

**COLETA, ETIQUETAGEM E ANOTAÇÃO DE INCOMPATIBILIDADES
INTERSEMIÓTICAS GERADAS POR TRADUÇÃO AUTOMÁTICA**
**COLLECTING, TAGGING, AND ANNOTATING INTERSEMIOTIC
MISMATCHES GENERATED BY MACHINE TRANSLATION**

Augusto Velloso dos Santos ESPINDOLA¹

Thiago Blanch PIRES²

Resumo: Os avanços tecnológicos impulsionaram as pesquisas de *corpora* multimodais. O desenvolvimento de ferramentas projetadas para coleta, etiquetagem e anotação, então, apresenta progressivos avanços. Ainda que a profusão destas ferramentas seja promissora, pesquisas voltadas para *corpora* multimodais de documentos texto-imagem ainda são escassas. Tendo em vista que a variedade de ferramentas disponíveis torna dispendiosa a escolha pelos pesquisadores, este estudo propõe uma análise de ferramentas que possam auxiliar os processos de coleta, etiquetagem e anotação de documentos multimodais texto-imagem em uma abordagem multidisciplinar envolvendo multimodalidade, tradução automática e intersemiótica. Para tanto, esta investigação baseia-se em uma adaptação dos achados em Rohlfing et al. (2006) e nos estudos de incompatibilidades intersemióticas geradas por tradução automática de Pires (2017). Os achados mostram que ferramentas não especializadas apresentam maior potencial de agilidade nos processos de anotação e etiquetagem de documentos multimodais estáticos do que as ferramentas especializadas.

Palavras chaves: Tradução automática. Incompatibilidades intersemióticas. Base de dados multimodais. Documentos multimodais estáticos. Etiquetagem multimodal.

Abstract: Technological advances have raised interest in multimodal corpus research. The development of tools designed for data collection, tagging, and annotation shows increasing progression. Although the profusion of such tools is promising, research aimed at text-image multimodal corpora is still incipient. Given that the variety of tools available makes it time-consuming for researchers to choose, this study proposes an analysis of tools that may assist collecting, tagging, and annotating processes of text-image multimodal documents in a multidisciplinary approach involving multimodality, machine translation and intersemiosis. To this end, this investigation is based on an adaptation of the findings in Rohlfing et al. (2006) and in the studies of intersemiotic mismatches generated by machine translation in Pires (2017). The findings point out that non-specialized tools have greater potential to make the processes of static multimodal documents annotation and tagging less time-consuming than the specialized tools.

Keywords: Machine translation. Intersemiotic mismatches. Multimodal database. Static multimodal documents. Multimodal tagging.

1 Introdução

O desenvolvimento de novas tecnologias é responsável pelo aparecimento de novas práticas da linguagem. Conforme os estudos realizados por Vieira e Silvestre (2015, p.

¹ Bacharel em Línguas Estrangeiras Aplicadas pela Universidade de Brasília. E-mail para contato: augustovse@gmail.com

² Doutor em Gestão da Informação pela Universidade de Brasília; Professor do Departamento de Línguas Estrangeiras e Tradução, vinculado ao Instituto de Letras da Universidade de Brasília. E-mail para contato: thiagocomaga@gmail.com

15), a interferência de tal modernização pode ser vista diretamente no processo de reorganização de novos gêneros discursivos, como é o caso do surgimento dos textos multimodais. Tais textos são caracterizados pela “presença de múltiplas semioses em sua composição” (*ibid.*, 2015) e, de acordo com os achados de Bateman (2008), dependem crucialmente da tecnologia.

Os textos multimodais são caracterizados como uma variedade de modos visualmente interpostos com o intuito de formar uma coleção orquestrada de modos comunicativos (BATEMAN, 2008, p. 1, tradução nossa³). Assim, estes textos são organizados de forma a relacionar semanticamente os elementos linguísticos e visuais presentes em um mesmo documento (PIRES, 2017, 2018), resultando na criação de novos significados entre os diversos modos presentes (BATEMAN, 2008), ou, segundo Lemke (1998), uma multiplicação de significados. Esta relação semântica entre elementos linguísticos e visuais está presente, por exemplo, em documentos multimodais formados por texto-imagem.

Por estar diretamente relacionado com o uso da tecnologia, o desenvolvimento de ferramentas voltadas para a gravação, armazenamento e consulta de documentos multimodais contribuiu para um aumento no interesse pelos estudos da linguística de *corpus* multimodal (KNIGHT, 2011, p. 392). Duncan (2013, p. 1017, tradução nossa⁴) aponta que o aumento de tais ferramentas disponíveis é promissor, pois “sinaliza um engajamento generalizado com o desafio de sistematizar o trabalho de observação e análise de dados de comunicação multimodal, bem como facilitar os aspectos morosos desse trabalho”. De acordo com Rohlfsing *et al.* (2006), pesquisas voltadas para o discurso multimodal são, em grande parte, realizadas com o suporte de tecnologias de edição de texto e mídia. Dentre os motivos para a escolha do uso destas ferramentas, estão a facilidade de uso, capacidade de aprendizado, confiabilidade, baixa probabilidade de obsolescência de *software* e capacidade de compartilhamento e de usabilidade a longo prazo dos *corpora* desenvolvidos a partir de tais ferramentas (ROHLFING *et al.*, 2006, p. 102), sendo a obsolescência de um *software* caracterizada por Duncan (2013, p. 1017) pelo curto período em que uma ferramenta é disponibilizada e mantida, por meio de atualizações, pelos desenvolvedores.

Apesar disso, pesquisas voltadas para os *corpora* multimodais, ainda que possuam esquemas diversos para marcação, codificação e anotação de dados (KNIGHT, 2011, p. 395), até o presente momento não possuem um esquema padronizado de anotação, como ferramentas, formatos ou esquemas (ABUCZKI; GHAZALEH, 2013, p. 87). Esse fator torna a natureza do processo de anotação dispendiosa, ademais de implicar tamanhos relativamente pequenos de *corpora* multimodais, sejam eles formados por documentos dinâmicos de áudio-vídeo, sejam eles formados por documentos estáticos texto-imagem, caso comparados aos *corpora* monomodais de texto.

Como abordado pelos achados em Duncan (2013, p. 1016-1017), a relativamente extensa variedade de ferramentas disponíveis, por vezes com funções aparentemente sobrepostas, torna o processo de escolha pelos pesquisadores difícil e dispendioso. Além disso, existe o fato de que muitas ferramentas de anotação multimodais são descontinuadas em um

³ Do original: “An orchestration of diverse presentational modes.” (BATEMAN, 2008, p.1)

⁴ Do original: “The proliferation of such tools is promising in that it signals widespread engagement with the challenge of systematizing the work of observing and analyzing multimodal communication data, as well as with facilitating the time-consuming aspects of this work.” (DUNCAN, 2013, p. 1017)

curto período de tempo, tendo sido desenvolvidas em plataformas que se tornaram obsoletas ou por pesquisadores que se deslocam para outros projetos, deixando-os sem suporte (*ibid.*, 2013). Há, também, a questão de que grande parte das ferramentas para anotação e etiquetagem multimodais está voltada para o tratamento de *corpora* dinâmicos, aqui descritos como aqueles formados por documentos multimodais de áudio e vídeo, sendo poucas as ferramentas voltadas para o tratamento de documentos multimodais estáticos, aqui descritos como aqueles formados por documentos texto-imagem.

Segundo Duncan (2013, p. 1017, tradução nossa⁵), outro obstáculo relacionado às ferramentas é o tempo despendido pelos pesquisadores na procura e análise de programas que possam servir de auxílio às pesquisas de *corpus* multimodal. Tal fato deve-se à questão de que os pesquisadores “podem não ser capazes de dizer, sem investir muito tempo de aprendizado, se pacotes específicos de ferramentas têm os recursos de que eles precisam ou irão combinar com seu estilo de pesquisa”.

Já as pesquisas em *corpora* multimodais voltadas para relação texto-imagem permanecem pouco exploradas (ESPINDOLA, 2019). Assim posto, o presente trabalho tem como objetivo principal analisar ferramentas que possam auxiliar e tornar menos dispendiosos os processos de coleta, anotação e etiquetagem de documentos multimodais de texto-imagem. Tal objetivo justifica-se pela significativa melhoria na eficiência e na confiabilidade de investigações multimodais a partir do uso de ferramentas (CHEN; ADOLPHS, 2017). Esta investigação possui, ainda, o objetivo específico de analisar tais ferramentas a partir da identificação de incompatibilidades intersemióticas, ou seja, desvios léxico semânticos de legendas que descrevem uma imagem ou parte dela, resultando em uma nova configuração semântica (PIRES, 2017), geradas por uso de tradutor automático.

De forma geral, o presente estudo apresenta relevância como estudo de caso para a investigação de métodos que agilizem a etiquetagem semi-automática de relações texto-imagem em grandes quantidades de documentos multimodais utilizando o Google Tradutor (ênfata-se aqui a relevância metodológica e exploratória dentro do âmbito da avaliação manual de tradução automática). Tais incompatibilidades são responsáveis pela alteração ou perda parcial do significado originalmente pensado de um texto multimodal, questões essas causadas pelo processo de tradução automática (PIRES, 2017, 2018).

Para tanto, este estudo utiliza-se de uma perspectiva multidisciplinar que envolve estudos da multimodalidade e da tradução automática, ao propor a análise de ferramentas de usos especializados e não especializados para a coleta, anotação e etiquetagem de documentos multimodais estáticos de natureza texto-imagem. Nesta investigação, as ferramentas de usos especializados são aquelas programadas para servirem de auxílio para os estudos de *corpora*, ou seja, voltadas para o uso do público de linguistas e linguistas computacionais. Já as ferramentas de usos não especializados são aquelas cujas funções foram adaptadas para o uso (mais intuitivo) em *corpus*.

Tal proposta baseia-se no fato de que o uso de tais ferramentas interfere diretamente na “velocidade com que os *corpora* multimodais são desenvolvidos e analisados”

⁵ Do original: “Researchers seeking an annotation tool appropriate for their work may not be able to tell, without investing a lot of learning time, whether particular packages have capabilities they require or will match their research style.” (DUNCAN, 2013, p. 1017)

(KNIGHT, 2011, p. 407, tradução nossa⁶), tornando menos dispendiosos tais processos para a construção de *corpora* de documentos multimodais estáticos (ESPINDOLA, 2019).

2 Incompatibilidades intersemióticas e tradução automática

Uma incompatibilidade intersemiótica gerada por tradução automática é resultado de uma nova configuração semântica entre os modos verbais e visuais (PIRES, 2017), sendo “modo” todo aspecto visual passível de apresentar informação (BATEMAN, 2008, p. 1). Com isso, a relação semântica entre texto e imagem presente no documento original é parcialmente alterada ou perdida no processo da tradução (PIRES, 2017, p. 73-74).

O estudo entre a relação da intersemiótica e a tradução é possível pois “a forma como aspectos visuais e verbais são combinados pode revelar significados em potencial para serem traduzidos” (PIRES, 2017, p. 17). No entanto, a ocorrência de incompatibilidades intersemióticas geradas por uso de tradutores automáticos ocorre devido ao fato de que tais ferramentas não são programadas para reconhecer essa relação semântica entre texto e imagem. Segundo Pires (2017, p. 99), a tradução automática lida somente com determinadas unidades linguísticas e, portanto, “não é desenvolvida para reconhecer padrões de ligações coesivas entre modo textual e visual”.

Com isso, o significado originalmente pensado no documento multimodal submetido à tradução automática é parcialmente alterado ou perdido no processo da tradução (BATEMAN, 2014), resultando em uma incompatibilidade intersemiótica gerada por uso de tradutor automático (PIRES, 2017). Nesse contexto, a incompatibilidade gerada é responsável por criar uma nova configuração intersemiótica, a qual modifica ou até mesmo anula a construção do significado em relação ao texto de partida (PIRES, 2017), conforme o exemplo presente na Figura 1, extraído da base de dados multimodal de Pires (2017). Nele, é possível identificar a ocorrência de uma incompatibilidade intersemiótica gerada por uso de tradutor automático. O documento em questão, originalmente extraído do *site* de notícias britânico *BBC News*, foi traduzido automaticamente pela ferramenta Google Tradutor, a fim de contribuir para os estudos sobre multimodalidade e tradução automática de Pires (2017). A tradução pela ferramenta Google foi primeiramente realizada no ano 2015, nos estudos de Pires (2017), e posteriormente em 2019, para fins do presente estudo. Em ambas ocasiões, o Google Tradutor gerou a mesma tradução.

⁶Do original: “Semi-automated processes of annotating data would also ease the speed at which multimodal corpora are developed and analyzed.” (KNIGHT, 2011, p.407)

Figura 1 – Artigo de notícia “Brexit: David Cameron to quit after UK votes to leave EU”.



Samantha Cameron viu como o PM anunciou que ele estava saindo

Tradução

Fonte: BBC News, com adaptações.

O exemplo de incompatibilidade intersemiótica presente na Figura 1 é criado a partir de uma mudança de significado da sigla “PM”. Na língua inglesa, idioma em que o documento multimodal em questão foi originalmente redigido, essa sigla é utilizada para referir-se a *Prime Minister*, palavra que pode ser livremente traduzida como Primeiro Ministro. A sigla, então, é utilizada para fazer menção à figura em destaque na imagem, o então Primeiro Ministro David Cameron.

Já na tradução automática gerada pela ferramenta Google Tradutor, a sigla foi mantida, causando, então, a incompatibilidade, pois na língua portuguesa, “PM” é comumente utilizado para referir-se à polícia ou policial militar, e não a Primeiro Ministro. Essa incompatibilidade intersemiótica, portanto, torna a imagem de difícil compreensão, já que aquilo que é lido no texto, resultado de um erro de tradução, não é o mesmo visto na imagem, ou seja, o texto gerado a partir da tradução automática é responsável por criar uma nova configuração semântica entre os modos visual e verbal, nomeadamente a relação texto-imagem. Essa nova configuração é responsável pela perda do significado originalmente pensado no documento multimodal, pois o novo modo verbal, resultado da tradução, não é mais compatível com o modo visual (ESPINDOLA, 2019).

3 Coleta, anotação e etiquetagem

Para a construção e análise de um *corpus* multimodal de documentos estáticos texto-imagem, são necessários, dentre outros, os processos de coleta, anotação e etiquetagem, a serem definidos a seguir.

A coleta é composta pela compilação dos documentos que farão parte do *corpus*. A coleta da base de dados multimodais extraída de páginas *on-line* desenvolvida em Pires (2017), por exemplo, utiliza-se do processo semiautomático de captura de telas, por meio de

ferramentas não especializadas, ou seja, programas que não foram desenvolvidos com o intuito de servirem aos estudos de *corpora* (ESPINDOLA, 2019). Os documentos coletados são, então, armazenados para futura anotação e etiquetagem.

A anotação de um *corpus* “compreende a inserção de cabeçalhos informativos nos arquivos” (BERBER SARDINHA, 2004, p. 145), sendo cabeçalhos “trechos demarcados contendo informação não veiculada verbalmente no evento comunicativo” (*ibid.*, p. 145). Esse processo, nos *corpora* de documentos multimodais estáticos de texto-imagem, ainda não conta com um sistema padrão, ou seja, cada pesquisador desenvolve um esquema próprio de anotação de acordo com seu estilo de análise, conforme exemplificado na Figura 2.

Figura 2 – Exemplo de anotação em página da *WikiHow*



Fonte: Pires (2017).

Nela, vê-se um exemplo de anotação em um documento texto-imagem, criado por Pires (2017). Na cor vermelha, o pesquisador ressalta uma anotação de imagem, feita a partir da inserção de formas geométricas, indicando a incompatibilidade intersemiótica gerada por tradução automática presente no documento.

Por fim, a etiquetagem de um *corpus*, segundo Berber Sardinha (2004, p. 150), “consiste na inserção de informações referentes a cada unidade texto”, as quais podem ser quantificadas. Essa etapa tem grandes diferenças em sua execução nos *corpora* monomodais, ou seja, coletâneas de textos que lidam com somente um modo, como é o caso de documentos textuais, e nos *corpora* multimodais, como por exemplo, coletâneas de documentos de texto-imagem. Em *corpora* monomodais de texto, a etiquetagem pode ser realizada a partir do uso de etiquetadores automáticos, o que permite ao pesquisador o tratamento de grandes quantidades de texto de maneira relativamente rápida (BERBER SARDINHA, 2004, p. 150).

Já nos *corpora* multimodais, sejam eles formados por documentos dinâmicos de áudio-vídeo ou documentos estáticos de texto-imagem, o processo de etiquetagem ainda não conta com ferramentas que permitam realizar tal atividade de forma automática, tendo os pesquisadores que realizá-la de forma manual ou semiautomática. Isso se dá pelo fato de que, diferentemente dos *corpora* monomodais, ainda não existe uma convenção que

determine um esquema universal (ABUCZKI; GHAZALEH, 2013, p. 94), ou seja, cada pesquisador cria um sistema de etiquetagem próprio, tendo em vista o tipo de pesquisa e suas características. Além disso, ao passo que documentos textuais são convertidos para os formatos eletrônicos XML ou TXT para que possam ser etiquetados automaticamente, documentos de texto-imagem, em sua maioria, têm a formatação JPEG ou PNG, não sendo possível sua etiquetagem a partir dos mesmos etiquetadores automáticos.

4 Metodologia

O presente trabalho tem caráter qualitativo, ao estudar particularidades de ferramentas escolhidas para análise, e exploratório, tendo-se em vista que a análise de ferramentas para coleta, anotação e etiquetagem de *corpora* de documentos multimodais estáticos texto-imagem conta com escassa pesquisa investigativa (PIRES, 2017). As análises visaram verificar se o uso destas ferramentas pode ou não tornar mais ágil o processo de construção e análise de *corpora* multimodais de documentos texto-imagem.

Para tanto, utilizaram-se como base para análise os documentos presentes no banco de dados⁷ de Pires (2017), formado por 15 pares de documentos multimodais texto-imagem extraídos de páginas *on-line*⁸, dentre elas os *sites* de notícias BBC News⁹ e de artigos procedimentais WikiHow¹⁰. Tais documentos foram traduzidos automaticamente da língua inglesa, idioma em que os documentos foram originalmente publicados, para a língua portuguesa, por meio da ferramenta *on-line* Google Tradutor, no ano de 2015, para estudo das relações texto-imagem nos casos em que tivesse ocorrido erro de tradução automática (PIRES, 2017). Para fins deste estudo, os documentos foram novamente submetidos à tradução automática por meio da ferramenta Google, nos anos 2018 e 2019.

Além disso, a presente investigação toma igualmente como base os procedimentos metodológicos e de análise presentes na pesquisa de Rohlfing *et al.* (2006), com enfoque para as questões pertinentes ao uso de ferramentas para análise multimodal, a seguir (ROHLFING *et al.*, 2006, p. 101, tradução nossa¹¹):

- a) Quais são as dificuldades e os tempos necessários para aprender a usar as ferramentas?
- b) Com que rapidez os dados podem ser anotados? (Velocidade de execução)
- c) O usuário precisa pensar com cuidado e ter muitas informações em mente ao usar a ferramenta? (Carga mental)
- d) Com que frequência ocorrem erros e qual a gravidade deles?

⁷ Base de dados disponível em <https://www.dropbox.com/s/g7ky2traywqxbek/corpus.zip?dl=0> (PIRES, 2017, p. 120). Acesso em: 10 abril. 2018.

⁸ Todos os direitos autorais foram mantidos, conforme explicitado em Pires (2017).

⁹ Disponível em <https://www.bbc.co.uk/news>. Acesso em: 11 abril. 2018.

¹⁰ Disponível em <https://www.wikihow.com/Main-Page>. Acesso em: 11 abril. 2018.

¹¹ Do original: “a) How difficult is it, and how long does it take to learn to use it? b) How quickly can the data be annotated? (speed of execution), c) Mental load: Does a user have to think carefully and have much of information in mind while using this tool? d) How often do errors occur, and how serious are they?” (ROHLFING *et al.*, p. 101, 2006)

As questões formuladas por Rohlffing *et al.* (p. 101, 2006) dizem respeito a funcionalidade, ou seja, o estudo das funções oferecidas pelos programas, e a usabilidade, tais como as funções podem ser utilizadas para auxiliar os usuários em suas pesquisas. Neste caso, os tópicos de Rohlffing *et al.* (2006) são destinados aos estudos das funções e limitações de ferramentas para a análise de documentos multimodais, em especial os processos de anotação e etiquetagem. Ademais, os autores (*ibid.*, 2006) apresentam como produto final de sua pesquisa a tabulação dos dados relacionados a outros quesitos das ferramentas, como informações gerais, modos de controle e informações técnicas.

Na presente investigação, optou-se por uma adaptação das questões de Rohlffing *et al.* (2006) para o contexto de documentos multimodais estáticos texto-imagem, em que se levaram em conta os tópicos supracitados. A análise de ferramentas foi realizada durante o período de um ano, entre 2018 e 2019, pelos próprios investigadores, voltados para os estudos da multimodalidade, tradução automática e incompatibilidades intersemióticas. As investigações foram realizadas no âmbito do grupo Complett¹², na linha intitulada “Tradução, Tecnologia e Multimodalidade”.

A primeira questão, relativa a possíveis dificuldades de uso das ferramentas pelos investigadores, foi avaliada pelo tempo necessário para que o usuário possa se familiarizar com as funções da ferramenta, ou seja, o tempo despendido para que se possa usufruir de todas as opções de usabilidade oferecidas pelo programa. A segunda questão, referente ao tempo necessário para realizar os processos de anotação e etiquetagem, está ligada ao tempo despendido para adicionar, excluir ou alterar tais processos, assim como a natureza deles, como, por exemplo, se são feitos de maneira individual ou se é possível reutilizar o mesmo esquema de etiquetagem em diversos documentos. A terceira questão, relativa a “carga mental”, pode ser definida como o grau de dificuldade enfrentado pelo usuário ao utilizar a ferramenta em análise; remete à complexidade dos comandos de cada ferramenta, como, por exemplo, a necessidade ou não de conhecimentos prévios específicos ou leitura atenta de manuais para que o usuário possa utilizar a ferramenta. A última questão, referente à frequência de erros, leva em consideração problemas técnicos como relatórios de erros, “congelamento” de tela e fechamento involuntário do programa (ESPINDOLA, 2019); sua gravidade, como descrito por Rohlffing *et al.* (2006), é medida com base na perda ou não, total ou parcial, dos dados.

Para a análise o presente estudo selecionou três ferramentas, nomeadamente *Nimbus Capture*, *Evernote* e *UAM ImageTool*. Tais ferramentas foram selecionadas por sua gratuidade e funções voltadas para documentos multimodais de texto-imagem. A análise foi dividida em três etapas. A primeira delas é voltada para a coleta dos documentos, realizada a partir da função de captura de tela. A segunda etapa está voltada para o processo de anotação dos documentos extraídos na etapa anterior, em que se analisa possíveis funções para a adição de texto e marcadores, como desenhos geométricos e caixas de texto. A última etapa analisa a função de etiquetagem, referida nesta investigação como informações referentes as incompatibilidades intersemióticas geradas por tradução automática do documento. A análise foi realizada com o uso de 10 pares

¹² COMPLETT - Corpus Multilíngue para Pesquisas em Línguas Estrangeiras, Tradução e Terminologia é um grupo de pesquisa do Instituto de Letras da Universidade de Brasília (UnB). Certificado pelo CNPQ, ele “tem como objetivo explorar as relações entre tradução, tecnologia e multimodalidade sob as perspectivas de língua e linguagem para desenvolver corpora multimodais que possibilitem analisar as estruturas e construções realizadas por tecnologias de tradução”. Disponível em: <http://dgp.cnpq.br/dgp/espelholinha/4492045805651960582297>. Acesso em 10 jul. 2018.

extraídos da base de dados de Pires (2017), em máquinas Windows 10, processador Intel Core, 4GB RAM, resolução 1366 x 768.

5 Análise e discussão

Nesta seção, discutem-se os achados das ferramentas escolhidas para análise nos processos de coleta, etiquetagem e anotação de incompatibilidades intersemióticas geradas por tradução automática em documentos multimodais estáticos de texto-imagem. A lista de programas estudados conta com as ferramentas de uso não especializado *Nimbus Capture*¹³ e *Evernote*¹⁴ e do programa de uso especializado *UAM ImageTool*¹⁵.

A avaliação de cada ferramenta foi baseada nos estudos de Rohlfig *et al.* (2006), apresentados na seção anterior, em que se levam em conta quatro etapas principais em relação a usabilidade e funcionalidade para análise multimodal a partir do uso de programas. A Tabela 1, a seguir, reúne um panorama de cada uma das ferramentas escolhidas para fins deste estudo, a serem descritas com maiores detalhes nas subseções seguintes.

Tabela 1 – Tabulação dos achados

	Nimbus Capture	Evernote	UAM ImageTool
Página inicial	https://nimbusweb.me	https://evernote.com/	http://www.wagsoft.com/ImageTool/
Versão utilizada	8.8.3	6.17.7.8474	2.1
Natureza	Extensão de navegador	Software de aplicação	Ferramenta de anotação de imagem
Ferramenta desenvolvida para:	Captura de tela e vídeos de páginas web, edição de imagem	Capturar e organizar notas, arquivos, fotos e lembretes. Armazenar e compartilhar informações	Anotação em corpora de imagens
Serventia para construção e análise de corpora de documentos multimodais estáticos	Coleta, anotação e etiquetagem	Coleta, anotação e etiquetagem	Anotação e etiquetagem

Fonte: Espindola (2019).

Tabela 1 – Tabulação dos achados (continuação) Fonte: Espindola (2019).

¹³ Disponível em <https://nimbusweb.me/>. Acesso em: 4 jul. 2018.

¹⁴ Disponível em <https://evernote.com/intl/pt-br/>. Acesso em: 19 abr. 2019.

¹⁵ Disponível em <http://www.wagsoft.com/ImageTool/>. Acesso em: 15 out. 2018.

Tabela 1 – Tabulação dos achados (continuação)

Usabilidade			
Características positivas	Oferece edição de imagem, como adicionar formas e texto. Possui a opção de borrar parte selecionada. Opção de enviar o arquivo gerado para o Google Drive ou Dropbox. Possível compartilhar o trabalho com outros usuários. Oferece cadastro gratuito de usuários em sua plataforma, na qual podem salvar, anotar e etiquetar documentos multimodais	Possibilita a captura de telas de páginas web, assim como sua edição e compartilhamento. Permite ao usuário criar um conta gratuita, com a qual é possível sincronizar os arquivos salvos com outros dispositivos e/ou usuários. Possui um sistema de etiquetagem semi-automática para os arquivos de imagens	Permite a classificação cruzada e a subespecificação de recursos, ou seja, alterações ou exclusões de esquemas são automaticamente atualizadas no demais arquivos. Possível a anotação de múltiplas imagens reutilizando o mesmo esquema de anotação. Suporte para uma série de análises estatísticas, facilitando o encontro de padrões
Limitações	Caso a página que se queira capturar ultrapasse o tamanho de 10Mb, parte do conteúdo não é processado. Não é possível extrair dados estatísticos automáticos do sistema de etiquetagem	O sistema de etiquetas não é separado por pastas/projetos. Não é possível extrair dados estatísticos automáticos do sistema de etiquetagem	Requer aprendizado prévio para usufruir de todas as opções que a ferramenta oferece. Não possui a opção de desfazer última(s) modificações

Fonte: Espindola (2019).

5.1 Ferramentas de uso não especializado

Para fins deste estudo, foram escolhidas duas ferramentas de usos não especializados, ou seja, ferramentas que não foram programadas para fins específicos de investigação em corpora. São elas a extensão de navegador *Nimbus Capture* e o *software* de aplicação *Evernote*. Suas escolhas foram baseadas na gratuidade das ferramentas e na presença de funções que permitem a anotação e a etiquetagem de documentos multimodais compostos por texto-imagem.

Optou-se por inclui-las na análise uma vez que, “entre as ferramentas que suportam a análise de dados do discurso multimodal, os editores de mídia e de texto, embora considerados de baixa tecnologia, ainda podem ter a mais ampla base de usuários” (ROHLFING *et al.*, 2006. p. 101, tradução nossa¹⁶). Além disso, ferramentas de uso especializado voltadas para a análise multimodal de documentos texto-imagem permanecem escassas, sendo a produção de tais voltada, em sua maioria, para análise de documentos formados por áudio-vídeo.

¹⁶ Do original: “Among tools that support analysis of multimodal discourse data, media and text editors, though regarded as ‘low-tech’, may still have the broadest user base.” (ROHLFING *et al.*, 2006, p.101)

5.1.1 *Nimbus Capture*

O programa *Nimbus Capture* é uma extensão de navegador cujo principal objetivo é a captura de telas, mas que se destaca por oferecer funções que permitem a anotação e etiquetagem de documentos multimodais estáticos texto-imagem (ESPINDOLA, 2019).

Por ter interfaces simples e intuitivas, ou seja, que não necessitam de um estudo prévio pelo investigador, a extensão é de uso fácil mesmo para aqueles ainda não familiarizados com a ferramenta. Com base nos estudos realizados pelos próprios investigadores, isso permite que o usuário possa se familiarizar com os comandos e funções dessa ferramenta à medida que insere, analisa e etiqueta os documentos multimodais de texto-imagem. Com isso, a “carga mental”, definida por Rohlfing *et al.* (p. 101, 2006) como a dificuldade e a quantidade de informações necessárias pelo usuário para que este possa utilizar uma ferramenta, é baixa, ou seja, o usuário não necessita memorizar comandos ao utilizar o programa. Além disso, com base na adaptação dos estudos de Rohlfing *et al.* (p. 101, 2006) sobre frequência de erros, o *Nimbus Capture* pode ser considerado uma ferramenta com potencial considerável de confiabilidade, pois são raros os erros de mal funcionamento como “congelamento” de tela ou fechamento involuntário do programa.

Ademais do sistema de captura de telas, que pode servir de auxílio para o processo de coleta de dados, o *Nimbus Capture* possui opções que permitem a edição de imagem. Essa função foi utilizada por Pires (2017) com o auxílio da ferramenta *Sketch*¹⁷, de uso não especializado, para criar o sistema de anotações nos documentos multimodais estáticos de texto-imagem presentes em sua base de dados. As anotações em questão são feitas a partir da inserção de caixas de texto e figuras geométricas para ressaltar e classificar a incompatibilidade intersemiótica presente no documento multimodal.

Além das funções que podem auxiliar os processos de coleta e anotação em documentos multimodais texto-imagem retirados de páginas *on-line*, a ferramenta *Nimbus Capture* oferece um sistema que permite a etiquetagem dos documentos. As etiquetas, referidas como *tags*, podem ser adicionadas, excluídas ou modificadas a qualquer momento. Além disso, esse processo é otimizado pela interface, que grava as etiquetas já cadastradas, tornando a etiquetagem dos documentos uma atividade semiautomática. As *tags* também facilitam a busca e agrupamento de documentos, uma vez que o sistema permite reagrupar dados, conforme etiquetagem, em diversas categorias (ESPINDOLA, 2019).

A ferramenta *Nimbus Capture* tem a vantagem de oferecer funções que permitem a coleta, anotação e etiquetagem de documentos multimodais estáticos extraídos de páginas *on-line* com o uso de um único programa. No entanto, por não ser uma ferramenta especializada, ou seja, que não foi primordialmente programada para ser utilizada em estudos de *corpora*, não possui certas funções de análises comuns a elas. Esse é o caso, por exemplo, das funções estatísticas, ou seja, a extração automática de padrões (ESPINDOLA, 2019).

5.1.2 *Evernote*

A ferramenta *Evernote* é um *software* de aplicação cujas principais funções são capturar, salvar, modificar e compartilhar arquivos *on-line*. De maneira similar ao *Nimbus Capture*, o *Evernote* possui uma interface simples e intuitiva, ou seja, que não requer

¹⁷ Disponível em <https://evernote.com/products/skitch>. Acesso em: 12 jul. 2018.

conhecimentos específicos de seus usuários, o que, segundo os estudos de Rohlffing *et al.* (p. 101, 2006), implica em uma baixa “carga mental” (vide seção anterior).

Segundo informações encontradas no *website*¹⁸ oficial da ferramenta *Evernote*, sua proposta é permitir que seus usuários salvem, editem e compartilhem arquivos de maneira rápida e simples, o que o torna igualmente eficaz para as etapas de coleta, anotação e etiquetagem em documentos multimodais estáticos extraídos de páginas *web* (ESPINDOLA, 2019). Ademais, com base nas análises realizadas no período de 2018 e 2019 pelos próprios investigadores, relatórios de erros técnicos como fechamento involuntário ou “congelamento” de tela são escassos, o que confere a ferramenta um potencial considerável de confiabilidade, sendo confiabilidade aqui descrita como a ausência de erros visíveis como os anteriormente mencionados.

Similarmente à outra ferramenta não especializada em estudo, a função de captura de tela do *Evernote* pode auxiliar na coleta de documentos multimodais estáticos texto-imagem, como proposto em Pires (2017) para o estudo de incompatibilidades intersemióticas geradas por uso de tradutor automático. Já as funções de edição de imagem permitem a anotação dos dados, também de forma similar à proposta por Pires (2017). O processo de etiquetagem de documentos também funciona de maneira similar ao *Nimbus Capture* e auxilia o usuário a localizar, agrupar ou até mesmo quantificar os arquivos, ainda que de forma semiautomática (ESPINDOLA, 2019).

A ferramenta *Evernote*, apesar de fornecer funções que permitam os processos de coleta, anotação e etiquetagem de documentos multimodais estáticos com o uso de uma única ferramenta, não conta com um sistema que permita a extração de dados estatísticos de maneira automática. Ademais, ainda que permita renomear e adicionar etiquetas a qualquer momento da pesquisa, não há uma divisão de projetos, ou seja, o sistema agrupa múltiplos projetos em uma mesma lista de etiquetas. Por exemplo, independentemente de quantas pastas, ou projetos, o usuário criar, as etiquetas são organizadas em ordem alfabética, em um único atalho. Tal limitação pode ser um aspecto negativo caso o investigador utilize a ferramenta para realizar estudos distintos, pois as informações seriam armazenadas pelo *Evernote* em uma mesma pasta, mesclando os conteúdos (ESPINDOLA, 2019).

5.2 Ferramenta de uso especializado

Para fins deste estudo, escolheu-se a ferramenta *UAM ImageTool* como exemplo de programa de uso especializado, ou seja, ferramenta primordialmente programada para servir de base e auxílio aos estudos de *corpora* multimodal. Sua escolha foi baseada primeiramente por sua gratuidade e, em segundo lugar, por ser uma ferramenta voltada para o tratamento de documentos texto-imagem.

5.2.1 c. *UAM ImageTool*

Trata-se de uma ferramenta especializada em anotação e etiquetagem de imagens, voltada para o desenvolvimento de *corpora*. Permite a etiquetagem e anotação de partes ou da totalidade de uma imagem, ademais de oferecer a possibilidade de “quantificar os tipos de categorias analisadas” (PIRES, 2017, p. 106), ou seja, de extrair dados estatísticos de forma automática.

¹⁸ Disponível em <https://evernote.com/intl/pt-br/>. Acesso em: 19 abr. 2019.

Apesar de não oferecer suporte para a coleta de documentos, o *UAM ImageTool*, diferentemente das ferramentas não especializadas anteriormente analisadas, dispõe de um sistema de anotação e etiquetagem voltado para o uso em *corpora*. Por exemplo, os esquemas realizados por meio desse programa são armazenados em formato XML, o que permite a exportação dos dados para outros programas de uso similar.

Além disso, o *UAM ImageTool* permite a “classificação cruzada e a subespecificação de recursos” (O’DONNELL, 2008, p. 15, tradução nossa¹⁹), função ausente nas demais ferramentas analisadas no presente estudo. Essa função remete à capacidade do programa em permitir a correlação entre os documentos do *corpus* em uso. Com isso, torna-se possível a ligação de uma série de dados que compartilharão os mesmos comandos. Por exemplo, alterações, exclusões ou adições no esquema de etiquetagem serão automaticamente repassadas para todos os documentos, não sendo necessária sua alteração manual e individual pelo pesquisador.

Com isso, faz-se possível a etiquetagem de múltiplos documentos multimodais texto-imagem, reutilizando-se o mesmo esquema de etiquetas. Isso torna o tempo despendido para realizar tal tarefa menor se comparado ao sistema individual das demais ferramentas analisadas neste estudo, como o *Nimbus Capture* e *Evernote* (ESPINDOLA, 2019).

Por fim, um dos maiores diferenciais do programa *UAM ImageTool* em comparação as demais ferramentas apresentadas é que ele oferece recursos para estudos estatísticos. Essa função é realizada de maneira automática: ainda que o investigador etiquete os dados, o programa agrupa dados similares em subconjuntos, o que permite a descoberta de padrões de maneira ágil, principalmente se comparada ao trabalho manual ou semiautomático exigido pelas ferramentas não especializadas. Os dados estatísticos automaticamente gerados são disponibilizados pela ferramenta em formato de tabelas e gráficos, podendo ser exportados para o computador do usuário.

Apesar disso, se comparado aos programas de uso não especializado anteriormente descritos, a ferramenta *UAM ImageTool* exige de seus usuários uma maior “carga mental”. Além de ser um programa voltado para linguistas, a gama de opções de análise oferecidas exige de seus usuários certo conhecimento prévio, devido à complexidade maior de comandos, ademais de uma atenta leitura de manuais para que se possa usufruir de todas as funções presentes na ferramenta.

Em certos casos, apesar de isolados, o programa pode apresentar relatórios de erro caso algum dos dados inseridos no *corpus* não tenha formato ou tamanho suportados pela ferramenta, que lida com arquivos em formato de imagem JPEG e PNG. Tal relatório é gerado somente ao acessar os dados estatísticos.

6 Considerações finais

A presente investigação teve como objetivo a realização de uma análise de ferramentas que possam auxiliar e tornar menos dispendiosos os processos de coleta, anotação e etiquetagem de documentos multimodais estáticos para o estudo de incompatibilidades intersemióticas geradas por tradução automática. Para tanto, tomaram-se os achados da base de dados de Pires (2017) como documentos utilizados para a realização de análise, guiada por uma adaptação dos fundamentos de Rohlifing et al (2006), a partir de questões

¹⁹ Do original: “[UAM ImageTool] allows cross-classification as well as sub-specification of features.” (O’DONNELL, 2008, p.15)

de estudos de usabilidade e funcionalidade de ferramentas para estudos e análise multimodal.

A fim de realizar a análise mencionada, escolheram-se três ferramentas, sendo duas delas de uso não especializado, ou seja, recursos que não foram programados com o fim de servirem como ferramentas de auxílio para estudos de *corpora*, nomeadamente as ferramentas *Nimbus Capture* e *Evernote*, e uma ferramenta de uso especializado, ou seja, que foi primordialmente criada para auxiliar em análises de *corpora* multimodais, nomeadamente o programa *UAM ImageTool*.

A partir da análise adaptada a partir dos estudos de Rohlfing *et al.* (2006), foi possível constatar que as ferramentas *Nimbus Capture* e *Evernote*, ambas de usos não especializados, apresentaram um maior potencial para tornar os processos de coleta, anotação e etiquetagem de incompatibilidades intersemióticas geradas por uso de tradutor automático menos morosos que aqueles do programa *UAM ImageTool*, de uso especializado. Tal afirmação é baseada principalmente na morosidade do processo para que o investigador possa aprender a utilizar todas as funções presentes nas ferramentas, definido por Rohlfing *et al.* (2006) como “carga mental”.

Os programas não especializados, de maneira geral, são mais simples e intuitivos que aqueles voltados para o público de linguistas ou linguistas computacionais, como é o caso do *UAM ImageTool*. Com isso, ferramentas como extensões de navegador e programas de compartilhamento e edição de imagens, como *Nimbus Capture* e *Evernote*, não requerem conhecimentos específicos para que o investigador possa começar a utilizá-los, fator que influencia diretamente na rapidez com que os processos são realizados (ESPINDOLA, 2019). Além disso, os programas de usos não especializados escolhidos permitem os três processos analisados, nomeadamente a coleta, anotação e etiquetagem, com o uso de uma única ferramenta. Com isso, é possível concentrar a pesquisa em somente um programa, fator que contribui para a rapidez com que os processos podem ser desenvolvidos pelo pesquisador.

Ainda que as ferramentas não especializadas apresentem funções que auxiliem na anotação e etiquetagem dos documentos texto-imagem, nota-se a ausência de funções importantes para as pesquisas em *corpora* multimodais, como é o caso das análises estatísticas automáticas. Sem essa função, investigadores devem, necessariamente, realizá-la de forma semiautomática, por meio das etiquetas, ou manual, o que pode tornar o encontro de padrões um processo moroso. Tendo-se em vista que grande parte das pesquisas em *corpora* tem caráter quantitativo, esse recurso mostra-se imprescindível. No entanto, vale ressaltar que a ferramenta de uso especializado apresenta, também, certas limitações de análise. Por exemplo, o programa *UAM ImageTool* não tem suporte que permita a coleta de dados, apenas a análise deles.

Conforme os estudos de Duncan (2013, p. 1020), nenhuma ferramenta é isenta de desvantagens em relação a certos objetivos analíticos. Tendo-se em vista tais limitações, contou-se, também, que a combinação de ferramentas não especializadas com o uso de ferramentas especializadas apresenta um potencial ainda maior de tornar as etapas de coleta, etiquetagem e anotação menos dispendiosas (ESPINDOLA, 2019). Tal afirmação leva em consideração a facilidade de anotação e coleta oferecidas pelas ferramentas *Nimbus Capture* e *Evernote*, de usos não especializados, e as funções semiautomática de etiquetagem e extração automática de dados estatísticos presentes no *UAM ImageTool*, de uso especializado.

Por fim, faz-se importante salientar que assim como achados de Rohlfing *et al.* (2006, p. 122), o objetivo desta investigação não foi apontar uma única ferramenta como uma

melhor opção, e sim analisar possíveis ferramentas que possam servir de auxílio para pesquisas voltadas para documentos multimodais estáticos texto-imagem, no contexto de incompatibilidades intersemióticas geradas por uso de tradução automática. Ademais, espera-se que os achados deste trabalho possam contribuir para futuras investigações sobre a multimodalidade, a tradução automática e para os estudos sobre incompatibilidades intersemióticas.

7 Referências

- ABUCZKI, A.; GHAZALEH, E. An overview of multimodal corpora, annotation tools and schemes. **Argumentum**, Debrecen, v. 9, n. 9, p. 86-98, 2013.
- BATEMAN, J. A. **Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents**. Nova York: Palgrave MacMillan, 2008.
- BATEMAN, J. A. **Text and Image: A Critical Introduction to the Visual/Verbal Divide**. Nova York: Routledge, 2014.
- BERBER SARDINHA, T. **Linguística de Corpus**. São Paulo: Manole, 2004.
- CHEN, Y.; ADOLPHS, S. Multimodal Corpus-based Investigations of the Gesture-Speech Relationship. In: International Corpus Linguistics Conference, 9., 2017, Birmingham. Disponível em: <<https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper281.pdf>>. Acesso em: 21 jan. 2019.
- DUNCAN, S. Multimodal annotation tools. In: **Body – Language – Communication: an international handbook on multimodality in human interaction**. Berlim: De Gruyter Mouton, 2013, cap. 66, p. 1015-1022.
- ESPINDOLA, A. **Incompatibilidades intersemióticas geradas por tradução automática: um estudo exploratório sobre coleta, etiquetagem e anotação em corpus de documentos multimodais estáticos**. 2019. 52 p. Programa de Graduação em Línguas Estrangeiras Aplicadas. Monografia – Universidade de Brasília, Distrito Federal, 2019.
- KNIGHT, D. The future of multimodal corpora. **Revista Brasileira de Linguística Aplicada**, v. 11, n. 2, p. 391-415, 2011.
- LEMKE, J. Multiplying Meaning: visual and verbal semiotics in scientific text. In: **Reading Science**. Londres: Routledge, 1998, p. 87-113.
- O'DONNELL, M. Demonstration of the UAM CorpusTool for text and image annotation. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, 2008, Columbus. Disponível em: <<https://www.aclweb.org/anthology/P08-4004/>>. Acesso em: 12 fev. 2019.
- PIRES, T. B. **Ampliando olhares sobre a tradução automática online: um estudo exploratório de categorias de erros de máquina de tradução gerados em documentos multimodais**. 2017. 169 p. Programa de Pós-Graduação em Ciência da Informação. Tese de Doutorado – Universidade de Brasília, Distrito Federal, 2017.

- PIRES, T. B. Multimodality and evaluation of machine translation: a proposal for investigating intersemiotic mismatches generated by machine translation in multimodal documents. **Texto livre: Linguagem e Tecnologia**, Belo Horizonte, v. 11, n. 1, p. 82-102, jan.-abr. 2018.
- ROHLFING, K.; LOEHR, D.; DUNCAN, S., et al. Comparison of multimodal annotation tools: Workshop report. In: *Gesprächsforschung*, 7., 2006, Bielefeld. Disponível em: < <http://www.gespraechsforschung-ozs.de/heft2006/tb-rohlfing.pdf>>. Acesso em: 10 fev. 2019.
- VIEIRA, J.; SILVESTRE, C. **Introdução à multimodalidade: contribuições da gramática sistêmico-funcional, análise de discurso crítica, semiótica social**. Brasília: J. Antunes Vieira, 2015.

