

ELABORAÇÃO DE UM GLOSSÁRIO DE TRADUÇÃO UTILIZANDO SOFTWARE LIVRE: UM ESTUDO DE VIABILIDADE A PARTIR DE TEXTOS-FONTE EM JAPONÊS¹
CREATING A TRANSLATION GLOSSARY USING FREE SOFTWARE: A STUDY OF ITS FEASIBILITY WITH JAPANESE SOURCE TEXT

Stephen Crabbe
Universidade de Portsmouth (Reino Unido)

David Heath
Universidade de Kanto Gakuin (Japão)

Tradução de Carolina Barcellos
Universidade de Brasília

RESUMO: Neste artigo, nós (a) explicamos como os tradutores podem se beneficiar com a criação de seus próprios glossários; e (b) avaliamos a facilidade com que um glossário de tradução pode ser criado a partir de texto-fonte em japonês usando *software* gratuito. Como mostra nosso estudo, um grande obstáculo decorre do fato de que o texto em japonês não inclui espaços, devendo ser segmentado, ou seja, dividido em “pedaços utilizáveis” (FAHEY, 2016), antes que um concordanciador (no nosso caso, AntConc 3.2.4) possa ser empregado para analisar esse texto e criar um glossário. Segmentamos nosso texto em japonês usando um *software* (ChaSen 2.1) projetado para esse fim. O resultado dessa operação foi problemático, o que nos obrigou a desenvolver soluções trabalhosas e demoradas. Nosso glossário (ver Apêndice 1) é adequado ao propósito que motivou a sua elaboração, mas as dificuldades enfrentadas no processo de criação colocam à prova a viabilidade do uso de *softwares* livres para criar glossários de tradução a partir de textos escritos em japonês.

PALAVRAS-CHAVE: criação de glossário da tradução; texto em japonês; concordanciadores; segmentação de texto; AntConc 3.2.4; ChaSen 2.1.

ABSTRACT: In this paper, we (a) explain how translators can benefit from creating their own glossaries; and (b) evaluate how easily a translation glossary can be created from Japanese source text using free software applications. As our study shows, a major hurdle arises from the fact that Japanese text does not include spaces; it must be segmented, i.e., broken into “usable chunks” (Fahey, 2016), before a concordancer (in our case, AntConc 3.2.4) can be used to analyze it for glossary creation. We segmented our Japanese text using an application (ChaSen 2.1) designed for this purpose. This application’s output was problematic, forcing us to devise workarounds that became labour-intensive and time-consuming. Our completed glossary (shown in Appendix 1) is fit

1 Crabbe, S.; Heath, D. Creating a Translation Glossary Using Free Software: A Study of Its Feasibility with Japanese Source Text. In: *International Journal of English Language & Translation Studies*. 5(3). 2017. Pp. 151-160. Disponível em: <<http://www.eltjournal.org/archive/value5%20issue3/18-5-3-17.pdf>>. Acesso em: 02 mar. 2018.

for purpose, but the complications in the process of creating it call into question the feasibility of using free software to make translation glossaries from text written in Japanese.

KEYWORDS: translation glossary creation; Japanese text; concordancers; text segmentation; AntConc 3.2.4; ChaSen 2.1.

1 Introdução

Neste artigo, partimos da nossa experiência como tradutores profissionais no par linguístico japonês-inglês e como pesquisadores da tradução para (a) explicar como tradutores podem se beneficiar com a criação de seus próprios glossários; e (b) avaliar a facilidade com que um glossário de tradução pode ser criado a partir de um texto-fonte em japonês usando o *software* gratuito AntConc 3.2.4 (ANTHONY, 2014) (um concordanciador) e ChaSen 2.1 (MATSUDA, 2000) (um segmentador para textos em japonês). Nós optamos por uma abordagem baseada em linhas de concordância para a seleção de termos do glossário, em vez de utilizar ferramentas de seleção automática de termos, pois, entre outras coisas, é extremamente “simples” (MUEGGE, 2013) e fornece algum grau de controle importante para abordar desafios “de ruído” (ou seja, candidatos inválidos a termo) e “silêncio” (ou seja, falta de candidatos legítimos a termo) (MUEGGE, 2013).

Para os tradutores (especialmente aqueles que trabalham com textos técnicos ou especializados), a chave para uma tradução de qualidade é a “congruência lexical” (STITT, 2017), ou seja, usar terminologia da língua alvo de maneira consistente. Para ser direto, é importante (e possivelmente essencial) usar sempre o mesmo termo como rótulo para a mesma coisa ou o mesmo conceito (STITT, 2017). Um método que tradutores usam para manter a “congruência lexical” (STITT, 2017) é desenvolver glossários. Então, o que é um glossário?

Um glossário é essencialmente uma lista de termos em um ou mais idiomas. [...] o glossário mais básico simplesmente irá conter listas de termos e seus equivalentes em uma ou mais línguas estrangeiras. [...] No outro extremo do espectro do glossário, encontrará glossários com muitos detalhes contendo definições, exemplos de uso, sinônimos, termos relacionados, notas de uso etc. Esses são os glossários que cada estudante de tradução [...] sonha em ter porque pode usá-los para entender termos, identificar equivalentes, aprender a usar termos [...] (BOWKER e PEARSON, 2002, pp. 137-138).

Um glossário tem algumas semelhanças com um dicionário. No entanto, dicionários geralmente não são tão úteis quanto glossários para traduções que envolvem idiomas para fins específicos/língua especializada. Um defeito dos dicionários “é sua incompletude inerente. O mundo à nossa volta e a linguagem usada para descrevê-lo estão evoluindo o tempo todo, isso significa que dicionários impressos ficam desatualizados muito rapidamente” (BOWKER e PEARSON, 2002, p. 15).

Outra deficiência dos dicionários é o seu tamanho. Bowker e Pearson (2002, p.15) deixam claro que “Embora seja possível compilar dicionários grandes, com vários volumes, na intenção de cobrir um assunto especializado, poucas pessoas vão conseguir

pagar o preço dessas obras e [...] não gostariam de carregá-las por aí”. Por conta de limitações de tamanho, “lexicógrafos que elaboram [...] dicionários têm que decidir quais informações devem ser incluídas e quais devem ser deixadas de fora. Infelizmente, suas escolhas não correspondem às necessidades dos usuários de idiomas para fins específicos” (Bowker & Pearson, 2002, p. 15).

Dicionários também são criticados por não fornecerem suficiente “informação sobre contexto e uso. Os alunos de idiomas para fins específicos/língua especializada devem prestar atenção a como termos são usados, o que significa que, além de informações sobre como usar aquele termo, também precisam de informações sobre como usar aquele termo em uma sentença” (BOWKER e PEARSON, 2002, p. 16). Não obstante, “a maioria dos dicionários [...] não pode fornecer, de forma simples, informações sobre a frequência com que um determinado termo é usado” (BOWKER e PEARSON, 2002, p. 16) mesmo que esse tipo de informação possa facilitar a tomada de decisões mais acertadas a respeito de quais escolhas lexicais podem ser adequadas (BOWKER e PEARSON, 2002, p. 16).

Um glossário criado a partir de um *corpus* [“um corpo de texto” (BOWKER e PEARSON, 2002, p. 9)] escolhido e projetado pelo próprio tradutor pode estar livre das deficiências mencionadas sobre dicionários. Mas como tradutores podem criar seus próprios glossários usando *softwares* livres? E quão complexo seria esse processo se o texto-fonte fosse escrito em japonês?

2 Revisão de Literatura

Lexicografia (atividade de editar e/ou compilar dicionários) foi, em suas origens, um processo lento e penoso. O esforço para definir uma palavra e classificar seus usos envolvia trabalhar com “um tipo de cartão de papel (chamado *citação*), cada um deles consistia de uma passagem citada contendo a palavra em discussão” (LANDAU, 2001, p. 44). A compilação da primeira edição do *Oxford English Dictionary* “levou setenta longos anos de muito trabalho” (WINCHESTER, 2004, p. XXV). E, apesar do esforço envolvido, dicionários baseados em citações eram fundamentalmente falhos. A seleção de conteúdo dependia muito da intuição dos lexicógrafos e estava, portanto, sujeita a seus “preconceitos e preferências” (KRISHNAMURTHY, 2002, p. 23). Além disso, eram inerentemente incompletos. Mesmo o *Oxford English Dictionary* “conseguia ter apenas uma cobertura fragmentada” (KRISHNAMURTHY, 2002, p. 23). Hoje, dicionários impressos ainda sofrem de “incompletude inerente” (BOWKER e PEARSON, 2002, p. 15), e da inclusão de “material linguístico morto” (BOWKER e PEARSON, 2002, p. 15).

A Lexicografia sofreu uma mudança dramática de meados da década de 1980 até meados da década de 1990, devido a aumentos consideráveis no poder de servidores e em também no poder dos discos rígidos dos computadores (LANDAU, 2001, p. 2). Talvez, mais importante ainda, computadores permitiram que lexicógrafos analisassem “conjuntos eletrônicos enormes de linguagem natural (chamados *corpora*, no singular *corpus*, significando “Corpo” em latim)” (LANDAU, 2001, p. 2) e os usassem “para estudar e analisar o uso da linguagem de maneiras que não eram possíveis até então” (LANDAU, 2001, p. 2). *Corpora* eletrônicos podem ser gigantescos. O *Collins Corpus*, por exemplo, contém mais de 4.5 bilhões de palavras (“The Collins Corpus”, 2016).

Um *corpus* eletrônico grande “pode ser muito mais abrangente e equilibrado que a experiência de língua de qualquer indivíduo” (KRISHNAMURTHY, 2002, p. 23). Talvez seu mérito principal seja fornecer evidência objetiva sobre o uso real da língua em termos de “como as palavras são usadas, o que querem dizer, quais palavras são usadas juntas e quantas vezes palavras são usadas” (“The Collins Corpus”, 2016).

Corpora eletrônicos podem ser de grande benefício aos tradutores. Em particular, podem beneficiar os tradutores técnicos, que precisam aprender e replicar o uso real de idiomas para fins específicos, ou seja, “a linguagem que é usada para discutir campos especializados de conhecimento” (BOWKER e PEARSON, 2002, p. 25). Conforme Bowker e Pearson (2002, p. 19) apontam:

Como *corpora* são compostos de textos que foram escritos por especialistas, os alunos de idiomas para fins específicos têm diante de si um corpo de evidências relativas à função, ao uso de palavras e expressões na área de idiomas para fins específicos. Além disso, com o auxílio de ferramentas de análise de *corpus*, é possível analisar contextos de modo que padrões significativos sejam revelados. Ademais, um *corpus* pode fornecer a um aluno de idiomas para fins específicos uma boa ideia sobre como um termo ou uma expressão *não* pode ser usado.

Um *corpus* eletrônico de idiomas para fins específicos/língua especializada e um concordanciador — um programa de computador que permite ao usuário ver cada ocorrência de uma determinada palavra acompanhada de seu contexto imediato na forma de linhas de concordância de palavras-chave em contexto (KWIC) e realizar análises estatísticas no *corpus* — podem permitir que o tradutor crie um glossário de língua especializada como auxílio para a produção de textos na língua alvo os quais estejam de acordo com o uso real de termos no contexto especializado dessa língua alvo.

Ao utilizar o concordanciador para (a) listar as palavras do *corpus* em ordem de frequência e/ou alfabética e (b) produzir, analisar e comparar as linhas de concordância com palavras-chave, o tradutor pode identificar candidatos a termo para o glossário, verificar quais candidatos a termo são termos reais [palavras e/ou expressões “que são usadas em um domínio especializado e ter um significado claro definido” (BOWKER e PEARSON, 2002, p. 145)] de maneira a justificar sua inclusão no glossário e possibilitar reunir exemplos de uso real dos termos. Utilizando as mesmas ferramentas, o tradutor também pode “ganhar informações de ordem conceitual, como o conhecimento sobre as características dos conceitos por trás dos termos e das relações que os conceitos têm um com o outro” (BOWKER e PEARSON, 2002, p. 39). O tradutor pode usar essa informação conceitual para produzir definições dos termos na língua de saída e/ou de chegada, escolhendo combinar informação com o seu próprio conhecimento e/ou com definições em outras fontes, por exemplo, dicionários convencionais de idiomas para fins específicos/língua especializada.

Com alguns programas de processamento de *corpus*, o processo de identificação de candidatos a termo pode ser semiautomatizado por meio de um recurso que identifica “palavras que ocorrem em uma frequência incomum e elevada em um texto ou *corpus* quando esse texto ou *corpus* é comparado a outro *corpus*” (BOWKER e PEARSON, 2002, pp. 114-115) e classifica as palavras “de acordo com sua ‘chavidade’ em vez de frequência” (BOWKER e PEARSON, 2002, p. 115) de forma que “as ‘palavras-chave’ vão direto para o topo da lista” (BOWKER e PEARSON, 2002, p. 115). (Não usamos esse

recurso em nosso estudo, pois criamos nosso glossário usando um único *corpus*.)

Dicionários convencionais monolíngues de língua especializada “tendem a se concentrar em fornecer informações sobre o significado em vez do uso de termos. Por consequência, em geral, não fornecem informação gramatical ou exemplos de uso” (BOWKER e PEARSON, 2002, p. 139). Em dicionários convencionais bilíngues ou multilíngues de língua especializada, “as definições raramente são fornecidas e a ênfase é dada principalmente a equivalentes e exemplos de uso” (BOWKER e PEARSON, 2002, p. 140). Um glossário de língua especializada produzido com o auxílio de um *corpus* eletrônico de língua especializada e um concordanciador pode estar livre de todas essas deficiências e pode, assim, ser muito mais útil. Os benefícios da compilação de glossários são enfatizados por empresas de tradução como *Integro Idiomas* (2017) e *Lionbridge* (2016). Ademais, a compilação de um *corpus* e a criação de um glossário são, como foi destacado pelo *European Graduate Placement Scheme* em suas orientações para estudantes de tradução na pós-graduação em processo de colocação no mercado de trabalho, habilidades práticas fundamentais para provedores serviços de tradução (European Graduate Placement Scheme, s.d.).

3 Metodologia

3.1 Compilação do *Corpus*

O *corpus* que selecionamos para o nosso estudo é o texto-fonte de um dos nossos projetos de tradução, já concluídos, do japonês para o inglês. Trata-se do manual de um produto, lançado em 2009, por uma montadora japonesa que tinha o objetivo de fornecer uma visão geral de um carro – na verdade, uma versão atualizada de um modelo já existente e que a montadora estava se preparando para lançar – aos seus distribuidores no exterior. (Para razões de confidencialidade, excluímos do presente artigo informações que pudessem identificar a montadora.) A data de publicação do manual é recente sugerindo que o *corpus* reflete adequadamente “o estado atual da língua e da área de conhecimento” (BOWKER e PEARSON, 2002, p. 54).

O *corpus* foi escrito por um especialista (redator profissional nativo de língua japonesa com experiência na área automotiva) sob a supervisão editorial de especialistas técnicos no assunto (equipe da sede da montadora responsável por fornecer materiais informativos e de marketing aos distribuidores no exterior). A autoria e a supervisão editorial sugerem que o *corpus* contém “exemplos mais autênticos de uso de língua especializada” (BOWKER e PEARSON, 2002, p. 54) do que conteria se tivesse sido escrito por pessoas que não são especialistas da área. Assumimos, a partir de nossa experiência traduzindo textos similares em língua japonesa, que os usuários do texto alvo também serão especialistas na mesma área de conhecimento.

O *corpus* contém cerca de dezoito mil caracteres. Baseando-se na regra geral dos tradutores do par linguístico japonês-inglês de que quatrocentos caracteres japoneses (número geralmente contido em uma página comum) de texto-fonte correspondem a cerca de duzentas palavras no texto alvo em inglês, o *corpus* seria composto de cerca de nove mil palavras em língua inglesa. Bowker e Pearson (2002, p. 48) afirmam que

corpora variando de cerca de dez mil palavras até centenas de milhares de palavras já se provaram úteis na investigação de língua especializada com base em análises estatísticas de frequência. Segundo essa medida, o tamanho do nosso *corpus* parece ser minimamente aceitável.

O manual do produto contém capítulos sobre o design do carro (ou seja, seu estilo); a dinâmica de condução (motores, transmissões e tecnologias relacionadas à direção, manuseio e qualidade de passeio); qualidade de acabamento (medidas tomadas pela montadora para criar um produto refinado); e segurança. Isso está em conformidade com a recomendação de Bowker e Pearson (2002, p. 49) para usar textos completos (em vez de extratos), a fim de evitar a exclusão acidental de conteúdo útil. No entanto, a amplitude de sua cobertura (o carro inteiro) sugeriu desde o início que o número de vezes que um determinado termo aparece – e o número de contextos em que aparece – pode ser pequeno.

Em parte, devido à nossa experiência com a tradução de textos semelhantes ao texto do nosso *corpus* e, também, à luz de outras literaturas pertinentes (por exemplo, TAKEUCHI, KAGEURA, KOYAMA, DAILLE, e ROMARY, 2003), assumimos, desde o início, que uma quantidade considerável (talvez, a maioria) do conteúdo lexical relevante para a produção do glossário consistiria de substantivos e/ou expressões cujos núcleos são substantivos. Além disso, nosso *corpus* reflete uma forte tendência dos japoneses em omitir sujeitos e deixar que o leitor infira essa informação a partir do contexto. Por exemplo, uma passagem sobre o estilo do carro contém a seguinte frase: 「エクステリアでは、新しいファミリーフェイスを採用しました。」 [Tradução literal: No exterior, [nós] adotamos [uma] nova face familiar.], em que o sujeito omitido pode ser inferido como sendo a própria montadora.

3.2 Expectativas do usuário e elaboração do glossário

Nossas suposições sobre o provável usuário do nosso glossário influenciaram nossos critérios para a seleção de termos e nosso projeto de entradas do glossário.

Nós traduzimos textos técnicos há décadas. Sabemos, por experiência própria, que um tradutor pode ficar sobrecarregado com o trabalho sob a intensa pressão de prazo e, assim, precisar da ajuda de outros tradutores. Nossa experiência também sugere que tradutores nativos de língua inglesa que trabalhem com o japonês como língua de partida e tenham conhecimento da área especializada automotiva são poucos e difíceis de encontrar. Portanto, o usuário alvo do nosso glossário, para efeitos deste estudo, é um tradutor *freelance* nativo do inglês e fluente em japonês com certa inclinação a trabalhos técnicos e interesse pessoal em carros, mas que não está completamente familiarizado com os termos e conceitos principais em textos para distribuidores de montadoras escritos em japonês pelas próprias montadoras japonesas. (Excluímos falantes nativos de japonês do nosso grupo de usuários hipotéticos por duas razões: (1) nossa experiência sugere que sua resposta é mais propensa a ser afetada pelo que Baker (1992, p. 54) chama de “efeito que absorve e amplia padrões do texto-fonte”. (2) A Federação de Tradução do Japão afirma, em seu guia para clientes de tradução, que 「外国語の文書を母国語に翻訳するのがプロの原則です」 [Tradução literal: É um princípio fundamental que tradutores profissionais trabalhem do idioma estrangeiro para o seu idioma nativo.] (Federação de Tradução do Japão, 2012, p. 15.) Tendo em mente a relevância da lista de

frequências produzida por nosso concordanciador, nós nos esforçamos para:

- excluir do glossário qualquer termo para o qual uma tradução literal estaria, mesmo que o tradutor não tivesse uma compreensão extensiva do conceito por trás desse termo, provavelmente correta;
- incluir qualquer termo para o qual uma tradução literal não estaria correta devido, por exemplo, ao uso idiossincrático do termo pela montadora ou pela indústria automobilística japonesa em geral; e
- excluir o que Bowker e Pearson (2002, p. 103) afirmam ser muitas vezes chamado de “vocabulário subtécnico, ou seja, vocabulário que é usado em domínios especializados, mas não exclusivamente em um dado domínio”.

Sabemos, a partir de nossa experiência profissional, que é possível saber o significado de um termo japonês em *kanji* (logogramas de origem chinesa usados na escrita japonesa) mesmo sem conseguir se lembrar de sua pronúncia (ou até mesmo sem saber sua pronúncia). No entanto, conhecer a pronúncia correta pode ser vital para reuniões e chamadas telefônicas relacionadas ao projeto de tradução. Para qualquer termo que inclua *kanji* (com ou sem um verbo auxiliar em *hiragana* – um dos dois silabários japoneses usados em conjunto com *kanji*), adicionamos a pronúncia do termo inteiro em *hiragana* entre parênteses. Partimos do princípio de que o usuário do glossário não precisaria de uma representação romanizada de qualquer termo japonês.

Cada entrada do nosso glossário começa com o termo japonês em questão (mostrado sem uma representação romanizada) e continua com a classe de palavras do termo (por exemplo, substantivo), nosso termo sugerido em inglês, o domínio em que a terminologia é usada, a fonte de nossa informação (na maioria dos casos, nossa própria pesquisa e/ou conhecimento, representado por nossas iniciais combinadas, SCDH), e um exemplo de contexto em que o termo em japonês ocorre dentro do nosso *corpus*. Algumas entradas também incluem uma nota sobre, entre outras coisas, o uso idiossincrático do termo em japonês pela montadora. Essa organização das entradas do glossário permitiu fornecer ao usuário informação abrangente para ser usada na tradução sem a necessidade de consulta a outras fontes. Um exemplo de entrada do glossário é mostrado a seguir².

緊急制動 [きんきゅうせいどう] *Grammar noun English* emergency braking *Domain* automobiles *Definition* Using a vehicle's brakes to bring the vehicle to a stop as quickly as possible (typically in order to avoid an accident). *Source* SCDH (July 2017) *Context* 「4輪のABS センサーから緊急制動を検出し、緊急制動信号を発信する。」

2 A entrada do glossário não foi traduzida para a língua portuguesa porque a proposta dos autores é criar um glossário no par linguístico inglês-japonês e isso envolve questões de segmentação de texto e regras gramaticais particulares ao projeto e a esse par linguístico. Ainda assim, a título de exemplo, a mesma entrada no par linguístico japonês-português seria: 緊急制動 [きんきゅうせいどう] *Gramática* substantivo *Português* freio de emergência *Domínio* automóveis *Definição* Usar os freios de um veículo para fazê-lo parar completamente o mais rápido possível (normalmente a fim de evitar uma colisão). *Fonte* SCDH (Julho 2017) *Contexto* 「4輪のABS センサーから緊急制動を検出し、緊急制動信号を発信する。」

3.3 Escolha do software

Nossa experiência profissional sugere que poucos tradutores *freelance* do par linguístico japonês-inglês estão dispostos a pagar por um *software* enquanto uma versão similar estiver disponível gratuitamente. Nossa experiência também sugere que poucos tradutores *freelance* do par linguístico japonês-inglês são versados em linguagens de programação (por exemplo, Python) ou conseguem utilizar interface de linha de comandos, sendo a maioria usuária dos sistemas operacionais Windows ou Macintosh. Além disso, nossa experiência sugere que exigências de confidencialidade impostas pelos clientes de tradução impedem qualquer *upload* de texto-fonte em serviços on-line realizados por terceiros. Portanto, decidimos que qualquer *software* utilizado para a criação do glossário deveria ser livre e compatível com Windows e/ou Macintosh, contando ainda com um processo de instalação simples de dois cliques e uma interface gráfica intuitiva (*user-friendly*).

Um recurso essencial foi o concordanciador. As regras para a utilização de um concordanciador na criação de glossários são, acreditamos, adequadamente explicadas na literatura pertinente como, por exemplo, em Bowker e Pearson (2002). Vários concordanciadores estão disponíveis para os sistemas operacionais mais utilizados. Nós selecionamos o concordanciador livre AntConc 3.2.4 (ANTHONY, 2014). A versão escolhida não é a mais recente, AntConc 3.4.4 (ANTHONY, 2016). Optamos pela versão anterior porque já estávamos familiarizados com ela e satisfeitos com sua funcionalidade para os propósitos do nosso estudo.

Outro recurso essencial foi um segmentador para textos em japonês. Selecionamos o segmentador livre ChaSen 2.1 (MATSUDA, 2000). O tempo decorrido desde o lançamento desse recurso e uma aparente falta de atualizações por parte de seu desenvolvedor inicialmente nos fez repensar nossa escolha. Entretanto, fomos tranquilizados pelas evidências de seu uso continuado em pesquisas linguísticas de língua japonesa como, por exemplo, Breen (2010, pp. 13-22). Somado a isso, o desenvolvedor do AntConc, Laurence Anthony, confirmou, por meio de comunicação pessoal com um dos autores do presente artigo, que ChaSen é um recurso de segmentação comum no Japão. Durante o nosso estudo, descobrimos que Anthony havia lançado um segmentador, SegmentAnt (ANTHONY, 2017), que também pareceu satisfazer os nossos critérios de pesquisa. Pretendemos, dessa forma, utilizar esse último *software* livre em um estudo futuro.

4 Análise dos resultados e discussão

Textos em japonês normalmente não incluem espaços para mostrar onde uma palavra termina e a próxima começa. Essa característica não é exatamente um problema para as linhas de concordância, mas nos obrigou a trabalhar extensivamente o *corpus* antes que pudéssemos utilizar o concordanciador AntConc 3.2.4 (ANTHONY, 2014) para criar listas de frequência e listas em ordem alfabética.

O desafio inicial neste estudo foi fazer o *parsing* (análise sintática) do *corpus*. O AntConc 3.2.4 (Anthony, 2014) não tem capacidade para executar *parsing* em textos. Mesmo com acesso ao ChaSen 2.1 (MATSUDA, 2000), inicialmente experimentamos

fazer uma segmentação manual, ou seja, fazer a etapa de *parsing* manualmente no *corpus* inserindo espaços. Já que tínhamos assumido, desde o início, que boa parte (talvez, a maioria) do conteúdo lexical relevante para a produção do glossário seria composta de substantivos e/ou expressões cujos núcleos são substantivos, nossa segmentação manual envolveu, entre outras coisas, separar substantivos de modificadores que os fizessem funcionar como verbos ou adjetivos. Nosso raciocínio para separar os substantivos dos modificadores era que, pelo menos, poderíamos usar o concordanciador para identificar todos os casos de expressões compostas por substantivos. A segmentação manual do *corpus* foi tediosa e demorada; envolveu pressionar cerca de dez mil vezes as teclas de barra de espaço e setas de direção no teclado do computador, somando mais ou menos dez horas de trabalho. Infelizmente, os resultados se provaram inúteis, pois, apesar de nossas melhores intenções, não tínhamos sido consistentes em nossa tarefa de separar substantivos de modificadores. Nesse ponto, decidimos optar por fazer o *parsing* do nosso *corpus* com ChaSen 2.1 (MATSUDA, 2000).

ChaSen 2.1 (MATSUDA, 2000) não produziu resultados imediatamente utilizáveis pois segmentou muitos termos de vários caracteres incorretamente. (Por exemplo, separou 「フェイスリフト」 [Tradução literal: *facelift*] em seus dois substantivos constituintes e os mostrou como termos separados.) Foi necessário limpar os resultados, entre outras coisas, removendo manualmente centenas de quebras de linha – um processo que levou várias horas. Ainda mais processos manuais se fizeram necessários quando as listas de frequência e em ordem alfabética mostradas pelo AntConc 3.2.4 (ANTHONY, 2014), nessa fase, foram geradas contendo uma grande quantidade de “ruído” (BOWKER e PEARSON, 2002, p. 169) sob a forma de numerais, palavras em inglês e modificadores de substantivos. (Uma amostra da lista de frequência dessa fase é mostrada no Apêndice 2.) Alguns dos modificadores de substantivos estavam escritos em *hiragana*. Nós consideramos mantê-los no *corpus*, usando o concordanciador para criar uma *stop list* para eles, mas percebemos que uma lista dessa natureza não seria viável, já que também retiraria dos nossos resultados bons candidatos a termo escritos em *hiragana*. Remover manualmente o “ruído” (BOWKER e PEARSON, 2002, p. 169) consumiu várias horas. A limpeza manual precisaria de mais trabalho ainda, mas estávamos confiantes de que, pelo menos, os resultados seriam internamente mais consistentes do que os resultados da nossa frustrada tentativa anterior. O conteúdo resultante desse *corpus* foi predominantemente de substantivos. Como tínhamos assumido, desde o início, que muitos ou todos os nossos candidatos a termo seriam substantivos, não estávamos exatamente preocupados com a perda de conteúdo de outra natureza gramatical.

Nesse ponto, foi possível usar o AntConc 3.2.4 (ANTHONY, 2014) para produzir uma lista de frequências utilizável (ver amostra no Apêndice 3) e uma lista alfabética utilizável (ver amostra no Apêndice 4). A lista de frequências foi essencial. No entanto, a lista em ordem alfabética sugeriu que a lista de frequências não fornecia base suficiente para decidir quais termos deveríamos incluir no glossário. Na verdade, a lista em ordem alfabética revelou que certos termos apareceram no *corpus* tanto isolados quanto integrando construções compostas. Enquanto que, por exemplo, a lista de frequências mostrou o termo 「減衰」 [tradução literal: *amortecimento*] em 904º lugar com uma única ocorrência, a lista em ordem alfabética revelou que o termo também apareceu em

compostos como 「減衰力」 [tradução literal: força de amortecimento] e 「振動減衰性」 [Tradução literal: desempenho de amortecimento de vibrações]. Ao utilizar também o AntConc 3.2.4 (ANTHONY, 2014) para produzir linhas de concordância, tanto ordenadas à esquerda quanto à direita, fomos capazes de identificar todos os compostos que continham candidatos a termo. Os modificadores de substantivos encontrados antes e/ou depois dos candidatos a termo pareciam ser “vocabulário subtécnico” (BOWKER e PEARSON, 2002, p. 103). Nós entendemos que a tradução literal de tais modificadores de substantivos estariam corretas desde que os substantivos por eles modificados tivessem sido traduzidos corretamente. Portanto, excluímos esses modificadores de substantivos do glossário.

Considerando nossas suposições sobre usuários potenciais, acreditamos que nosso glossário (ver Apêndice 1) atinge seu objetivo de uso. Com certeza, está livre das principais deficiências dos dicionários (conforme discutido anteriormente neste artigo). Um aprimoramento possível ao nosso glossário diz respeito à sua formatação. Criamos esse glossário como blocos de texto (um bloco por entrada) com o objetivo de termos liberdade máxima para alongar, encurtar e manipular entradas à medida que as refinamos. Se tivéssemos criado o glossário em uma planilha do *Microsoft Excel*, provavelmente o glossário poderia ser convertido de maneira mais fácil em uma base de dados para *softwares* de tradução.

5 Considerações finais

As vantagens de um glossário baseado em *corpus* sobre um dicionário convencional são ressaltadas pela observação de Firth (1957, p. 179, citado por STORJOHANN, 2010, p. 6) de que nós “sabemos o significado de uma palavra pela companhia que ela mantém”. Dito isto, nossa experiência neste estudo, ao tomar uma abordagem baseada em *corpus* para a criação de um glossário de tradução, sugere que esse empreendimento é desafiador quando o *corpus* está em japonês. O principal desafio parece concentrar-se no fato de que a língua japonesa normalmente não usa espaços para marcar limites entre palavras. A necessidade de segmentar o *corpus* usando ChaSen 2.1 (MATSUDA, 2000) e, depois, gastar muitas horas executando a limpeza manual dos resultados antes de poder analisá-los com AntConc 3.2.4 (ANTHONY, 2014) fez com que a produção do glossário fosse extremamente demorada. Isso nos fez suspeitar também que o japonês não seria uma língua adequada para empreendimentos dessa natureza. Nossa suspeita foi reforçada pela existência de um site da Universidade de Tóquio (“Senmon yōgo kīwādo jidō chūshutsu sābisu gensen web”, s.d.) que dá acesso a um sistema de extração automática de termos específicos em domínios de língua especializada a partir de textos em japonês – aparentemente, tornando desnecessário o uso de segmentadores como ChaSen 2.1 (MATSUDA, 2000), sua posterior limpeza manual e análise com um concordanciador.

No entanto, continuamos convencidos do valor fundamental de glossários para a tradução. Não há razão para duvidar que tradutores do par linguístico japonês-inglês (especialmente aqueles que trabalham com textos técnicos ou especializados) podem se beneficiar em longo prazo ao despendendo o tempo necessário para a criação desses glossários. Para um próximo estudo, portanto, planejamos investigar se outras técnicas

e/ou outros *softwares* livres como, por exemplo, SegmentAnt (ANTHONY, 2017), seriam capazes de permitir a criação de glossários de tradução a partir de textos-fonte em língua japonesa de maneira mais rápida e fácil.

Referências

ANTHONY, L. AntConc (Version 3.2.4) [Programa de computador]. Tóquio, Japão: Waseda University, 2014. Disponível em: <<http://www.laurenceanthony.net>>. Acesso em: 13 mar. 2017.

ANTHONY, L. AntConc (Version 3.4.4) [Programa de computador]. Tóquio, Japão: Waseda University, 2016. Disponível em: <<http://www.laurenceanthony.net>>. Acesso em: 13 mar. 2017.

ANTHONY, L. SegmentAnt (Version 1.1.2) [Programa de computador]. Tóquio, Japão: Waseda University, 2017. Disponível em: <<http://www.laurenceanthony.net>>. Acesso em: 13 mar. 2017.

BAKER, M. *In other words*. Abingdon, UK: Routledge. 1992.

BOWKER, L.; PEARSON, J. *Working with specialized language: a practical guide to using corpora*. London, UK: Routledge, 2002.

BREEN, J. Identification of neologisms in Japanese by corpus analysis. In: S. Granger, & M. Paquot (Eds.), *Proceedings of eLex 2009: eLexicography in the 21st century: new challenges, new applications* (p. 13- 22). Louvain, Bélgica: Presses Universitaires de Louvain, 2010.

EUROPEAN GRADUATE PLACEMENT SCHEME. (s.d.). *Occupational standards for European postgraduate translation students on work placement*. Disponível em: <http://www.e-gps.org/wp-content/uploads/2014/05/Occupational_ESF.pdf>. Acesso em: 13 out. 2017.

FAHEY, R. *Japanese text analysis in Python*, 2016. Disponível em: <<http://www.robfahey.co.uk/blog/japanese-text-analysis-in-python>>. Acesso em: 19 ago. 2017.

GOODSELL, D.L. Damping. In: *Dictionary of Automotive Engineering*. (p. 57). Warrendale, PA: Society of Automotive Engineers, 1995.

INTEGRO LANGUAGES. *4 reasons why glossary creation before translation is so important*, 2017. Disponível em: <<http://www.integrolanguages.com/4-reasons-why-glossary-creation-before-translation-is-so-important>>. Acesso em: 13/10/2017.

JAPAN TRANSLATION FEDERATION. 翻訳で失敗しないために翻訳発注の手引き [Tradução literal. *For not getting it wrong with translation: a guide to ordering translation*],

2012. Disponível em: <http://www.jtf.jp/pdf/translation_order.Pdf>. Acesso em: 01 ago. 2017.

KRISHNAMURTHY, R. The corpus revolution in EFL dictionaries. *Kernerman Dictionary News*. Pp. 23-27, Julho, 2002.

LANDAU, S. *Dictionaries: the art and craft of lexicography*. (2a. ed.) Cambridge, UK: Cambridge University Press, 2001.

LIONBRIDGE. *How to create a translation style guide and terminology glossary*. 2016. Disponível em: <<http://content.lionbridge.com/how-to-create-a-translation-style-guide-and-terminology-glossary>>. Acesso em: 13 out. 2017.

MATSUDA, H. ChaSen (2.1) [Programa de computador]. Nara, Japan: Nara Institute of Science and Technology, 2000. Disponível em: <<https://ja.osdn.net/projects/chasen-legacy/releases/27515>>. Acesso em: 13 mar. 2017.

MUEGGE, U. *10 things you should know about automatic terminology extraction*, 2013. Disponível em: <<http://linguagrega.com/blog/2013/09/automatic-terminology-extraction>>. Acesso em: 01 ago. 2017.

SENMON YŌGO KĪWĀDO JIDŌ CHŪSHUTSU SĀBISU GENSEN WEB [Serviço on-line de extração automática de terminologia/palavras-chaves]. (s.d.). Disponível em: <<http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html>>. Acesso em: 20 jul 2017.

STITT, R. *The essentials of consistent terminology in academic and professional translation*, 2016. Disponível em: <<https://www.ulatus.com/translation-blog/the-essentials-of-consistent-terminology-in-academic-and-professional-translation>>. Acesso em: 19 jul 2017.

STORJOHANN, P. Lexico-semantic relations in theory and practice. In: P. Storjohann (Ed.), *Lexical-semantic relations: theoretical and practical perspectives*. (pp. 5-13). Amsterdã, Holanda: John Benjamins Publishing, 2010.

TAKEUCHI, K.; KAGEURA, K.; KOYAMA, T.; DAILLE, B.; ROMARY, L. Pattern based term extraction using ACABIT system. *Language Processing*, 10(4), 2003. Disponível em: <<https://arxiv.org/ftp/arxiv/papers/0907/0907.2452.pdf>>. Acesso em: 18/07/2017.

THE COLLINS CORPUS, 2016. Disponível em: <[https://collins.co.uk/page/The+Collins %20+Corpus?](https://collins.co.uk/page/The+Collins+%20+Corpus?)>. Acesso em: 15 ago. 2017.

WINCHESTER, S. *The meaning of everything: the story of the Oxford English Dictionary*. Oxford, UK: Oxford University Press, 2004.

Recebido em dia 30 de abril de 2018.
Aprovado em dia 30 maio de 2018.

APÊNDICE 1: GLOSSÁRIO³

Observações:

1. Por respeito a questões confidenciais, essa apresentação de nosso glossário mostra o nome da montadora como “ABC”, o nome do modelo do carro como “XYZ” e os nomes das cores com registro de propriedade como “Cor 1” e “Cor 2”.
2. SCDH significa Stephen Crabbe e David Heath.

か)

緊急制動 [きんきゅうせいどう] **Grammar** noun English **emergency braking** Domain automobiles Definition Using a vehicle's brakes to bring the vehicle to a stop as quickly as possible (typically in order to avoid an accident). **Source** SCDH (July 2017) **Context** 「4輪のABSセンサーから緊急制動を検出し、緊急制動信号を発信する。」

減衰 [げんすい] **Grammar** noun English **damping** Domain automobiles Definition Dissipation of energy in a vibrating system, usually by mechanical friction or fluid flow through an orifice. **Source** Dictionary of Automotive Engineering (1995) **Context** 「サスクロスの振動減衰性をアップ。」 **Note** 「減衰」 is used not only in isolation but also in compounds such as 「振動減衰性」 [しんどうげんすいせい] (typically rendered as “vibration-damping performance”) and 「減衰力」 [げんすいりょく] (typically rendered as “damping force”).

さ)

サスクロス **Grammar** noun English **suspension crossmember** Domain automobiles Definition A beam that forms a solid link between suspension components on a left-hand wheel and suspension components on the opposite, right-hand wheel. **Source** SCDH (July 2017) **Context** 「外力が加わった際のサスクロス位置決め剛性を高めるとともに、サスクロスの内力を増加させ、振動減衰性を向上させています。」 **Note** ABC typically writes “crossmember” as one word in product-information publications for distributors. It is possible that the term is written as two words, i.e., “cross member”, in other ABC publications and in publications by other automakers.

しっかり感 [しっかりかん] **Grammar** noun English **stability** Domain automobiles Definition The feeling of steadiness given by a suspension system that adequately isolates the body from external forces. **Source** SCDH (July 2017) **Context** 「操舵時のリアのしっかり感を向上させた。」

集中ディスプレイ [しゅうちゅうでいすぷれい] **Grammar** noun English **centre display** Domain automobiles Definition A display that is positioned approximately in the centre of a vehicle's instrument panel (typically separate from the speedometer and any other meter) and shows various types of information (e.g., the current time, the temperature setting of

3 Como o artigo trata de uma proposta de criação de glossário especializado no par linguístico inglês-japonês, a amostra do glossário fornecida pelos autores no texto-fonte não foi traduzida e encontra-se reproduzida aqui tal como foi apresentada no texto-fonte.

the air conditioner, and the settings of the audio system). **Source** SCDH (July 2017) **Context** 「集中ディスプレイでは、時計の時刻調整を簡単に出来るよう、新たに時計調整スイッチを採用しました。」

浄化性能 [じょうかせいのう] **Grammar noun English emission-reduction performance** Domain automobiles Definition The effectiveness with which a vehicle's exhaust system minimizes emissions of harmful substances. **Source** SCDH (July 2017) **Context** 「冷間始動時の早期活性・浄化性能を向上させ、排出ガスのクリーン化を追求しています。」

ステアリングスイッチ **Grammar noun English steering-wheel switch; switch on the steering wheel** Domain automobiles Definition Any of the switches incorporated into a steering wheel to enable the driver to control vehicle systems (e.g., the audio system) without letting go of the steering wheel. **Source** SCDH (July 2017) **Context** 「ステアリングスイッチに、ABC車として初めてハンズフリーストック操作専用ボタンを採用。」 **Note** Some ABC publications use 「ステアリングスイッチ」 regardless of whether the switch is a rocker switch or a push-button. If the switch is a push-button, “steering-wheel button” or “button on the steering wheel” is a more appropriate rendering.

設定する [せっていする] **Grammar verb English to make available** Domain automobiles Definition To make a vehicle feature, e.g., a colour or technology, available with a particular model. **Source** SCDH (July 2017) **Context** 「エクステリアカラーは全10色を設定しています。」 **Note** Some ABC publications include this usage of 「設定する」 in addition to the more conventional usage, which typically refers to establishing a setting, e.g., setting a temperature with an air conditioner.

操縦安定性 [そうじゅうあんていせい] **Grammar noun English handling stability** Domain automobiles Definition A measure (usually expressed in terms of a cline from worse to better rather than numerically) of how faithfully a vehicle responds to the driver's steering inputs and how stable the vehicle remains when subjected to forces from outside. **Source** SCDH (July 2017) **Context** 「フェイスリフトXYZでも、新しいファミリーフェイスを始めとする新デザインと、操縦安定性に寄与するCD値のさらなる低減を両立させるべく、空力性能の開発を行いました。」 **Note** 「操縦安定性」 is sometimes shortened to 「操安性」 [そうあんせい]. The term “handling stability” is the established rendering for ABC product-information publications aimed at distributors. A rendering that better reflects the etymology of the Japanese term and appears to have greater currency is “handling and stability”. It may be advisable to ask the source-text author whether s/he has a preference.

た)

チューニングする **Grammar verb English to tune** Domain automobiles Definition To adjust the design and/or operating variables of an engine or other vehicle system (e.g., the steering system) to achieve optimal performance. **Source** SCDH (July 2017) **Context** 「パワーステアリングのアシスト特性をチューニングし、制御マップを変更。」 **Note** Where the source text does not explicitly state the purpose of the tuning, “optimize” or “enhance” may be a more suitable rendering.

トップレベル **Grammar** adjective *English among the best; some of the best Domain automobiles* Definition An arguably disingenuous description used by ABC for a vehicle attribute (e.g., fuel economy or engine power) that is better than the corresponding attributes of most competing vehicles but is not the best. **Source** SCDH (July 2017) **Context** 「超小型タービンの採用により、クラストップレベルのレスポンスを実現。」 **Note** ABC uses the term 「トップレベル」 not only by itself but also in compounds such as 「クラストップレベル」 and 「世界トップレベル」.

トレーリングブッシュ **Grammar** noun *English trailing-arm bush Domain automobiles* Definition A bush (a cylindrical sleeve forming a bearing surface for a shaft or pin) in one of the trailing arms of a vehicle's rear suspension. **Source** Dictionary of Automotive Engineering (1995) and SCDH (July 2017) **Context** 「タイヤの動きを後上方に逃がしてショックを緩和するため、リアのトレーリングブッシュのすぐりを前傾。」

は)

ピアノブラック **Grammar** noun *English piano black Domain automobiles* Definition A smooth, glossy, black finish that looks and feels like the finish on the black keys of a piano. **Source** SCDH (July 2017) **Context** 「センターパネル周辺では、現行XYZで採用していたシルバーペイントの加飾を廃止し、新たに艶感や厚み、滑らかさのあるピアノブラックの加飾を採用。」

フェイスリフト **Grammar** noun *English facelift Domain automobiles* Definition A change (or collection of changes) to a vehicle model mid-way through the model's production run. A facelift is less extensive than a full redesign. It typically consists of aesthetic updates but may also include updates to technologies such as the engine. It enables an automaker to freshen an aging model and thereby maintain customer interest in it until the next full redesign. **Source** SCDH (July 2017) **Context** 「今回のフェイスリフトからの新採用色としては、Colour 1、Colour 2 を用意しています。」 **Note** ABC often uses 「フェイスリフト」 as part of a compound noun, e.g., 「フェイスリフト XYZ」. In this case, the established English rendering is the adjective “refined”, e.g., “the refined XYZ”.

踏み換える [ふみかえる] **Grammar** verb *English See Definition. Domain automobiles* Definition To release the brake pedal and press the accelerator pedal or vice versa. **Source** SCDH (July 2017) **Context** 「ドライバーがブレーキを踏み換え、発進に必要なトルクが発生するまでの間、停車状態を維持。」

フラット感 [ふらっとかん] **Grammar** noun *English smoothness Domain automobiles* Definition A feeling of levelness given by a vehicle's suspension system. **Source** SCDH (July 2017) **Context** 「ストラットマウント特性変更：バネ定数ダウンにより、上下方向の振動を適度に逃がしてフラット感を向上。」 **Note** 「フラット感」 tends to be used to describe smoothness in terms of a ride whereby the body does not tip, roll, or bounce to any extent that could be felt by occupants. 「マイルド感」 is also rendered as “smoothness” but tends to be used to describe smoothness in terms of an absence of vibration and harshness in the ride.

ブルブル感 [ぶるぶるかん] **Grammar noun English shake; judder** Domain automobiles
Definition An unpleasant, juddering sensation resulting from failure of a vehicle's suspension system to adequately damp vibration and/or from flexing of an insufficiently stiff body. **Source** SCDH (July 2017) **Context** 「キャビンの小刻みな動きとして感じやすい4Hz~9Hzの上下振動エネルギーと、ブルブル感や減衰の悪さを感じやすい10~14Hzの前後振動エネルギーを低減」 **Note** If the source text explicitly states that the 「ブルブル感」 results from flexing of an insufficiently stiff body when the vehicle goes over bumps, the appropriate rendering is “scuttle shake”.

プレミアム感 [ぶれみあむかん] **Grammar noun English premium identity** Domain automobiles
Definition A sense of superior quality conveyed by a vehicle or by some feature(s) of a vehicle. **Source** SCDH (July 2017) **Context** 「スポーティでありながらアグレッシブになりすぎず、プレミアム感が漂うフロントビューを創り上げました。」 **Note** If 「プレミアム感」 clearly applies to the appearance and/or tactile quality of a physical object, “premium look”, “premium feel”, or “premium look and feel” may be a more appropriate rendering.

ま)

マイルド感 [マイルドかん] **Grammar noun English smoothness** Domain automobiles
Definition An absence of vibration and harshness in the ride given by a vehicle. **Source** SCDH (July 2017) **Context** 「ブッシュのストッパーを廃止して荷重がかかった際のたわみ特性をリニアにし、マイルド感を向上」 **Note** 「フラット感」 is also rendered as “smoothness” but tends to be used to describe smoothness in terms of a ride whereby the body does not tip, roll, or bounce to any extent that could be felt by occupants.

ら)

リアコンビランプ **Grammar noun English rear combination lamp** Domain automobiles
Definition A rear lamp unit containing a number of lamps with separate functions, e.g., making the vehicle visible from behind in darkness, showing when the vehicle is turning (or about to turn) a corner, and showing when the driver is pressing the brake pedal. **Source** SCDH (July 2017) **Context** 「リアビューでは、リアコンビランプのデザインを新しくしています。」

Apêndice 2: Amostra da lista de frequência antes da remoção do ruído

16 59 より

17 58 する

18 53 現行

19 50 まし

20 46 せ

21 45 が
22 44 な
23 41 感
24 40 性
25 40 採用
26 39 向上
27 36 変更
28 33 R
29 33 図
30 32 操作
31 30 から
32 29 grade

Observação: A primeira coluna mostra a posição de cada termo em ordem de frequência de ocorrência no texto-fonte. A segunda coluna mostra o número de ocorrências.

Apêndice 3: Amostra da lista de frequência utilizável

27 10 ステアリングスイッチ
28 10 バネ定数
29 10 従来
30 10 新
31 10 発進
32 10 色
33 9 インチ
34 9 エンジン
35 9 ディーゼルエンジン
36 9 加速度
37 9 走行
38 9 車
39 8 インテリア
40 8 チューニング

Observação: A primeira coluna mostra a posição de cada termo em ordem de frequência

de ocorrência no texto-fonte. A segunda coluna mostra o número de ocorrências.

Apêndice 4: Amostra da lista alfabética utilizável

73 1 アクリルカバー
74 1 アグレッシブ
75 2 アシスト特性
76 2 アスファルト
77 1 アスレティックさ
78 1 アッパーボディ
79 7 アップ
80 7 アルミホイール
81 1 アンダーボディ
82 2 イメージ
83 9 インチ
84 8 インテリア
85 5 インテリアカラー
86 1 インテリアデザイン

Observação: A primeira coluna mostra a posição de cada termo (frequência) no texto-fonte em ordem alfabética. A segunda coluna mostra o número de ocorrências.