

LINGUÍSTICA DE *CORPUS* E TERMINOLOGIA BILÍNGUE: O PROGRAMA *ANTCONC* E A EXTRAÇÃO DE TERMOS EM ALEMÃO

Corpus Linguistics and Bilingual Terminology: The *AntConc* Program and Term Extraction in German

Cristina ALBERTS-FRANCO (Faculdade de Tecnologia de São Paulo, São Paulo, Brasil)

Resumo:

Este artigo aborda três pesquisas envolvendo Linguística de Corpus e Terminologia Bilíngue Alemão/Português com uso do Programa AntConc. Na primeira, objetivou-se extrair termos de Economia em jornais eletrônicos alemães; na segunda e na terceira, o objetivo foi a extração de termos da cultura material indígena em obras de Koch-Grünberg para elaboração de glossário bilíngue. Nas duas primeiras, utilizando-se das ferramentas Word List, Concordance e File View, extraiu-se em pouco tempo um expressivo número de termos cuja frequência foi estabelecida de maneira confiável. A terceira pesquisa, em fase inicial, deverá ampliar os resultados da segunda, atualmente com 257 termos e suas traduções.

Palavras-chave: Linguística de *Corpus*, Terminologia, Programa *AntConc* e Língua Alemã.

Abstract:

This paper discusses three researches involving Corpus Linguistics and Terminology German/Portuguese using program AntConc. The first aimed to extract terms related to Economics in German electronic newspapers. The second and third aimed to extract terms related to the Indian material culture in works by Koch-Grünberg; the purpose: compile a bilingual glossary. The first and second researches, using tools Word List, Concordance, and File View, shortly allowed extraction of large number of terms, whose frequency was reliably established. The third, in initial phase, will allow us to enlarge the results produced with the second, presently with 257 terms and their translation.

Keywords: Corpus Linguistics, Terminology, *AntConc* Program and German Language.

1. Introdução

Este artigo tem como objetivo descrever três pesquisas terminológicas bilíngues envolvendo as línguas alemã e portuguesa e o uso do programa *AntConc* para extração de termos nesses idiomas. No primeiro caso, a finalidade foi extrair termos em Alemão de textos da área de Economia que passariam a fazer parte de um banco de dados multilíngue dessa especialidade; no segundo e terceiro casos, a finalidade foi a extração, em Alemão e em Português, de termos relativos a objetos indígenas do Brasil presentes em obras escritas pelo etnólogo alemão Theodor Koch-Grünberg e em suas traduções para o português para posterior elaboração de um glossário bilíngue alemão/português desses objetos.

No entanto, apesar de, nos dois primeiros casos, ter sido adotado o mesmo programa computacional e metodologia semelhante, os resultados obtidos em uma e na outra pesquisa revelaram evidentes diferenças qualitativas. As causas de tais diferenças serão aqui apresentadas e comentadas, bem como os motivos que me levaram a adotar de novo o *software AntConc* na terceira pesquisa, atualmente em fase de montagem do *corpus*. Antes, porém, de abordar as pesquisas acima mencionadas, considero relevante esclarecer os conceitos de Terminologia, termo e Linguística de *Corpus*, assim como a inter-relação entre Terminologia e Linguística de *Corpus*.

A Terminologia estuda de modo empírico o uso de termos dentro de um contexto pré-estabelecido. Segundo Pavel e Nolet, em sentido restrito, a Terminologia constitui-se em uma “disciplina lingüística consagrada ao estudo científico dos **conceitos** e **termos** usados nas línguas de especialidade” (PAVEL e NOLET, 2002: xvii, grifo das autoras). Ainda de acordo com as autoras, os termos, ou unidades terminológicas, podem ser definidos como “palavras (...) que designam um conceito de uma área específica.” (PAVEL e NOLET, 2002, p. 131).

O levantamento do léxico é a base da pesquisa terminológica, e os textos coletados devem atender a critérios previamente estabelecidos que estejam de acordo com os objetivos da pesquisa. A pesquisa em Linguística de *Corpus*, utilizando-se de programas computacionais, possibilita a coleta, o armazenamento e a análise desses dados. Segundo Berber Sardinha:

“A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados lingüísticos textuais coletados criteriosamente, com o

propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador” (BERBER SARDINHA, 2004, p. 3).

Há muito a Terminologia faz uso de grande quantidade de textos pelo simples fato de que estes são a fonte primária da pesquisa terminológica. No entanto, com o surgimento dos computadores, aliado ao desenvolvimento da Linguística de *Corpus*, a coleta, o armazenamento e a análise de grande quantidade de dados tornou-se mais confiável. Além disso, como afirmam Bowker e Fisher;

O “desenvolvimento tecnológico tem permitido tamanho progresso na produtividade, qualidade e acessibilidade, que se tornou virtualmente impensável àqueles que fazem pesquisa terminológica não tirar proveito do amplo leque de recursos e ferramentas eletrônicos à sua disposição.” (BOWKER e FISHER, 2012, p. 1)¹.

Além das vantagens acima mencionadas por Bowker e Fisher, outra importante contribuição da Linguística de *Corpus* à pesquisa terminológica, quer seja ela monolíngue, bilíngue ou multilíngue, são os quatro pré-requisitos para formação de um *corpus* computadorizado elencados por Berber Sardinha:

- “1) O corpus deve ser composto de textos autênticos, em linguagem natural. [...];
- 2) Autenticidade dos textos subentende textos escritos por falantes nativos. [...];
- “3) O conteúdo do corpus deve ser escolhido criteriosamente. Os princípios da escolha dos textos devem seguir, acima de tudo, condições de naturalidade e autenticidade. Mas devem também obedecer a um conjunto de regras estabelecidas por seus criadores de modo que o *corpus* coletado corresponda às características desejadas. [...];
- 4) Representatividade. [...] (BERBER SARDINHA, 2004, p. 19).”.

Ao atender a esses pré-requisitos, em combinação com os recursos oferecidos pelos programas computacionais desenvolvidos para se fazer análise linguística, a pesquisa

¹ Technological “developments have permitted such improvements in productivity, quality, and accessibility that it has become virtually unthinkable for those carrying out terminological research not to take advantage of the wide range of electronic resources and tools available to them.” (Tradução minha).

terminológica dispõe de condições que favorecem muito a produção de resultados confiáveis. Foram essas condições que beneficiaram as três pesquisas envolvendo Linguística de *Corpus* e Terminologia bilíngue Alemão/Português que passo agora a descrever.

2. Orientação de Pesquisa de Iniciação Científica (FATEC-SP)

No primeiro semestre de 2009, tendo ingressado havia pouco tempo como Docente de Língua Alemã no Curso de Automação de Escritórios e Secretariado (AES) da Faculdade de Tecnologia de São Paulo (FATEC-SP), fui convidada a participar de um projeto de pesquisa no âmbito da terminologia de Economia envolvendo quatro idiomas: Alemão, Francês, Inglês e Português. O objetivo do projeto era a elaboração de um glossário multilíngue de termos de Economia que iria contemplar esses quatro idiomas, sendo que para cada idioma haveria um orientador e um aluno desenvolvendo pesquisa em nível de Iniciação Científica, durante a qual seriam extraídos termos de Economia de textos disponibilizados na *Internet* para posterior elaboração do glossário.

Ficou a critério de cada orientador e de seu orientando decidir com quais textos trabalhariam, desde que fossem textos recentes e confiáveis da área de Economia, em formato digital e disponíveis na *Internet* para *download*. Após efetuar pesquisa na *Internet*, minha orientanda e eu optamos por trabalhar com os textos de Economia dos jornais eletrônicos alemães *Die Welt* (<http://www.welt.de>), *Tagesspiegel* (<http://www.tagesspiegel.de>) e *Taz.de* (<http://www.taz.de>). A fim de que o *corpus* tivesse maior amplitude, a coleta foi feita ao longo da pesquisa para não haver grande repetição de termos relacionados a um único fato, tendo sido iniciada em 16/05/2009 e concluída em 26/03/2010, perfazendo um total de cem textos.

A metodologia inicial adotada para a pesquisa com os termos em Alemão consistiu na leitura e análise individual dos textos coletados, os quais eram, então, copiados em um documento em branco do *Word*. A seguir, era feita a extração manual dos termos de Economia, tanto substantivos, quanto adjetivos e verbos. Montou-se um glossário para cada texto; posteriormente, os glossários individuais foram anexados a um glossário geral e os termos foram classificados segundo suas classes de palavras, ou seja, substantivo, verbo e adjetivo. Além disso, os substantivos foram classificados de acordo com o gênero, que em Alemão são três, feminino, masculino e neutro, e receberam a terminação da sua forma no

plural; já os verbos foram classificados em transitivo, bitransitivo ou intransitivo. As classificações adotadas seguiram a tipologia dos dicionários *Michaelis*. Por exemplo:

Substantivo: Abfindung Sf, -en indenização (SF – substantivo feminino, -en – terminação da palavra no plural);

Verbo: einsparen Vtr economizar (Vtr – verbo transitivo direto);

Adjetivo: wertbeständig Adj de valor estável (Adj – adjetivo).

Mais tarde, verificou-se, também manualmente, se havia repetição, ou não, de cada termo encontrado, o qual recebeu uma tradução provisória. Nessa fase inicial da pesquisa, que durou de maio a novembro de 2009, foram analisados quarenta textos em Língua Alemã, o que possibilitou a extração de um total de 636 termos, equivalendo a uma média de 15,9 termos por texto. Dos 636 termos encontrados, 571 eram substantivos, vinte e nove eram verbos e trinta e seis eram adjetivos. No entanto, devido às mudanças adotadas no projeto e que serão comentadas mais adiante, não houve tempo hábil para se consultar um especialista em Economia para revisão final de cada termo em Português.

Alguns meses após o início das pesquisas, os orientadores participantes do projeto, procurando agilizar o processo de coleta e análise dos dados e obter maior confiabilidade, optaram por uma metodologia baseada em *corpora* eletrônicos cujos dados seriam analisados por programa de computador desenvolvido para esse tipo de pesquisa. O objetivo do projeto também foi modificado, passando a ser a elaboração de um banco de dados multilíngue dos termos de Economia destinado à pesquisa pública, ou seja, disponibilizado na *Internet*.

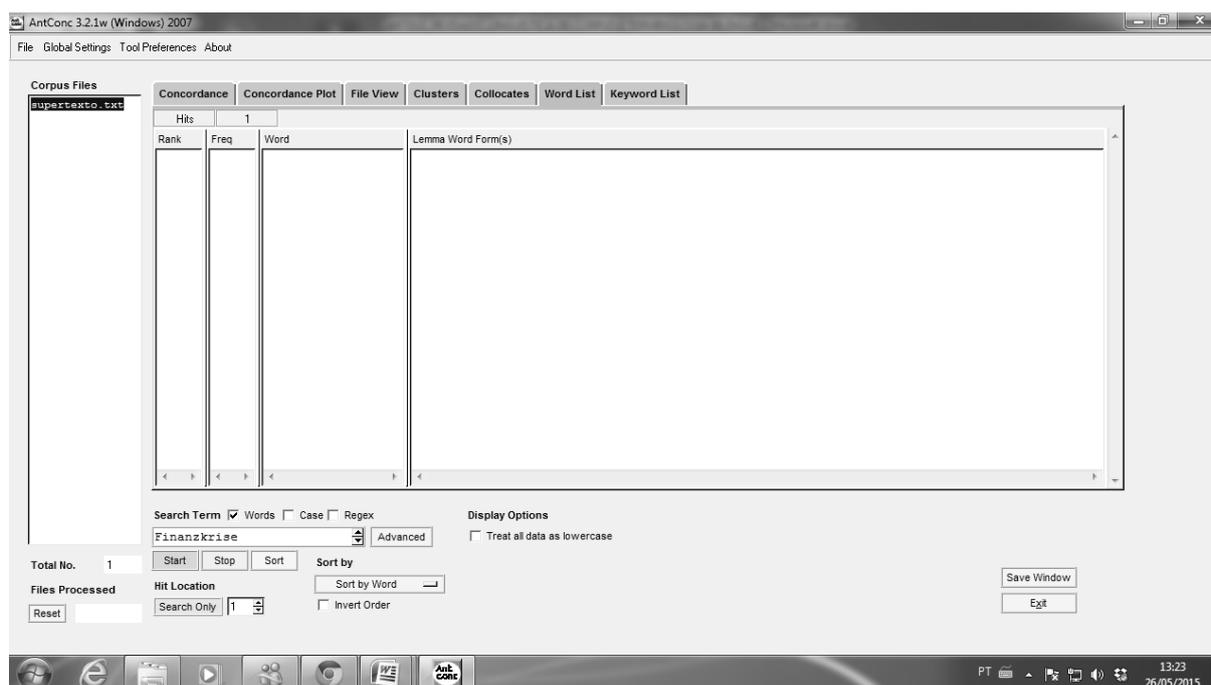
Após análise dos *softwares* que poderiam auxiliar-nos em nossa pesquisa, optou-se pela adoção do programa *AntConc 3.2.1*, que é um *software* de uso livre desenvolvido por Laurence Anthony, PhD em Linguística Aplicada pela Universidade de Birmingham, Inglaterra, e, atualmente, docente na *Faculty of Science and Engineering* da Universidade de Waseda, no Japão (<http://www.laurenceanthony.net>).

O *AntConc* é um programa escrito em *Perl 5.8* utilizando o editor de texto *ActiveState Komodo* e roda em qualquer ambiente *Windows*, bem como em *Macintosh OSX* e *Linux*. As ferramentas que compõem o *AntConc* são: *Concordance*; *Concordance Plot*; *File View*; *Clusters*; *Collocates*; *Word List* e *Keyword List*. *Concordance* gera linhas de concordância de um determinado termo a partir dos textos analisados pelo pesquisador. *Concordance Plot* é

Linguística de *Corpus* e terminologia bilíngue: o Programa *AntConc* e a extração de termos em alemão

um gráfico, semelhante a um “código de barras”, que mostra a distribuição, no arquivo em questão, do termo que se está pesquisando. *File View* localiza, no arquivo com o qual se trabalha, as diferentes ocorrências do termo pesquisado. *Clusters* gera uma lista do termo que se pesquisa em ordem ou de frequência, alfabética, de probabilidade ou de suas terminações. *Collocates* gera uma lista ordenada das palavras próximas ao termo pesquisado. *Word List* gera uma lista, ou em ordem alfabética ou por frequência, de todas as palavras que constam dos arquivos selecionados para a análise linguística. Por fim, *Keyword List* gera uma lista de palavras-chave comparando-se a frequência das palavras do *corpus* de estudo com a frequência das palavras do *corpus* de referência. Nas três pesquisas abordadas neste artigo, as ferramentas utilizadas foram *Concordance*, *File View* e *Word List*. A figura a seguir apresenta a tela inicial do *AntConc* com as ferramentas acima mencionadas.

Figura 1: Tela inicial do programa *AntConc*.



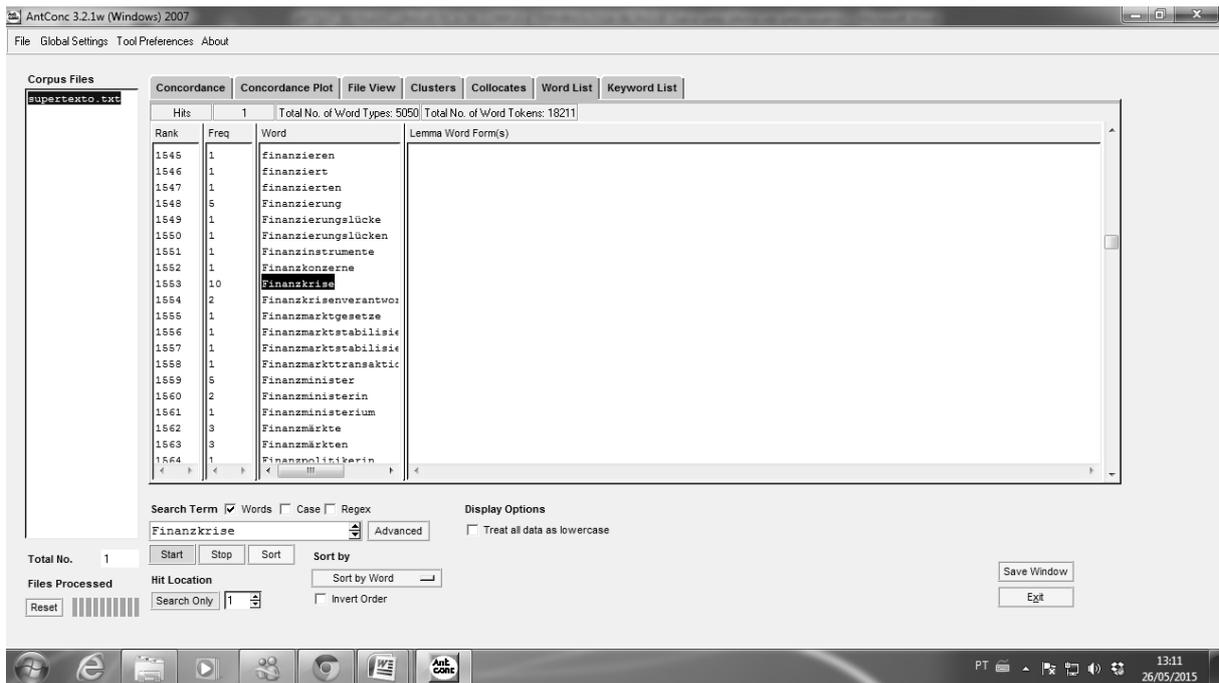
Fonte: Dados da autora.

Para a segunda fase da pesquisa com os termos em Alemão, foram coletados sessenta textos de Economia disponíveis nos jornais eletrônicos anteriormente mencionados, *Die Welt* (<http://www.welt.de>), *Tagesspiegel* (<http://www.tagesspiegel.de>) e *Taz.de* (<http://www.taz.de>). A nova metodologia consistiu em se fazer uma cópia de cada texto coletado, depois as cópias eram copiadas num único arquivo, formando um só grande

documento em branco do *Word*, que foi salvo no formato txt. Com o *AntConc 3.2.1* abria-se esse grande arquivo em txt, e com auxílio da ferramenta *Word List*, verificavam-se os termos individualmente. Caso houvesse dúvida quanto ao emprego de algum termo, recorria-se à ferramenta *Concordance*, que permite que se visualize o termo dentro do contexto da frase. Persistindo alguma imprecisão quanto ao uso do termo dentro do contexto, utilizava-se a ferramenta *File View*, que exhibe, inserida no texto maior, a frase da qual o termo faz parte.

A título de ilustração, apresento, a seguir, na Figura 2, a tela do *AntConc* na janela da ferramenta *Word List* com a lista de palavras em ordem alfabética e na qual aparece, em destaque, o termo *Finanzkrise* (crise financeira), à esquerda, na coluna do meio, aparece a frequência do termo *Finanzkrise* e, mais à esquerda, a coluna *Rank*, que traz o *ranking* de cada palavra do texto pesquisado.

Figura 2: Tela da *Word List* com a palavra *Finanzkrise* em destaque.



Fonte: Dados da autora.

Na segunda e final fase da pesquisa, extraíram-se 500 termos dos sessenta textos em Alemão coletados. No entanto, devido à falta de tempo por conta da adaptação ao novo método e ao término da vigência da bolsa de Iniciação Científica de minha orientanda, não foi possível fazer a revisão dos termos extraídos e estabelecer a distinção por gênero dos substantivos, nem a inclusão da terminação dos substantivos no plural, nem a tradução

provisória dos termos para o português. Como mencionado anteriormente, também não foi possível consultar um especialista em Economia para revisão final de todos os termos em Português, tanto daqueles extraídos na primeira fase da pesquisa, quanto dos que foram extraídos na segunda fase.

Ao final de um ano de pesquisa, tendo utilizado tanto o método manual (primeira fase), quanto o programa *AntConc* (segunda fase), contabilizamos um total de 1.136 termos extraídos, 90% dos quais eram substantivos, 4,5% eram verbos e 5,5 eram adjetivos. 50% dos substantivos eram compostos, ou seja, apresentavam em sua composição a justaposição de dois ou mais elementos, como é o caso de *Geldpolitik* (política monetária), que é a aglutinação de dois substantivos, *Geld* (dinheiro) e *Politik* (política), assim como *Industrieproduktion* (produção industrial), composto por *Industrie* (indústria) e *Produktion* (produção), ou, então, *Immobilienmarkt* (mercado imobiliário), composto por *Immobilien* (imóveis) e *Markt* (mercado), para ficarmos com apenas alguns exemplos. Já os adjetivos e verbos foram encontrados em pequeno número, o que denota sua pouca utilização nos textos coletados. No entanto, seria necessário um prazo maior para que pudéssemos analisar mais detidamente sua presença e importância em textos da área de Economia. Infelizmente, a pesquisa com os termos de Economia em Alemão não teve continuidade, pois os demais alunos que fizeram Iniciação Científica sob minha orientação optaram por realizar pesquisas em outras áreas do conhecimento.

Neste ponto, eu gostaria de fazer algumas observações sobre o uso do programa *AntConc* para extração dos termos em Alemão que explicam, em parte, a demora em nos adaptarmos à nova metodologia. Devido a certas características gramaticais da língua alemã, bem como a algumas convenções ortográficas desse idioma, o uso do programa *AntConc* exige atenção redobrada do pesquisador, além de um conhecimento do Alemão que, via de regra, nossos alunos não possuem. No caso da pesquisa com extração dos termos de Economia, foram duas as maiores dificuldades encontradas para se determinar com confiabilidade esses termos: as declinações e os verbos separáveis.

O Alemão faz parte do grupo das línguas flexivas, em que as relações gramaticais entre as palavras são reveladas ou por meio de afixos, ou por meio da mutação total da palavra, geralmente um pronome pessoal. Em Alemão, há quatro casos que determinam a inclusão, ou não, desses afixos: nominativo, acusativo, dativo e genitivo. Para simplificar, costumo dizer aos alunos que o nominativo é o caso do sujeito da oração, o acusativo é o caso do objeto direto, o dativo é o caso do objeto indireto, e o genitivo revela a relação de posse

entre um/a “dono/a” e seu “objeto possuído”. Como exemplo, vejamos o caso do substantivo masculino “céu”, em Alemão, *der Himmel*.

Nominativo: *Heute ist der Himmel bedeckt.* – Hoje o céu está encoberto.

Acusativo: *Ich sehe den Himmel nicht.* – Não vejo o céu.

Dativo: *In dem Himmel sind viele Wolken.* – Há muitas nuvens no céu.

Genitivo: *Das Blau des Himmels ist schön.* – É belo o azul do céu.

Em Alemão, dependendo do caso gramatical, declinam-se artigos definidos e indefinidos, pronomes pessoais e possessivos, os adjetivos, bem como todos os substantivos no dativo plural, todos os substantivos masculinos e neutros no genitivo singular e um número não muito extenso de substantivos masculinos com algumas terminações específicas no nominativo que são modificadas nos demais casos.

Pois bem, as declinações alemãs não são automaticamente reconhecidas pelo programa *AntConc* como declinações. É fato que os afixos acima mencionados aparecem nas palavras, qualquer que seja a ferramenta utilizada, *File View*, *Concordance*, ou *Word List*, mas é preciso que o pesquisador conte a frequência do termo em cada um dos casos em que ele aparece no texto, nominativo, acusativo, dativo e genitivo, para, no fim, fazer a soma final. Essa contagem, além de demandar mais tempo, exige que o pesquisador tenha um domínio muito bom da sintaxe e do léxico da língua alemã. Quando se trata de pesquisa terminológica, seu domínio da área de especialidade facilita muito a identificação e extração dos termos.

Outra dificuldade apresentada pelo programa *AntConc* para extração dos termos alemães de Economia ocorreu com os chamados verbos separáveis. Em Alemão, existe um grande número de verbos formados por um prefixo e um verbo. Esses são chamados verbos separáveis porque, ao conjugá-los, seus prefixos se “despreendem” e, via de regra, assumem a última posição na frase. Em geral, esses prefixos são preposições ou advérbios. Por exemplo, nos verbos *mitnehmen* (levar junto) e *hingehen* (ir), *mit* é a preposição, *hin* é o advérbio. *Mitnehmen*: *Er **nimmt** seinen Sohn ins Fußballstadion **mit**.* Ele leva seu filho ao estádio de futebol. *Hingehen*: *Er **geht** aber heute nicht **hin**.* Mas hoje ele não irá lá.

Além disso, ao se adicionar um prefixo ao verbo, o sentido original deste último é alterado. Seguem abaixo exemplos de como um prefixo pode alterar o sentido do verbo: *machen* – fazer, produzir, fabricar, criar; *aufmachen* – abrir, apresentar; *zumachen* – fechar,

tapar, abotoar; *anmachen* – colocar, fixar, amassar, preparar, acender, ligar; *ausmachen* – apagar, desligar, combinar, ajustar, avistar, atingir um valor. Dessa maneira, é importante que se leiam atentamente as frases inteiras em Alemão, a fim de se identificar de modo correto o verbo empregado.

Pois bem, como o programa *AntConc* conta as palavras uma por uma, ele acaba por contar um verbo separável cujo prefixo se “desprende” como se se tratasse de dois termos, o verbo e o prefixo, que muitas vezes pode ser confundido com uma preposição. Assim, o pesquisador terá de procurar um desses verbos na *Word List* e encontrar uma das partes do verbo, depois terá de clicar nessa parte do verbo, encontrá-la na ferramenta *Concordance* e clicar novamente na parte do verbo que foi localizada; só então conseguirá visualizar o verbo e seu prefixo no *File View*. Vale, aqui, o que foi comentado anteriormente sobre as declinações: além de essa identificação demandar mais tempo, é preciso que o pesquisador tenha um conhecimento muito bom da língua alemã. No caso da pesquisa terminológica, ele deverá conhecer bem a área de especialidade com a qual está trabalhando.

Além dos problemas com as declinações e os verbos separáveis, atribuo parte das dificuldades que tivemos com o *AntConc* ao fato de ter sido a primeira vez que o utilizamos, ou seja, quando ainda não tínhamos familiaridade com ele, bem como ao fato de termos empregado esse programa na fase final da pesquisa. Além disso, quando estávamos começando a dominar a metodologia, expirou o prazo da bolsa de minha orientanda. No entanto, considero bem-sucedida a segunda fase dessa pesquisa, pois o programa *AntConc* possibilitou a coleta de expressivo número de termos em pequeno intervalo de tempo, permitindo, também, que se estabelecesse a frequência de cada termo de maneira confiável. No geral, considero o saldo final foi positivo e, devido à familiaridade adquirida com o *AntConc* nessa pesquisa, não hesitei em adotar esse mesmo programa, só que em uma versão mais recente, quando realizei meu estágio de pós-doutoramento.

3. Pós-Doutoramento no LAEL da PUC-SP

Pouco mais de um ano e meio após a conclusão da pesquisa que acabo de descrever, foi aprovado pelo Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem da Pontifícia Universidade Católica de São Paulo meu projeto de Pós-Doutoramento intitulado “Elaboração de um glossário bilíngue Alemão/Português dos objetos

indígenas na obra *Do Roraima ao Orinoco*, de Theodor Koch-Grünberg”, a ser realizado sob a supervisão do Prof. Tony Berber Sardinha.

O objetivo do projeto era o levantamento dos termos em Alemão e em Português dos objetos indígenas constantes dos três primeiros volumes da obra *Do Roraima ao Orinoco*, em alemão *Vom Roroíma zum Orinoco* (1916-1928), de autoria do etnólogo alemão Theodor Koch-Grünberg, e a finalidade última dessa pesquisa era a elaboração de um glossário bilíngue dos objetos indígenas nesses três volumes. A metodologia da pesquisa envolveria as áreas da Terminologia e da Linguística de *Corpus*, pois o *corpus* a ser analisado era extenso, e a localização dos termos em Alemão e sua tradução para o Português seriam grandemente facilitadas pelo uso das ferramentas computacionais empregadas nas pesquisas em Linguística de *Corpus*.

No entanto, antes de abordar o desenvolvimento de minha pesquisa de pós-doutoramento, eu gostaria de apresentar um pouco mais de perto o autor com o qual trabalhei e sua obra *Do Roraima ao Orinoco*, bem como minha motivação para realizar essa pesquisa.

Theodor Koch-Grünberg (1872-1924) foi um dos mais destacados etnólogos alemães das duas primeiras décadas do século XX que realizaram pesquisas junto a grupos indígenas da América do Sul, principalmente no extremo norte e extremo noroeste do Brasil. Foram quatro suas viagens ao Brasil, duas das quais expedições por ele chefiadas, cada qual com duração de dois anos e cujos resultados deram origem a obras, hoje consideradas clássicas, sobre os grupos indígenas por ele pesquisados.

Do Roraima ao Orinoco, considerada sua obra-prima, traz em cinco volumes os resultados obtidos na segunda expedição ao Brasil chefiada por Koch-Grünberg, que o levou, de 1911 a 1913, do monte Roraima, na divisa do Brasil, Venezuela e Guiana Inglesa, ao médio curso do rio Orenoco, na Venezuela, depois de subir de barco o rio Urariquera e atravessado a pé e de barco extensa parte do sul da Venezuela.

O primeiro volume da obra *Do Roraima ao Orinoco* traz a descrição da viagem; o segundo volume apresenta mitos e lendas indígenas inéditos naquela época; o terceiro volume apresenta a etnografia dos grupos indígenas Makuxí, Taulipáng, Waíka, Xirianá e Yekuaná; o quarto volume traz listas de palavras de vinte e três línguas e dialetos indígenas, bem como a gramática da língua Taulipáng e o relato em Taulipáng de parte da expedição de Koch-Grünberg; por fim, o quinto volume contém 180 fotografias de representantes indígenas, sozinhos ou em grupos, das etnias pesquisadas pelo etnólogo alemão nessa expedição. Para

nós, brasileiros, a obra *Do Roraima ao Orinoco* tem especial interesse pelo fato de o escritor paulistano Mário de Andrade a ter lido, adotando em *Macunaíma*, o herói sem nenhum caráter importantes personagens e ações descritos e relatados no segundo volume de *Do Roraima ao Orinoco*.

Como parte de meu doutoramento, traduzi integralmente o primeiro volume da obra *Do Roraima ao Orinoco*, ou seja, o relato da expedição. Trata-se de um texto multidisciplinar e interdisciplinar muito extenso, 406 páginas só de texto, excetuando-se os índices, as ilustrações e o prefácio, e cuja tradução, em época anterior à popularização da *Internet*, exigiu minuciosa pesquisa junto a obras de referência sobre astronomia, botânica, cultura espiritual e material de diferentes grupos indígenas, geografia, história, línguas indígenas, toponímia e zoologia.

De todos esses campos do conhecimento, aquele que ofereceu maior dificuldade à tradução foi o da cultura material indígena, ou seja, dos objetos indígenas, e isso pelos seguintes motivos: vários dos objetos mencionados e/ou descritos não constam dos dicionários que consultei a princípio, o bilíngue *Dicionário de Alemão/Português*, da Porto Editora, edição de 1992, e o monolíngue *Deutsches Universal Wörterbuch*, da Duden Verlag, de 1983, nem mesmo o *Dicionário de Etnologia e Sociologia*, de autoria de Herbert Baldus e Emilio Willems, publicado em 1939 pela Editora Nacional; vários objetos são apenas citados em meio à descrição dos locais ou à narração dos fatos, mas não descritos por Koch-Grünberg, que também não indica de que material são feitos, nem o seu uso, nem a sua finalidade; vários objetos recebem mais de um nome em Português nas obras de referência que consultei, como é o caso dos volumes II (*Artesanato Indígena*, 1986) e III (*Arte Índia*, 1986) da *Suma Etnológica Brasileira*, editados por Darcy Ribeiro e coordenados por Berta G. Ribeiro.

Para solucionar essas dúvidas, as obras que mais me auxiliaram na tradução dos objetos indígenas foram o *Dicionário do Artesanato Indígena* (1988), de autoria de Berta G. Ribeiro, e o terceiro volume em alemão da obra *Do Roraima ao Orinoco*, que é ricamente ilustrado e dedica importantes seções à descrição de vários objetos que fazem parte do cotidiano indígena.

Diante das dificuldades encontradas, minha tese de doutoramento (ALBERTS FRANCO, 1999) abordou justamente a tradução dos objetos indígenas no primeiro volume da obra *Do Roraima ao Orinoco*, buscando as causas dessas dificuldades e propondo soluções para elas, bem como apresentando um glossário dos 140 objetos indígenas citados e/ou

descritos nesse volume. Após concluir o doutoramento, vendi minha tradução à EDUNESP, que a publicou em 2006 sob o título *Do Roraima ao Orinoco. Observações de uma viagem pelo Norte do Brasil e Venezuela nos anos de 1911 a 1913. Vol. I.*, e assinei contrato com essa editora para traduzir o segundo e terceiro volumes da mesma obra.

Foi ao longo do trabalho de tradução desses dois volumes, em especial do terceiro volume, que decidi realizar um estágio de Pós-Doutoramento do qual faria parte a elaboração de um glossário bilíngue Alemão/Português dos objetos indígenas nessa obra, pois mais uma vez eu encontrava dificuldade para traduzir adequadamente os objetos indígenas mencionados e/ou descritos nesses dois volumes, em especial do terceiro volume, o qual continha um número significativamente maior de objetos indígenas do que o primeiro volume.

Sabendo que o Prof. Tony Berber Sardinha realiza e orienta pesquisas também em Terminologia, conversei com ele acerca do meu projeto de Pós-Doutoramento, que ele concordou em supervisionar. Assim, apresentei ao Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem da Pontifícia Universidade Católica de São Paulo o meu projeto de pesquisa em nível de Pós-Doutoramento, o qual foi aprovado e realizado no período de março de 2012 a junho de 2014.

Optei por trabalhar somente com os três primeiros volumes da obra *Do Roraima ao Orinoco* e deixar os dois últimos volumes para uma pesquisa posterior, pois até aquela data não havia tradução deles da qual eu poderia extrair os termos em português. Baixei os três primeiros volumes em alemão em pdf do site www.brasiliana.usp.br e os transformei em formato txt, fazendo o mesmo com minha tradução para o português dos respectivos volumes, que eu possuía em formato *Word 2003*. Criei seis arquivos em formato txt: um arquivo para cada volume em Alemão e um arquivo para cada volume traduzido para o Português; a seguir, fiz a limpeza dos seis arquivos em txt.

Originalmente, eu trabalharia com o programa *SynchroTerm*, que deveria alinhar paralelamente, frase por frase, o texto em Alemão e o texto em Português para que eu pudesse localizar e extrair os termos relativos aos objetos indígenas nas duas línguas. No entanto, constatei que esse programa não alinhava satisfatoriamente os dois textos, o que dificultava a localização e extração dos termos.

Após conversar com o Prof. Tony Berber Sardinha, decidimos mudar a metodologia de extração dos termos, bem como o *software* que me auxiliaria na análise dos termos inseridos nos textos em alemão e em português. Primeiro, extraí manualmente os termos

relativos aos objetos indígenas constantes dos três volumes, em ambos os idiomas, o que resultou em duas listas, uma em Alemão e uma em Português, cada qual contendo 257 objetos, vários dos quais apresentavam um ou mais sinônimos. Em cada lista, os objetos foram divididos por categorias, de acordo com sua matéria-prima, ou a finalidade a que se destinam: cerâmica; trançados; cordões e tecidos; adornos plumários; adornos de materiais ecléticos, indumentária e toucador; instrumentos musicais e de sinalização; armas; utensílios e implementos de madeira e outros materiais; e, por fim, objetos rituais, mágicos e lúdicos.

Neste ponto da pesquisa, passei a utilizar o programa *AntConc* 3.2.4, que me foi de grande auxílio, primeiro para estabelecer a frequência de cada termo, depois para verificar se a palavra em questão era mesmo um termo, por fim, para localizar os trechos da obra *Do Roraima ao Orinoco* dos quais seriam extraídos os exemplos, em Alemão e em Português, da presença dos objetos indígenas dentro do contexto da obra. É importante mencionar que, nesta pesquisa, trabalhei exclusivamente com substantivos, pois o glossário deverá conter apenas os objetos em Alemão e em Português, mas não aspectos descritivos a eles referentes, nem as técnicas ou processos com que são fabricados.

Para estabelecer a frequência de cada termo, empreguei a ferramenta *Word List*, clicando na coluna *Freq*; para localizar os termos que efetivamente se referem a objetos indígenas, pois há ocasiões em que ocorre sinonímia, e o termo encontrado diz respeito a outro domínio, empreguei a ferramenta *Concordance*; por fim, para localizar os trechos dos textos em txt de alguns exemplos de objetos indígenas não encontrados pela ferramenta *Localizar* dos programas *Adobe Reader XI* (cópia em pdf da obra em Alemão) e *Word 2003* (tradução dos três volumes), utilizei novamente a ferramenta *Concordance*.

A seguir, os termos e seus sinônimos foram importados para o banco de dados do programa *FileMaker*, o qual permite, por meio da formatação desejada, o armazenamento de todos os dados relevantes que constarão de cada ficha terminológica. Nele foram criados dois bancos de dados, um com os termos em Alemão, o outro com os termos em Português. No total, cada um dos bancos de dados dessa pesquisa contém 257 fichas, cada uma delas apresentando os seguintes dados: o termo em questão, a categoria a que ele pertence, seu gênero gramatical, a frequência tanto do termo, quanto de cada um de seus sinônimos em cada um dos três volumes e, por fim, um pequeno trecho da obra exemplificando a presença do objeto dentro do contexto, bem como a indicação de qual volume da obra e de qual página o exemplo foi extraído. Dessa maneira, os dados coletados durante a pesquisa foram organizados e armazenados no banco de dados do *FileMaker*, o qual, apesar de não se

constituir propriamente no glossário, possui todas as condições para se formatarem os dados e se produzir instantaneamente o glossário. A título de exemplo, apresento, a seguir, a figura da ficha do objeto de número 201 (*Bratrost/moqué*m) em Alemão. Nela é possível visualizar os campos acima mencionados.

Figura 3: Ficha do objeto de número 201 do *FileMaker*.

The screenshot shows the FileMaker Pro interface for a record with ID 201. The main data fields are as follows:

Field	Value
Id	201
Term	Bratrost
Gender	der
Category	utensílios e implementos de madeira e Gebrauchsgegenstände aus Holz und
Equivalent	Bratständer
Gender	der
V1	4
V2	6
Equivalent	Rost
Gender	der
V1	7
V2	1
V3	7

Additional fields shown include: Timestamp: 22/02/2014 07:51:06; Term_Freq: 20+12+13; Example: Unter dem Gerüst brennen die Feuer, stehen die Bratrost mit den Fischen, die sie unterwegs gefangen haben.

Fonte: Dados da autora.

Concluída a coleta e a organização dos dados, passei a pesquisar modelos de glossários junto a obras de referência e artigos sobre Terminologia e Lexicografia, além de ter trocado informações com o Prof. Tony Berber Sardinha e com os colegas do Grupo de Estudos em Linguística de *Corpus* (GELC), do qual passei a fazer parte ao iniciar minha pesquisa de pós-doutoramento. Por fim, optei por montar um glossário que deverá apresentar as seguintes características: será um glossário Alemão/Português e Português/Alemão; os verbetes virão em ordem alfabética; o glossário virá em dois formatos: impresso e digital (ou como aplicativo, ou como *e-book*).

Apresento, a seguir, o modelo escolhido para cada verbete, também elaborado com base nas leituras feitas junto às obras de referência e nas conversas acima mencionadas. A fim de apresentar com mais clareza o aproveitamento dos dados constantes do *FileMaker* acima ilustrados, o objeto escolhido para exemplo é, novamente, o de número 201,

Bratrost/moquém, e faz parte da categoria “Utensílios e implementos de madeira e outros materiais”.

Modelo de verbete:

Bratrost (s.m) moquém, grelha

sin. **Bratständer** (s.m)

sin. **Rost** (s.m)

z.B. „Unter dem Gerüst brennen die Feuer, stehen die Bratrost mit den Fischen, die sie unterwegs gefangen haben.“ (Vol.I, p. 228)

Trad.” Sob o andaime ardem as fogueiras e ficam os moquéns com os peixes que eles pescaram durante o trajeto.” (Vol. I, Cap. 15).

No modelo de verbete escolhido, em caso de sinonímia, os termos virão em ordem decrescente de frequência. Além disso, optei por não incluir a definição de cada objeto, pois, muitas vezes, os objetos indígenas na obra *Do Roraima ao Orinoco* são descrições que não receberam de Koch-Grünberg uma denominação concisa, dispensando, portanto, uma definição. Considero, no entanto, importante a inclusão de frases que apresentem o contexto em que os termos aparecem, pois assim o pesquisador que consultar o glossário terá uma ideia melhor de como os objetos se inserem no cotidiano indígena.

Após concluir a pesquisa realizada no pós-doutoramento, satisfeita com os resultados obtidos e tendo tomado conhecimento da tradução de outras três importantes obras de Koch-Grünberg, decidi não modificar ainda a formatação do banco de dados do *FileMaker* e acrescentar os dados obtidos na continuação do projeto, que se encontra em fase de montagem do *corpus* e que passo agora a descrever.

4. Dando Continuidade à Pesquisa Iniciada no Pós-Doutoramento

Como mencionado anteriormente, no projeto inicial de minha pesquisa de pós-doutoramento, eu pretendia elaborar um glossário que contemplasse somente os objetos indígenas dos três primeiros volumes da obra *Do Roraima ao Orinoco*, pois, no que diz respeito aos termos em Português, eu dispunha apenas de minha tradução desses volumes. No entanto, ao saber da tradução parcial para o Português do quarto e do quinto volumes da obra

Do Roraima ao Orinoco, bem como da tradução recente de três outras importantes obras de Theodor Koch-Grünberg intituladas *Anfänge der Kunst im Urwald* (1905), *Zwei Jahre unter den Indianern* (1909/1910) e *Südamerikanische Felszeichnungen* (1907), optei por dar continuidade ao “Glossário bilíngue dos objetos indígenas”, ampliando-o para as outras obras de Koch-Grünberg já traduzidas para o Português e que não foram contempladas em minha pesquisa de pós-doutoramento.

Em meados de 2013, eu soube que Gabriela Kühnel, aluna de Letras/Alemão da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, havia traduzido as 180 legendas das pranchas fotográficas do quinto volume da obra *Do Roraima ao Orinoco* e estava empenhada na tradução parcial do “Esboço da Gramática do Taulipáng”, que faz parte do quarto volume dessa obra. Tanto as 180 legendas do quinto volume, quanto o trecho traduzido do quarto volume fazem parte do Trabalho de Graduação Individual (TGI) de Gabriela Kühnel intitulado *Subsídios para a Tradução de Excertos da Obra “Do Roraima ao Orinoco” de Theodor Koch-Grünberg* (2014), defendido por ela em dezembro de 2014 e de cuja banca examinadora participei.

Sabendo de minha pesquisa em nível de pós-doutoramento, em comunicação pessoal, Gabriela Kühnel concordou que eu incluísse em meu glossário os objetos indígenas que constam do excerto do “Esboço da Gramática do Taulipáng” por ela traduzido. Decidi, então, ampliar o glossário dos objetos indígenas da obra *Do Roraima ao Orinoco* e incluir os objetos constantes dos volumes IV e V, mesmo aqueles que não fazem parte dos trechos já traduzidos para o Português, pois trata-se, sem exceção, de objetos que já constam do banco de dados do *FileMaker* e, assim, só precisarei verificar a frequência com que são mencionados nesses dois volumes e somá-la à frequência com que aparecem nos demais volumes.

Já as três outras obras mais importantes de Koch-Grünberg mencionadas anteriormente receberam tradução e publicação recentes em português: *Anfänge der Kunst im Urwald* recebeu o título de *Começos da arte na selva* (2009), *Zwei Jahre unter den Indianern* foi publicada em português em 2005 sob o título de *Dois anos entre os indígenas* e *Südamerikanische Felszeichnungen* recebeu o título de *Petróglifos Sul-Americanos* (2010). As duas primeiras traduções foram publicadas pela EDUA (Editora da Universidade Federal do Amazonas) em parceria com FSDB (Faculdade Salesiana Dom Bosco), a terceira foi publicada pelo Museu Paraense Emilio Goeldi em parceria com o Instituto Socioambiental.

Atualmente dedico-me à ampliação de minha pesquisa iniciada no pós-doutoramento, que agora se denomina “Glossário Bilíngue Alemão/Português dos Objetos Indígenas na Obra

de Theodor Koch-Grünberg” e que, como mencionado anteriormente, encontra-se em fase de montagem do *corpus*. A metodologia será a mesma que adotei no pós-doutoramento: coleta dos textos em Alemão e em Português, que deverão ser digitalizados, transformados em formato txt e limpos; será criado um arquivo para cada texto em txt em Alemão e em Português; será feita a extração manual dos termos em Alemão e em Português referentes aos objetos indígenas citados e/ou descritos nessas obras; então, virá a elaboração de duas listas em txt, uma em Alemão e uma em Português, contendo todos os objetos divididos por categorias; por fim, empregando o programa *AntConc* 2.3.4, utilizarei a ferramenta *Word List* para verificar a frequência de cada termo em cada obra e em sua tradução e a ferramenta *Concordance* para verificar se trata-se mesmo de um termo quando houver sinonímia e, também, para localizar e extrair os exemplos em Alemão e em Português dos objetos indígenas presentes nos textos.

Todos esses dados serão importados para o banco de dados do *FileMaker*, em cujas fichas serão criados novos campos com informações relativas às outras obras e volumes não contemplados no pós-doutoramento, bem como a novos termos que provavelmente surgirão. Terminada essa fase, será feita a apresentação final do glossário mudando-se automaticamente a formatação do banco de dados.

5. Considerações Finais

Neste artigo, procurei apresentar minha trajetória como pesquisadora em Terminologia e Linguística de *Corpus* utilizando o programa *AntConc* em três momentos diferentes e em situações também diferentes. A primeira delas foi como orientadora de uma pesquisa em nível de Iniciação Científica em que fui apresentada a esse *software* quando a pesquisa em si já havia sido iniciada empregando-se outra metodologia, tendo, portanto, primeiro de aprender a utilizá-lo, para então dar prosseguimento à pesquisa. Essa contingência, aliada a certas características gramaticais e ortográficas da língua alemã e ao fato de eu ter trabalhado com um léxico que não era de minha especialidade, tornou mais demorado o processo de extração, armazenamento e análise dos dados.

A segunda foi no meu estágio de pós-doutoramento, desta vez sob condições bem mais favoráveis, ou seja, eu já havia adquirido experiência com o *software AntConc*, além do fato de o léxico com que trabalhei ser do meu domínio, pois eu o havia pesquisado intensivamente

ao longo do doutoramento. Além disso, realizei o pós-doutoramento sob a supervisão do Prof. Tony Berber Sardinha, que possui vasta experiência como pesquisador em Linguística de *Corpus* e cujas observações sobre meu trabalho e sugestões me auxiliaram a desenvolver minha pesquisa de forma segura.

Por fim, na pesquisa que desenvolvo atualmente em condições muito parecidas com as do pós-doutoramento, mas com expectativas parcialmente diferentes, pois, apesar de estar trabalhando com textos em Alemão do mesmo autor, o etnólogo Theodor Koch-Grünberg, a tradução de suas outras obras não é de minha autoria. Portanto, a expectativa é de que a tradução dos objetos indígenas nem sempre coincidirá com a minha, assim como espero encontrar termos relativos a objetos que não foram contemplados em minhas traduções e, por isso, não estão incluídos no banco de dados produzido em meu pós-doutoramento, cujos resultados serão, assim, ampliados.

Nessas três situações de pesquisa, apesar das dificuldades enfrentadas na primeira delas, o saldo foi positivo, pois o programa *AntConc* agilizou grandemente a localização dos termos e a confirmação, ou não, de seu emprego, o estabelecimento de sua frequência e, no caso de minha pesquisa iniciada no pós-doutoramento e à qual, atualmente, dou continuidade, a localização e extração dos exemplos em que eles ocorrem no texto, facilitando, portanto, o armazenamento dos dados no banco de dados do *FileMaker* e a posterior elaboração final do *Glossário Bilíngue Alemão/Português dos Objetos Indígenas na Obra de Theodor Koch-Grünberg*.

Recebido em: 04/2015; Aceito em: 06/2015

Referências Bibliográficas

- ANDRADE, Mário de. 1988. *Macunaíma. O herói sem nenhum caráter*. Ed. Crítica. Coord. Telê P. Ancona Lopez. Florianópolis: Ed. da UFSC.
- BALDUS, Herbert & WILLEMS, Emilio. 1939. *Dicionário de Etnologia e Sociologia*. São Paulo: Nacional.
- BERBER SARDINHA, T. 2004. *Linguística de Corpus*. Barueri, SP : Manole.
- BOWKER, L., FISHER, D. 2012. *Technology and Terminology. The Encyclopedia of Applied Linguistics*. Wiley Online Library.
- Dicionário de Alemão/Português*. 1992. Porto: Porto Editora.

- DROSDOWSKI, G.(Hrsg.). 1983. *DUDEN*. Deutsches Universalwörterbuch. Mannheim/Wien/Zürich: Dudenverlag.
- ALBERTS FRANCO, Cristina C. 1999. *Aspectos lingüístico-culturais na tradução dos objetos etnográficos no Vol. I de Vom Roroíma zum Orinoco, de Theodor Koch-Grünberg*. Tese de Doutorado, Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.
- KOCH-GRÜNBERG, T. 1905. *Anfänge der Kunst im Urwald*. Berlin: Ernst Wasmuth.
- _____ 2009. *Começos da arte na selva: desenhos manuais de indígenas colecionados por Dr. Theodor Koch-Grünberg em suas viagens pelo Brasil*. Trad. Pe. Casimiro Beksta. Manaus: EDUA/IGHA.
- _____ 1907. *Südamerikanische Felszeichnungen*. Berlin: Ernst Wasmuth.
- _____ 2010. *Petróglifos Sul-Americanos*. Trad. João Batista Poça da Silva. Org. Edithe Pereira. Belém: Museu Paraense Emílio Goeldi/São Paulo: Instituto Socioambiental.
- _____ 1917. *Vom Roroíma zum Orinoco*. Bd. I. Schilderung der Reise. Berlin: Dietrich Reimer (Ernst Vohsen).
- _____ 1916. *Vom Roroíma zum Orinoco .Bd. II. Mythen und Legenden der Taulipáng- und Arekuná-Indianer*. Stuttgart: Verlag Strecker und Schroeder.
- _____ 1923. *Vom Roroíma zum Orinoco. Bd. III. Ethnographie*. Stuttgart: Verlag Strecker und Schroeder.
- _____ 1928. *Vom Roroíma zum Orinoco. Bd. IV. Sprachen*. Stuttgart: Verlag Strecker und Schroeder.
- _____ 1923. *Vom Roroíma zum Orinoco. Bd. V. Typenatlas*. Stuttgart: Verlag Strecker und Schroeder.
- _____ *Do Roraima ao Orinoco. Vol. I*. Trad. Cristina Alberts-Franco. São Paulo: EDUNESP, 2006.
- _____ 1909/1920. *Zwei Jahre unter den Indianern: Reisen in Nordwest-Brasilien*. Berlin: Ernst Wasmuth.
- _____ 2005. *Dois anos entre os indígenas: viagens ao noroeste do Brasil (1903/1905)*. Trad. Pe. Casimiro Beksta. Manaus: EDUA/FSDB.
- KÜHNEL, Gabriela. 2014. *Subsídios para a Tradução de Excertos da Obra “Do Roraima ao Orinoco” de Theodor Koch-Grünberg*. Trabalho de Graduação Individual, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.
- MICHAELIS. 1999. *Dicionário de Economia e Direito Alemão-Português/ Português-Alemão*. São Paulo: Hemus.
- PAVEL, S. & NOLET, D. 2002. *Manual de Terminologia*. Trad. Enilde Faulstich. Direção de Terminologia e Normalização. Departamento de Tradução do Governo Canadense. Ministério de Obras Públicas e Serviços Governamentais do Canadá.
- RIBEIRO, Berta G. 1988. *Dicionário do Artesanato Indígena*. Belo Horizonte: Edusp/Itatiaia.
- RIBEIRO, Darcy (Ed.). 1986. *Suma Etnológica Brasileira. II Artesanato Indígena*. Coord. Berta G. Ribeiro. Petrópolis: Vozes/finep.

RIBEIRO, Darcy (Ed.). 1986. *Suma Etnológica Brasileira. III Arte Índia*. Coord. Berta G. Ribeiro. Petrópolis: Vozes/finep.

Sites dos Jornais Alemães:

<http://www.tagesspiegel.de>.

<http://www.taz.de>.

<http://www.welt.de>.

Cristina Alberts-Franco teaches German language at FATEC- SP since 2008. She did postdoctoral research in Applied Linguistics under the direction of Professor Tony Berber Sardinha. Her main areas of study and research are: German teaching, German/Portuguese translation, Theodor Koch-Grünberg, Terminology, Corpus Linguistics and Germanic intercultural studies. E-mail address: crisalberts@fatecsp.br.