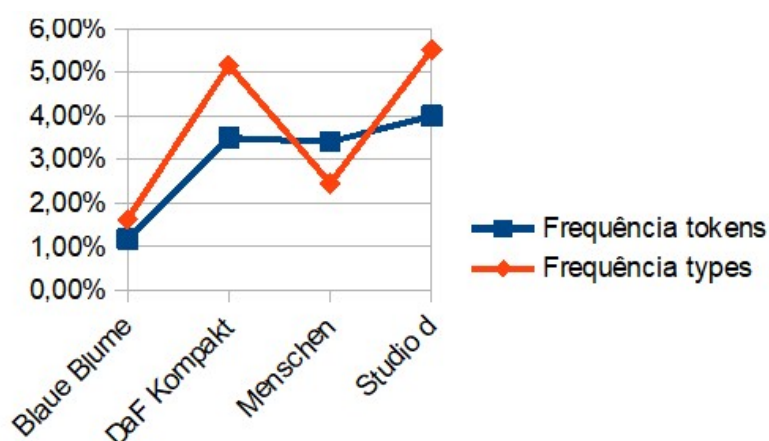
analisar as seguintes frases de nível A1: “*Hallo! Ich bin George. Ich komme aus Brasilien!*”.¹⁰ Nesse caso temos 8 *tokens* e 7 *types* (há repetição do pronome *ich*). Isso implicaria uma razão entre *types* e *tokens* de 0,875, o que indicaria (erroneamente) um texto nível C1.

Tendo em vista que, o conhecimento do número de *tokens* e *types* pode ser útil para a discriminação de textos a serem utilizados em sala de aula, passemos agora a aplicar essas informações no caso da investigação com os pronomes reflexivos. Para tanto, a primeira tarefa será discernir como identificar esses elementos, já que os pronomes possuem mais de uma função verbal, ou seja, “*mich*” pode ser tanto um pronome reflexivo, quanto pronome pessoal no caso acusativo. Para conseguir fazer essa distinção, foi essencial a utilização da ferramenta *Concord* conforme explicado na seção anterior.

Após a identificação dos pronomes reflexivos presentes nos diversos textos, foi possível obter sua frequência de ocorrência (razão entre o número de ocorrência dos pronomes reflexivos e número total de palavras), por meio de uma análise baseada apenas nos *tokens*. Não obstante, essa análise pode ser generalizada para abranger também os *types*. Para tanto, consideramos a estrutura completa de um verbo reflexivo (verbo + pronome reflexivo), identificando como *types* aquelas ocorrências que tenham verbos distintos, independentemente do pronome reflexivo vinculado. A oração “*ich fühle mich ...*” (eu me sinto) e “*ich erinnere mich an*” mais o objeto no acusativo (eu me lembro de) representam, por exemplo, dois *types* com relação à ocorrência de verbos/pronomes reflexivos no texto, em “*ich fühle mich ...*” e “*sie fühlen sich ...*” encontramos um único *type*. Empregando-se essa abordagem, obtemos os resultados esquematizados a seguir:

¹⁰ Tradução própria: “Oi! Eu sou o Georg. Eu sou do/venho do Brasil”.

Gráfico 3 - indicando a frequência de ocorrência para *tokens* (azul) e *types* (vermelhos) para os quatros LDs analisados




Fonte: autoria própria

A frequência de ocorrência dos verbos/pronomes reflexivos, em azul, tem por base *tokens* e, em vermelho, por *types*. Esse gráfico apresenta alguns padrões importantes, como por exemplo, a análise da frequência de ocorrência dos *tokens*. Uma característica marcante é que apenas o LD *Blaue Blume* apresenta uma frequência de ocorrência próxima a 1%, já os demais possuem números próximos de 3%. Uma maior ocorrência de *tokens* pode indicar que houve uma didatização do material, de modo a se conseguir uma frequência artificialmente mais alta, em comparação ao que seria esperado em um texto autêntico. Logo, com os dados obtidos podemos levantar uma primeira hipótese, que o texto do LD *Blaue Blume* terá uma maior autenticidade, ou menor didatização, do que os demais.

Ao considerar os *types*, notamos uma clara discrepância entre *Blaue Blume* e os demais LDs (hipótese de autenticidade), no entanto, *Blaue Blume*, *DaF kompakt* e *Studio d* apresentam um aumento na frequência de ocorrência dos verbos/pronomes reflexivos, *Menschen* tem um decréscimo superior a 25%. Esse dado indica que o texto adotado por *Menschen* possui um grande número de repetições de verbos reflexivos. No caso, há 3 ocorrências de “*sich ängern*”; 2 ocorrências de “*sich fühlen*”; 2 de “*sich erinnern*”; e uma de “*sich freuen*”. Em todos os outros textos temos ocorrências únicas de verbos reflexivos e seus pronomes. Essa informação pode auxiliar a(o) docente na tarefa de se decidir entre textos que abordem o mesmo tema de maneira mais ou menos didatizada, como ocorre no LD *Menschen* com várias repetições do mesmo verbo e pronome reflexivo, ou mais abrangente, como apresentado no *DaF kompakt* e *Studio d* com 8 e 7 verbos reflexivos distintos, respectivamente.


Conforme mencionado anteriormente, a ferramenta *Concord* foi essencial para a distinção entre pronomes reflexivos e pessoais. Entretanto, essa ferramenta também fornece outras informações, que podem ser utilizadas para a escolha de determinado texto a ser empregado em sala de aula. Nesse sentido, apresentamos a seguir os resultados fornecidos pelo *Concord* para os quatro textos analisados:¹¹

Figura 3 - Resultados gerados pela ferramenta *Concord* para o texto do LD *Blau Blume*

 Concord

File Edit View Compute Settings Windows Help

N	Concordance
1	mit Mahmut jetzt ein Jahr zusammen. An dem gefällt mir einfach alles. Aber heiraten will ich nicht. Nie! Das
2	. Ich hab mein Herz noch nie verloren. Er hat mir in der Fleischerei immer mit den schweren Kisten
3	. Jetzt geht ihm das Schreien auf die Nerven. Ich hatte mir das alles anders vorgestellt! Ich würde mein Herz
4	jeder hatte seinen Dickkopf. Als er gegangen ist, hat mir das sehr wehgetan. Seither bin ich vorsichtig


 Concord

File Edit View Compute Settings Windows Help

N	Concordance
1	Monaten war Mandy da. Erst hat er gesagt, er freut sich drauf. Jetzt geht ihm das Schreien auf die Nerven.


Fonte: *WordSmith Tools*

Figura 4 - Resultados gerados pela ferramenta *Concord* para o texto do LD *DaF kompakt*

 Concord

File Edit View Compute Settings Windows Help

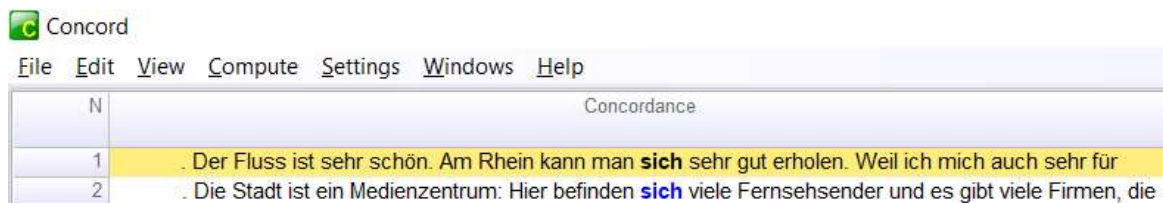
N	Concordance
1	Köln Köln ist eine sehr interessante Stadt. Ich habe mich schon ein bisschen in diese Stadt verliebt! Am
2	. Am Rhein kann man sich sehr gut erholen. Weil ich mich auch sehr für moderne Kunst interessiere, habe
3	und interaktive Videospiele. Als Computerfan freue ich mich schon sehr auf die Messe. Dann kann ich die
4	Besucher auf der Messe – Wahnsinn! Also, ich fühle mich wirklich wohl in Köln. Ich kann mir auch jetzt

 Concord

File Edit View Compute Settings Windows Help

N	Concordance
1	wir eine Schiffstour auf dem Rhein gemacht, das hat mir sehr gut gefallen. Der Fluss ist sehr schön. Am
2	in einem Jahr verschiedene Ausstellungen. Ich habe mir dieses Mal viele Bilder der Expressionisten
3	Also, ich fühle mich wirklich wohl in Köln. Ich kann mir auch jetzt schon vorstellen, hier länger als nur für

¹¹ Por motivo de espaçamento optamos por apresentar apenas os trechos das orações mais relevantes. O texto completo pode ser encontrado no anexo ao final do artigo.



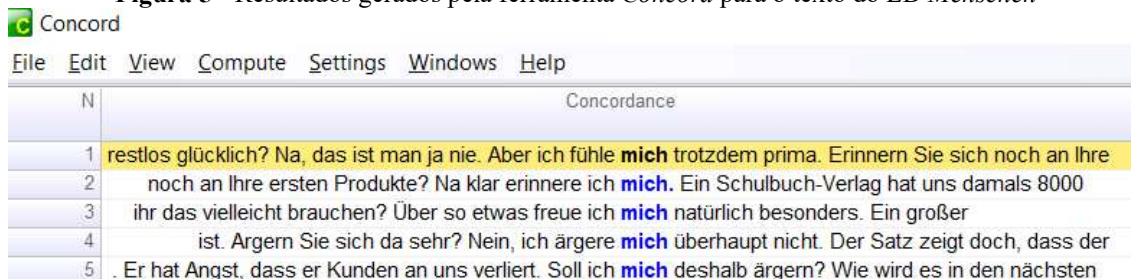
Concord

File Edit View Compute Settings Windows Help

N	Concordance
1	. Der Fluss ist sehr schön. Am Rhein kann man sich sehr gut erholen. Weil ich mich auch sehr für
2	. Die Stadt ist ein Medienzentrum: Hier befinden sich viele Fernsehsender und es gibt viele Firmen, die

Fonte: *WordSmith Tools*

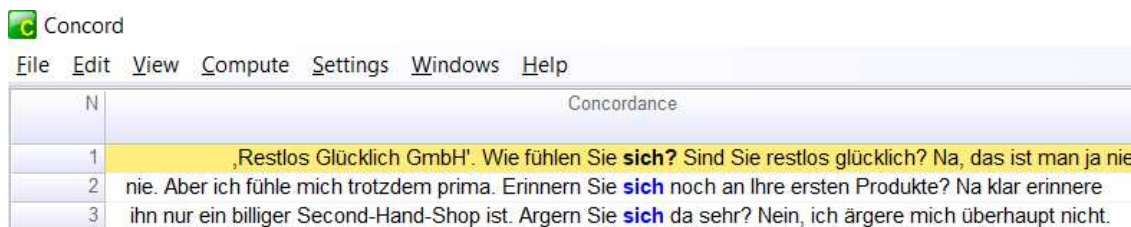
Figura 5 - Resultados gerados pela ferramenta *Concord* para o texto do LD *Menschen*



Concord

File Edit View Compute Settings Windows Help

N	Concordance
1	restlos glücklich? Na, das ist man ja nie. Aber ich fühle mich trotzdem prima. Erinnern Sie sich noch an Ihre
2	noch an Ihre ersten Produkte? Na klar erinnere ich mich . Ein Schulbuch-Verlag hat uns damals 8000
3	ihr das vielleicht brauchen? Über so etwas freue ich mich natürlich besonders. Ein großer
4	ist. Ärgern Sie sich da sehr? Nein, ich ärgere mich überhaupt nicht. Der Satz zeigt doch, dass der
5	. Er hat Angst, dass er Kunden an uns verliert. Soll ich mich deshalb ärgern? Wie wird es in den nächsten



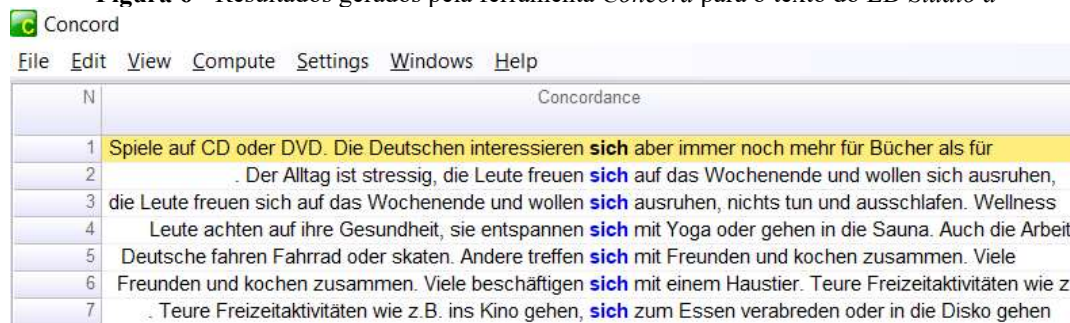
Concord

File Edit View Compute Settings Windows Help

N	Concordance
1	„Restlos Glücklich GmbH“. Wie fühlen Sie sich ? Sind Sie restlos glücklich? Na, das ist man ja nie
2	nie. Aber ich fühle mich trotzdem prima. Erinnern Sie sich noch an Ihre ersten Produkte? Na klar erinnere
3	ihn nur ein billiger Second-Hand-Shop ist. Ärgern Sie sich da sehr? Nein, ich ärgere mich überhaupt nicht.

Fonte: *WordSmith Tools*

Figura 6 - Resultados gerados pela ferramenta *Concord* para o texto do LD *Studio d*



Concord

File Edit View Compute Settings Windows Help

N	Concordance
1	Spiele auf CD oder DVD. Die Deutschen interessieren sich aber immer noch mehr für Bücher als für
2	. Der Alltag ist stressig, die Leute freuen sich auf das Wochenende und wollen sich ausruhen,
3	die Leute freuen sich auf das Wochenende und wollen sich ausruhen, nichts tun und ausschlafen. Wellness
4	Leute achten auf ihre Gesundheit, sie entspannen sich mit Yoga oder gehen in die Sauna. Auch die Arbeit
5	Deutsche fahren Fahrrad oder skaten. Andere treffen sich mit Freunden und kochen zusammen. Viele
6	Freunden und kochen zusammen. Viele beschäftigen sich mit einem Haustier. Teure Freizeitaktivitäten wie z
7	. Teure Freizeitaktivitäten wie z.B. ins Kino gehen, sich zum Essen verabreden oder in die Disko gehen

Fonte: *WordSmith Tools*

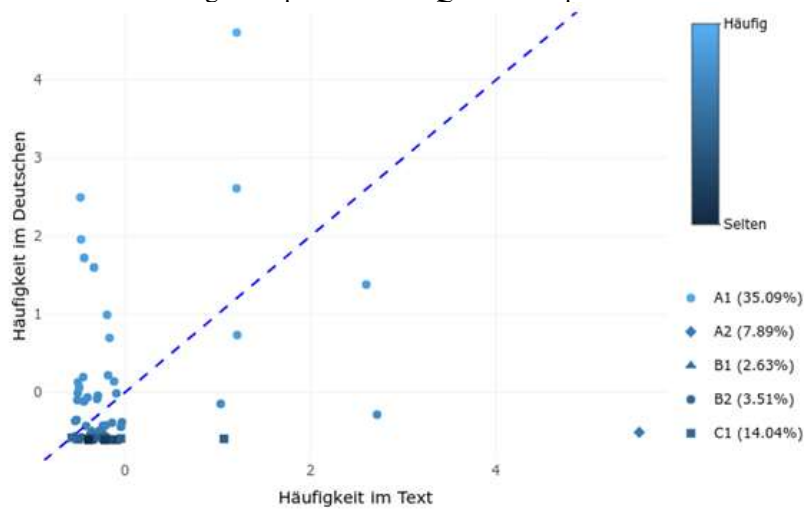
Como podemos perceber, o LD *Blaue Blume* apresenta apenas duas orações com pronomes reflexivos, sendo que em uma delas existe um distanciamento no posicionamento entre o pronome e o verbo. Esse resultado é semelhante em *DaF kompakt*. Já nos LDs *Menschen* e *Studio d* quase todas as ocorrências são introduzidas por meio de

*chunks*¹² fixos na oração. Tal observação ilustra que ao utilizar textos como os apresentados no *Blaue Blume* e *DaF kompakt*, a(o) aluna(o) é exposta(o) a estruturas gramaticais mais complexas, mais próximas da linguagem em uso. No caso dos textos retirados dos LDs *Menschen* e *Studio d*, a linguagem mostra-se truncada e simplificada, indicando uma maior didatização do material.

Em resumo, ao empregarmos a ferramenta *Concord* do *WordSmith Tools* no estudo dos quatro textos, podemos levantar a hipótese de que o texto do LD *Blaue Blume* é o menos didatizado de todos tendo por base a menor frequência de ocorrência da área de interesse de pesquisa. Além disso, percebemos que os insumos de *Menschen* e *Studio d* abordam apenas estruturas com *chunks*, o que torna a sua abordagem mais simplificada.

Para verificar (qualitativamente) a hipótese de autenticidade levantada, empregaremos o software *QuAX-DaF*, já que ele busca realizar uma comparação entre frequência de ocorrência de cada palavra no texto alvo (*Häufigkeit im Text*) e a frequência de ocorrência esperada no *Corpus LCC* (*Häufigkeit im Deutsch*), conforme apresentamos na seção anterior. A análise com o *QuAX-DaF* oferece os seguintes resultados:

Gráfico 4 - Gráfico como gerado pelo software *QuAX-DaF* para o texto do LD *Blau Blume*



Fonte: *QuAX-DaF*

¹² Ensino de estruturas e palavras por meio da memorização com orações e exemplos prototípicos (THURMAIR 2010: 3).

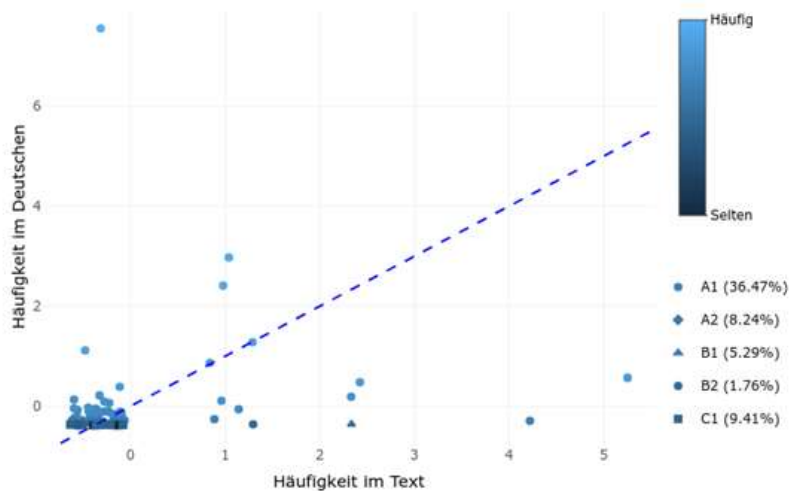
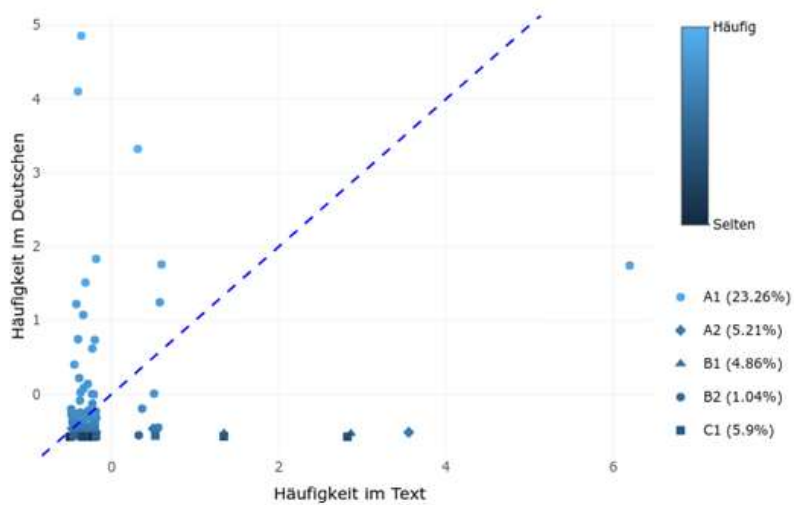
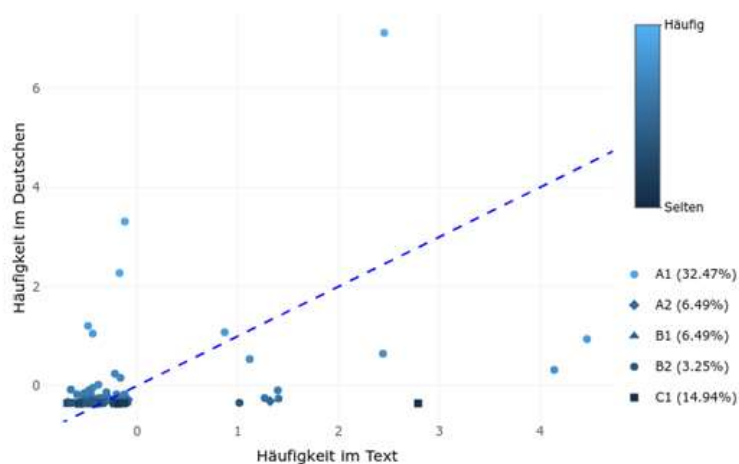
Gráfico 5 - Gráfico como gerado pelo software *QuAX-DaF* para o texto do LD *DaF kompakt*Fonte: *QuAX-DaF***Gráfico 6** - Gráfico como gerado pelo software *QuAX-DaF* para o texto do LD *Menschen*Fonte: *QuAX-DaF*

Gráfico 7 - Gráfico como gerado pelo software *QuAX-DaF* para o texto do LD *Studio*

Fonte: *QuAX-DaF*

Primeiramente, é necessário discutir sobre os critérios de identificação de autenticidade, segundo a ferramenta. Um primeiro critério (critério I) seria notar quão dispersos estão os pontos com relação à reta ideal. A reta é dita ideal, pois os pontos que estão sobre ela têm a mesma frequência no texto de entrada e no *corpus*, ou seja, o texto estaria próximo ao uso autêntico da língua. O critério I será mais preciso quanto maior o texto de entrada for. Um segundo critério (critério II) a se adotar é a quantidade de pontos que ficam abaixo da reta ideal, ou seja, um maior número de pontos nessas condições implica que o texto apresenta várias palavras com frequência superior àquela esperada no *corpus*. Esse critério deve ser aplicado de maneira complementar ao anterior.

Sendo assim, classificaremos os textos, analisados neste artigo com relação a sua autenticidade, tendo por base a dispersão (critério I) de pontos que estejam abaixo da reta ideal (critério II) e que possuam frequência superior a 0 (nas unidades definidas pela ferramenta).¹³ Essa abordagem busca otimizar a análise. Uma descrição quantitativa mais elaborada foge ao escopo da pesquisa atual, mas será desenvolvida futuramente.

Na tabela 2 apresentamos, para cada texto, o número de pontos abaixo da reta ideal e com frequência superior a 0 (X), o número de *types* (Y), e a razão entre X e Y. Primeiramente, é possível notar que o texto em *Blaue Blume* apresenta o menor número de pontos abaixo da reta ideal, tanto em números absolutos (6), quanto relativos (0,05). De acordo com o critério II, esse fato corroboraria a hipótese que esse insumo é o que

¹³ A ferramenta *QuAX-DaF* adota uma escala logarítmica, logo, é possível encontrar frequências negativas. Maiores detalhes podem ser encontrados no manual.

AQUINO, M.; CHERCHIGLIA, A. – Textos didatizados e não didatizados

apresenta maior autenticidade, ou não didatização. Para os demais textos, analisando-se os números absolutos e relativos, percebemos que *Studio d* e *Menschen* apresentam os textos mais didatizados, também corroborando as análises anteriores. Vejamos agora quais informações podem ser obtidas ao considerar o critério I.

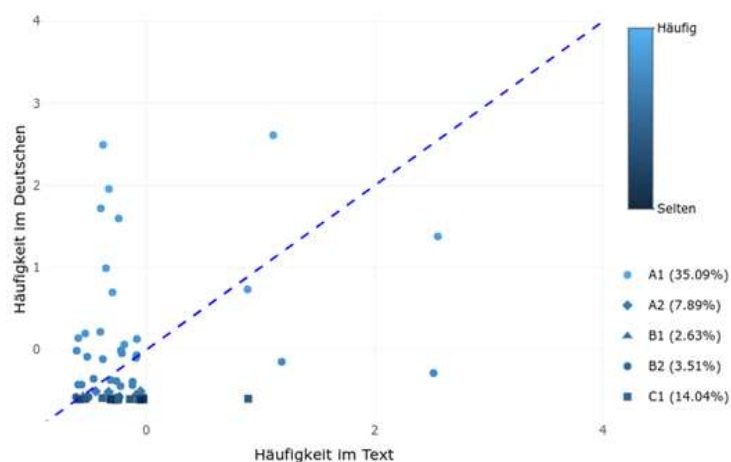
Tabela 2 - Número de pontos abaixo da reta ideal com frequência superior a 0 para os textos dos quatro LDs

Material Didático	Número de pontos abaixo da reta ideal e com frequência superior a 0 (X)	Número <i>types</i> (Y)	Razão X/Y
<i>Blaue Blume</i>	6	123	0,05
<i>DaF kompakt</i>	9	155	0,06
<i>Menschen</i>	12	163	0,07
<i>Studio d</i>	10	127	0,08

Fonte: autoria própria

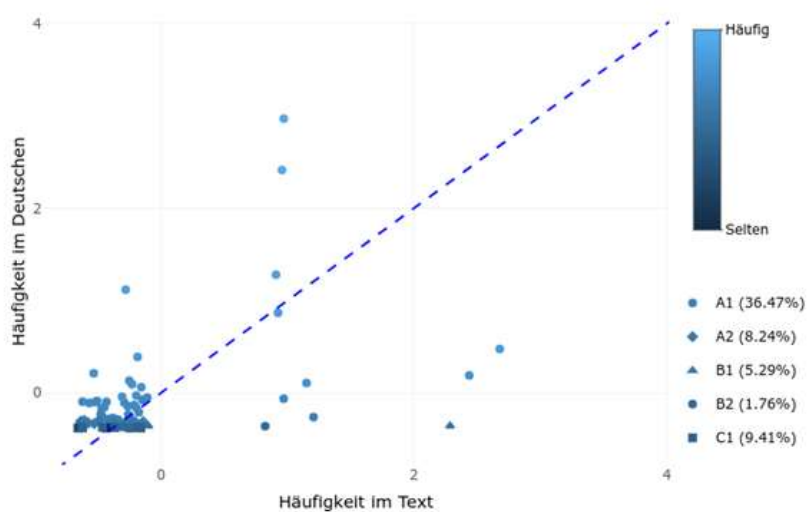
Dessa forma, a primeira tarefa será adotar uma convenção com relação ao que entendemos por dispersão. Como adotamos uma abordagem qualitativa, buscaremos identificar apenas visualmente qual(is) textos apresentam pontos mais distantes da reta ideal. No sentido de facilitar esse processo, apresentamos a seguir os quatro gráficos novamente, onde agora adotamos a mesma escala.

Gráfico 8 - Gráfico com frequência máxima igual a 4, para efeitos de comparação. Gerado pelo software *QuAX-DaF* para o texto do LD *Blau Blume*



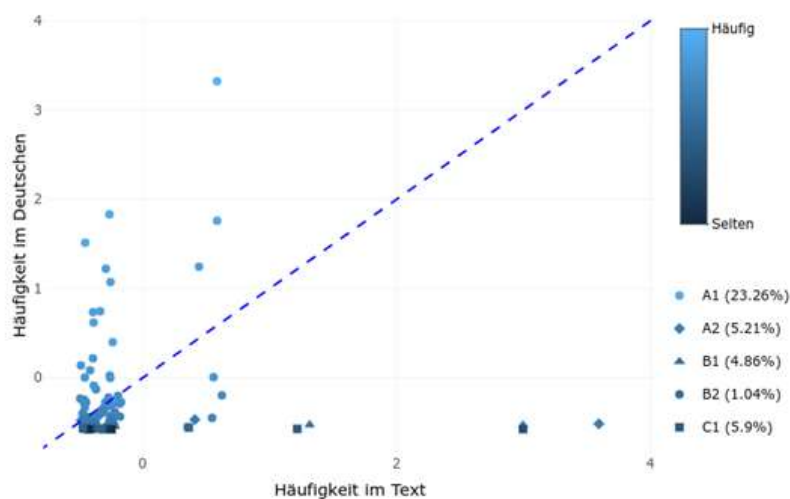
Fonte: *QuAX-DaF*

Gráfico 9 - Gráfico com frequência máxima igual a 4, para efeitos de comparação. Gerado pelo software *QuAX-DaF* para o texto do LD *DaF kompakt*



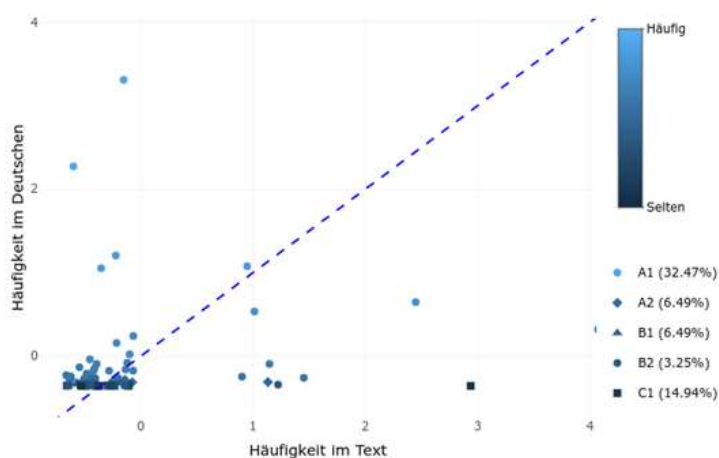
Fonte: *QuAX-DaF*

Gráfico 10 - Gráfico com frequência máxima igual a 4, para efeitos de comparação. Gerado pelo software *QuAX-DaF* para o texto do LD *Menschen*



Fonte: *QuAX-DaF*

Gráfico 11 - Gráfico com frequência máxima igual a 4, para efeitos de comparação. Gerado pelo software *QuAX-DaF* para o texto do LD *Studio d*



Fonte: *QuAX-DaF*

Inicialmente, podemos observar que os pontos abaixo da reta ideal para *Menschen* e *Studio d* estão majoritariamente concentrados ao longo do eixo horizontal, ou seja, para esses dois textos há uma alta dispersão dos pontos o que, de acordo com o critério I, indica grande didatização. Tal fato corrobora, uma vez mais, que tais textos são os mais didatizados em relação ao *corpus* investigado. Já os gráficos dos textos de *Blaue Blume* e *DaF kompakt* indicam que a tendência dos pontos é semelhante em ambos, embora *DaF kompakt* apresente uma concentração ligeiramente superior de pontos mais afastados da reta ideal. Esse fato indicaria, também segundo o critério I, que o texto de *Blaue Blume* seria o mais autêntico de todos os analisados. Assim, podemos concluir que a análise fornecida por *QuAX-DaF* corrobora nossa hipótese, que o texto do LD *Blaue Blume* é o mais autêntico, ou não didatizado, do *corpus* de trabalho.

Gostaríamos de finalizar essa seção apresentando um panorama dos resultados encontrados, assim como os métodos utilizados. Inicialmente, fazendo uso do software *Word Smith Tools* conseguimos realizar uma comparação entre a frequência de ocorrência da área de interesse (pronomes reflexivos) dos insumos dos quatro LDs. O texto em *Blaue Blume* apresenta uma frequência relativa bastante inferior aos demais textos do *corpus*. Com esse dado foi possível levantar a hipótese sobre a sua autenticidade/não-didatização. A seguir, utilizamos a ferramenta *Concord*, que permitiu vislumbrar que os textos apresentados em *Menschen* e *Studio d* realizavam uma abordagem por *chunks*, o que novamente poderia indicar uma maior didatização do material, corroborando os resultados obtidos ao analisar apenas a frequência de ocorrência. Além disso, essa

ferramenta se mostrou essencial para distinção entre pronomes pessoais e reflexivos. Um exemplo interessante pode ser encontrado no texto do LD *Blaue Blume* que apresentava várias ocorrências de “*mir*”, mas apenas uma delas tinha a função de pronome reflexivo. Finalmente, empregamos o software *QuaX-DaF*, desenhado especificamente para indicar o nível de autenticidade de textos em alemão. Com tal análise, concluímos que o texto do LD *Blaue Blume* apresentou a menor didatização ao se comparar com os demais, corroborando a hipótese de autenticidade.

Cabe salientar que o objetivo desta pesquisa é oferecer ferramentas para o(a) professora/professor e pesquisador(a) para uma tomada de decisão consciente quanto à utilização de materiais didáticos mais adequados para o seu próprio contexto local. Assim, tendo por base informações já conhecidas – textos empregados no LD *Blaue Blume* são autênticos – fomos capazes de confirmar empiricamente essa afirmação. Ou seja, o essencial, no espírito deste trabalho, não seria defender que o LD *Blaue Blume* apresenta textos autênticos, mas sim demonstrar como ferramentas computacionais podem nos fornecer essa informação. Essa verificação é importante ao propiciar consistência à metodologia aqui desenvolvida, sobretudo caso queiramos estudar/classificar textos que *a priori* não tenhamos conhecimento acerca de sua didatização.

Na próxima seção vamos desenvolver uma discussão final sobre a relação entre os resultados obtidos e as consequências para o ensino de ALA, com enfoque na contribuição deste estudo para a seleção de textos tanto para o trabalho em sala de aula, como para o desenvolvimento de materiais didáticos, como é o caso do projeto *Zeitgeist*.

5 Conclusão

Não há dúvida de que a competência textual é parte importante das habilidades linguísticas e, portanto, os tipos de textos e suas características linguísticas ou culturais específicas devem ser levados em consideração para a escolha e/ou avaliação dos materiais didáticos. Assim, o LD não deve apenas fornecer textos e temas, mas responder ao conjunto de exigências que tange o trabalho linguístico nos campos de pronúncia, léxico, vocabulário, gramática, unindo isso à variedade de textos, exercícios e tarefas.

De modo a apoiar o(a) professora/professor e pesquisador(a) no desenvolvimento e escolha de materiais didáticos mais adequados a seu meio local, demonstramos neste trabalho como ferramentas computacionais podem ser grandes aliadas nessas tarefas. Em

particular, tivemos a intenção de apresentar em detalhes como extrair dados numéricos que identifiquem e corroborem o conhecimento sobre a relação de materiais didatizados e não didatizados para o ensino de línguas. Como objetos de estudos, selecionamos quatro livros didáticos de ALA, dos quais extraímos textos escritos sobre o mesmo tema gramatical (pronomes reflexivos) e adotamos dois *softwares* para realização da análise, *Word Smith Tools* e *QuAX-DaF*. Dentre os livros didáticos selecionados, apenas um deles apresentava textos não didatizados (autênticos) e pudemos demonstrar como essa informação pode ser extraída por meio das ferramentas computacionais. Tal fato é essencial para que tenhamos confiança na metodologia desenvolvida.

Com relação às perspectivas para o ensino e aprendizagem de ALA, a investigação aqui desenvolvida apresenta diversas possibilidades. Dentro do contexto de projetos maiores como o *Zeitgeist*, que almeja criar um LD didático tendo em vista contextos locais, as ferramentas descritas neste artigo poderiam ser empregadas para criação de um banco de dados com textos de diferentes vertentes e matizes, classificados, por exemplo, em torno de sua didatização, apresentação de estruturas gramaticais complexas, adequação para determinado nível GER. Todas essas classificações podem ser realizadas de maneira análoga a que realizamos nesta pesquisa para o tema dos pronomes reflexivos.

Do ponto de vista de projetos individuais, a nossa análise poderia ser empregada para tratar textos não-didatizados que determinado(a) docente tenha intenção de trabalhar em sala de aula, mas que tenha incertezas acerca de sua adequação lexical e/ou eficácia no tratamento de determinado tema gramatical. Assim, de maneira relativamente simples, o(a) professora/professor teria acesso à frequência de ocorrência do tema de seu interesse no texto por ele(a) selecionado e poderia tomar uma decisão consciente da sua implementação (ou não) em sala de aula.

Reiteramos, por fim, a importância do conhecimento de variados meios para compreender e questionar o uso de materiais didáticos em diversos contextos de aprendizagem. Nesse sentido, buscamos discutir sobre algumas ferramentas que podem auxiliar à professora/o professor nessa missão, tendo em vista as escolhas de textos para diferentes necessidades, nível de proficiência, entre outros. Não obstante, muitas pesquisas ainda precisam ser realizadas focando na discussão do trabalho com textos autênticos para o ensino de línguas, buscando, entre outros, encontrar uma relação ideal entre a frequência de ocorrência de aspectos lexicais e gramaticais e, o emprego de materiais didatizados e não didatizados em sala de aula.

Documentos suplementares

Anexo 1

Blaue Blume (2011: 96): Haben Sie jemals Ihr Herz verloren?

A Mein Herz gehört meinem Freund. Ich bin mit Mahmut jetzt ein Jahr zusammen. An dem gefällt mir einfach alles. Aber heiraten will ich nicht. Nie! Das geht immer schief.

B Sein Herz verschenken? Völliger Blödsinn! Na ja, vielleicht einmal. In jungen Jahren, da hatte ich so eine Affäre. Die Frau sah ziemlich gut aus, Liebe auf den ersten Blick, kurz und stürmisch. Aber irgendwie war alles wie ausgedacht.

C Ich hab mein Herz noch nie verloren. Er hat mir in der Fleischerei immer mit den schweren Kisten geholfen. Nach neun Monaten war Mandy da. Erst hat er gesagt, er freut sich drauf. Jetzt geht ihm das Schreien auf die Nerven. Ich hatte mir das alles anders vorgestellt!

D Ich würde mein Herz nie mehr verschenken, ich würde es nur noch verborgen! Er ist weg, weil er noch ein Kind wollte, eine eigene Familie. Ich hatte aber schon eins und jeder hatte seinen Dickkopf. Als er gegangen ist, hat mir das sehr wehgetan. Seither bin ich vorsichtig geworden.

Anexo 2

DaF kompakt (2016: 30): Notizen aus Köln

Köln ist eine sehr interessante Stadt. Ich habe mich schon ein bisschen in diese Stadt verliebt! Am letzten Wochenende haben wir eine Schiffstour auf dem Rhein gemacht, das hat mir sehr gut gefallen. Der Fluss ist sehr schön. Am Rhein kann man sich sehr gut erholen.

Weil ich mich auch sehr für moderne Kunst interessiere, habe ich gestern das Museum Ludwig besucht. Es zeigt in einem Jahr verschiedene Ausstellungen. Ich habe mir dieses Mal viele Bilder der Expressionisten angesehen.

Die Stadt ist ein Medienzentrum: Hier befinden sich viele Fernsehsender und es gibt viele Firmen, die in der Medienbranche arbeiten. Bei jungen Leuten sehr bekannt sind auch die vielen YouTuber, junge Leute, die auf YouTube einen eigenen Kanal haben und dort über ihre Themen reden, Comedy machen oder Tipps geben. Viele von diesen YouTubern leben in Köln, manche sogar in einem Haus. Im nächsten Monat findet hier die Gamescom statt, eine Messe für Computerspiele und interaktive Videospiele. Als Computerfan freue ich mich schon sehr auf die Messe. Dann kann ich die neuesten Spiele kennenlernen. Letztes Jahr waren über 340.000 Besucher auf der Messe – Wahnsinn!

Also, ich fühle mich wirklich wohl in Köln. Ich kann mir auch jetzt schon vorstellen, hier länger als nur für das Studium zu bleiben. Nur mit dem Dialekt habe ich doch meine Probleme – die kölschen Witze verstehe ich leider gar nicht.

Ciao, euer Bernhard

Anexo 3

Menschen A2 (2013: 65): Sind Sie restlos glücklich?

Die Allgemeine: Frau Bauer, seit zehn Jahren sind Sie nun selbstständig mit Ihrem Unternehmen ‚Restlos Glücklich GmbH‘. Wie fühlen Sie sich? Sind Sie restlos glücklich?

Luisa Bauer: Na, das ist man ja nie. Aber ich fühle mich trotzdem prima.

Die Allgemeine: Erinnern Sie sich noch an Ihre ersten Produkte?

Luisa Bauer: Na klar erinnere ich mich. Ein Schulbuch-Verlag hat uns damals 8000 große alte Landkarten geschenkt und wir haben Geschenkpapier und Briefumschläge daraus gemacht.

Die Allgemeine: Ist das normal, dass Sie so einfach Altmaterial von anderen Unternehmen bekommen?

Luisa Bauer: Am Anfang war es nicht leicht, weil ich nur wenige Kontakte hatte. Inzwischen kehne ich aber viele Betriebe. Manche kommen von selbst und fragen: Das soll eigentlich auf den Müll, könnt ihr das vielleicht brauchen? Über so etwas freue ich mich natürlich besonders.

Die Allgemeine: Ein großer Designmöbelhändler hier in der Stadt hat mal gesagt, dass die ‚Restlos Glücklich GmbH‘ für ihn nur ein billiger Second-Hand-Shop ist. Ärgern Sie sich da sehr?

Luisa Bauer: Nein, ich ärgere mich überhaupt nicht. Der Satz zeigt doch, dass der Mann uns als Konkurrenz sieht. Er hat Angst, dass er Kunden an uns verliert. Soll ich mich deshalb ärgern?

Die Allgemeine: Wie wird es in den nächsten Jahren weitergehen? Haben Sie schon neue Ideen?

Luisa Bauer: Oh ja! Zum Beispiel hätte ich gern eine Internetplattform für Firmen wie unsere. Einen internationalen ‚Aus-alt-mach-neu-Markt‘, verstehen Sie? Das wäre doch toll, oder?

Anexo 4

Studio d (2006: 61): Freizeitaktivitäten

Seit den 70er Jahren gehören Fernsehen, Radio hören und Zeitung lesen zu den beliebtesten Freizeitaktivitäten. Heute sind auch die elektronischen Freizeitmedien sehr beliebt, z.B. Musik und Spiele auf CD oder DVD. Die Deutschen interessieren sich aber immer noch mehr für Bücher als für Computerspiele. Der Alltag ist stressig, die Leute freuen sich auf das Wochenende und wollen sich ausruhen, nichts tun und ausschlafen. Wellness ist in: Immer mehr Leute achten auf ihre Gesundheit, sie entspannen sich mit Yoga oder gehen in die Sauna. Auch die Arbeit im Garten ist beliebt und hilft gegen Stress. Ein Trend im dritten Jahrtausend: Auf der einen Seite gibt es mehr Freizeitangebote als früher. Auf der anderen Seite müssen viele Menschen aber sparen. Immer mehr Deutsche fahren Fahrrad oder skaten. Andere treffen sich mit Freunden und kochen zusammen. Viele beschäftigen sich mit einem Haustier. Teure Freizeitaktivitäten wie z.B. ins Kino gehen, sich zum Essen verabreden oder in die Disko gehen sind out. „Die Bundesbürger machen eine neue Erfahrung: Freizeit muss nicht immer Geld kosten“, so Prof. Dr. Horst W. Opaschowski, der Leiter des B.A.T.-Freizeitforschungsinstituts.

Referências bibliográficas

- AQUINO, M. A contribuição da ferramenta de corpus para o ensino das PMs denn, doch, halt e wohl. *Diálogo das Letras*, v. 8, n. 1, 144-158, 2019.
- AQUINO, M. Mudando o ritmo das aulas de alemão como língua adicional por meio de músicas e mídias digitais. *Pandaemonium Germanicum*, v. 24, n. 42, 22-47, 2021.
- ANDRADE E SILVA, M. K. Autenticidade de materiais e ensino de línguas estrangeiras. *Pandaemonium Germanicum*, v. 20, n. 31, 1-29, 2017.
- ARANTES, P. Imagens de aprendizes de ALE em livros didáticos e o disciplinamento dos saberes. *Pandaemonium Germanicum*, v. 21, n. 34, 1-30, 2018.
- BAKHTIN, M. M. *Estética da criação verbal*. São Paulo: Martins Fontes. 1997.
- BOOIJ, G. *The grammar of words: an introduction to linguistic morphology*. Oxford: Oxford University Press, 2005.
- BRAUN, B. *DaF kompakt neu A2*. Deutsch als Fremdsprache für Erwachsene. Stuttgart: Klett, 2016.
- BYBEE, J. Mechanisms of change in grammaticization: the role of frequency. In: BRIAN, D. J.; RICHARD, D. J. (ed.). *Handbook of Historical Linguistics*. Oxford: Blackwell. 2002.
- BYBEE, J. *Language, Usage and Cognition*. New York: Cambridge University Press, 2010.
- CONSELHO DA EUROPA. *Quadro Europeu Comum de Referência para as Línguas – Aprendizagem, Ensino, Avaliação*. Porto: Edições Asa, 2001. Online: <http://www.goethe.de/z/50/commeuro/deindex.htm> (05/08/2020)
- EICHHEIM, H. *Blaue Blume*. Deutsch als Fremdsprache. Kursbuch. Portugiesische Ausgabe. Übersetzung Paulo Oliveira, Susana Kampff Lages. São Paulo: Editora Unicamp, 2011.
- FUNK, H. *Studio d A2*. Kurs- und Übungsbuch. Berlin: Cornelsen, 2006.
- GLAS-PETERS, S. *Menschen A2*. Kursbuch. München: Hueber, 2013.
- GOLDHAHN, D.; ECKART, T.; QUASTHOFF, U. Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA), 2012.
- HALLET, W. Umgang mit Texten und Medien. In: HALLET, W.; KÖNIGS, F. G. (ed.). *Handbuch Fremdsprachendidaktik*. Seelze: Kallmeyer, 2010, 173-177.
- JACH, D. *QuAX-DaF*: Quantitative Analyse von Texten für den Deutsch-als-Fremdsprache-Unterricht. 2019. Online: <https://daniel-jach.github.io/quax-daf/documentation/quax-daf-documentation.pdf> (05/08/2020)
- KRAMSCH, C. *Language and Culture*. Oxford: Oxford University Press, 1998.
- LUCHTENBERG, S. Language Awareness. In: AHRENHOLZ, B.; OOME-WELKE, I. (Org.). *Deutsch als Zweitsprache*. Baltmannsweiler: Schneider Verlag, 2008, 107-117.
- NATION, P; WARING, R. Vocabulary size, text coverage and word lists. In: SCHMITT, N.; MCCARTHY, E. (ed.). *Vocabulary: description, acquisition and pedagogy*. New York: Cambridge University Press, 1997, 6-19.
- NEUNER, G.; HUNFELD, H. *Methoden des Fremdsprachlichen Deutschunterrichts: eine Einführung*. Langenscheidt: Berlin, 1993.
- OLIVEIRA, P.; LADEL L. Zeitgeist: Modelando um projeto editorial com interface digital. *Pandaemonium Germanicum*, no prelo.
- RÖSLER, D. *Deutsch als Fremdsprache*. Eine Einführung. Stuttgart: J. B. Metzler, 2012.
- SCHMIDT, C. Estudo de livros didáticos de língua alemã no ensino superior brasileiro: foco nos aspectos gramáticos. *Revista Expectativa*, v. 16, n. 1, 36-55, 2017.
- SCOTT, M. *WordSmith Tools Version 7.0*. Oxford: Oxford University Press, 2016.

AQUINO, M.; CHERCHIGLIA, A. – Textos didatizados e não didatizados

THURMAIR, M. Alternative Überlegungen zur Didaktik von Modalpartikeln. *Deutsch als Fremdsprache*, v. 47, n. 1, 3-9, 2010.

TSCHIRNER, E. Der rezeptive Wortschatzbedarf im Deutschen als Fremdsprache. Edição de T. Studer, I. Thonhauser & E. Peyer. In: STUDER, T.; THONHAUSER, I.; PEYER, E. (Hrsg.). *Akten der XVI. Internationalen Deutschlehrertagung (IDT)*. Berlin: Erich Schmidt, 2019, 98-111.

Recebido em 5 de agosto de 2020

Aceito em 7 de janeiro de 2021