

UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE LÍNGUAS ESTRANGEIRAS E TRADUÇÃO

AUGUSTO VELLOSO DOS SANTOS ESPINDOLA

**INCOMPATIBILIDADES INTERSEMIÓTICAS GERADAS POR TRADUÇÃO
AUTOMÁTICA: UM ESTUDO EXPLORATÓRIO SOBRE COLETA,
ETIQUETAGEM E ANOTAÇÃO EM CORPUS DE DOCUMENTOS
MULTIMODAIS ESTÁTICOS**

BRASÍLIA/DF

2019

AUGUSTO VELLOSO DOS SANTOS ESPINDOLA

**INCOMPATIBILIDADES INTERSEMIÓTICAS GERADAS POR TRADUÇÃO
AUTOMÁTICA: UM ESTUDO EXPLORATÓRIO SOBRE COLETA,
ETIQUETAGEM E ANOTAÇÃO EM CORPUS DE DOCUMENTOS
MULTIMODAIS ESTÁTICOS**

Trabalho de conclusão de curso apresentado junto ao curso de Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação da Universidade de Brasília como requisito parcial à obtenção do título de Bacharel.

Orientador: Prof. Dr. Thiago Blanch Pires

BRASÍLIA/DF

2019

AUGUSTO VELLOSO DOS SANTOS ESPINDOLA

**INCOMPATIBILIDADES INTERSEMIÓTICAS GERADAS POR TRADUÇÃO
AUTOMÁTICA: UM ESTUDO EXPLORATÓRIO SOBRE COLETA,
ETIQUETAGEM E ANOTAÇÃO EM CORPUS DE DOCUMENTOS
MULTIMODAIS ESTÁTICOS**

Trabalho de conclusão de curso apresentado junto ao curso de Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação da Universidade de Brasília como requisito parcial à obtenção do título de Bacharel.

Orientador: Prof. Dr. Thiago Blanch Pires

COMISSÃO EXAMINADORA

Prof.^a. Dra. Carolina Pereira Barcellos (LET/UnB)

Prof.^a. Dra. Elisa Duarte Teixeira (LET/UnB)

Prof. Dr. Thiago Blanch Pires (LET/UnB)

Brasília, 04 de julho de 2019

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador, o professor Dr. Thiago Blanch Pires, sem o qual este trabalho não seria possível, pela disponibilidade, atenção e inúmeras oportunidades.

Às professoras Dras. Carolina Pereira Barcellos e Elisa Duarte Teixeira, por aceitarem fazer parte da banca examinadora.

À minha mãe, Tatiana, que sempre me incentiva em meus estudos, e aos meus amigos Jeferson, Anahy, Renan, Anna, Cris e Sofia, que me acompanharam nesta jornada.

**INCOMPATIBILIDADES INTERSEMIÓTICAS GERADAS POR TRADUÇÃO
AUTOMÁTICA: UM ESTUDO EXPLORATÓRIO SOBRE COLETA,
ETIQUETAGEM E ANOTAÇÃO EM CORPUS DE DOCUMENTOS
MULTIMODAIS ESTÁTICOS**

RESUMO

Os avanços tecnológicos impulsionaram o interesse em corpora multimodais, estimulando o desenvolvimento de ferramentas projetadas para coleta, etiquetagem e anotação. Ainda que a profusão destas ferramentas seja promissora, softwares voltados para os corpora de documentos multimodais estáticos permanecem pouco explorados. Uma vez que a extensa variedade de ferramentas torna dispendiosa a escolha por parte dos pesquisadores, este trabalho propõe a testagem exploratória de ferramentas que possam agilizar os processos de coleta, etiquetagem e anotação de corpora de documentos multimodais estáticos, a partir da adaptação dos achados em Rohlfing et al. (2006). O estudo utiliza-se de uma abordagem multidisciplinar envolvendo áreas da multimodalidade e tradução automática e toma como base para testagem o banco de dados desenvolvido em Pires (2017). Os achados mostram que as ferramentas cuja manipulação não requer conhecimentos prévios de programação apresentam maior potencial de agilidade no processo de anotação de documentos multimodais estáticos que as ferramentas especializadas.

PALAVRAS-CHAVE: Corpus de documentos multimodais estáticos; relação texto-imagem; incompatibilidades intersemióticas; tradução automática; ferramentas de anotação e etiquetagem.

**INTERSEMIOTIC MISMATCHES GENERATED BY AUTOMATIC
TRANSLATION: AN EXPLORATORY STUDY ON COLLECTION, TAGGING AND
ANNOTATION IN STATIC MULTIMODAL DOCUMENTS CORPUS**

ABSTRACT

Technological advances boosted interest in multimodal corpora, stimulating the development of tools designed for their collection, tagging, and annotation. Although the profusion of these tools is promising, softwares of static corpora-oriented multimodal documents is still little explored. Considering that the wide variety of tools makes the choice of researchers time-consuming, this study proposes an exploratory testing of software that may speed up the collection, tagging and annotation process in corpora of static multimodal documents, from an adaptation of the findings in Rohlfing et al. (2006). The study uses a multidisciplinary approach involving areas of multimodality and automatic translation, as it uses as source the database developed in Pires (2017). The findings show that the tools whose manipulation does not require previous programming knowledge present greater potential for agility in the process of annotating static multimodal documents than specialized tools.

KEYWORDS: Corpus of static multimodal documents; text-image relationship; intersemiotic mismatches; automatic translation; annotation and tagging tools.

LISTA DE FIGURAS

Figuras 1 e 2 – Brexit: David Cameron to quit after UK votes to leave EU	15
Figura 3 – An index of multimodal corpora (Knight, 2011)	18
Figura 4 – Exemplo de anotação e etiquetagem (PIRES, 2017)	24
Figura 5 – Edição de imagem Nimbus Capture	30
Figura 6 – Sistema de etiquetagem e anotação Nimbus Capture	31
Figura 7 – Parte da interface de uma conta no Evernote	33
Figura 8 – Exemplo de anotação e etiquetagem no programa UAM ImageTool	35
Figura 9 – O olhar feminino em propagandas de cosméticos	36
Figura 10 – Tentativa de réplica da classificação de erros de tradução	37

SUMÁRIO

1. INTRODUÇÃO	9
2. INCOMPATIBILIDADES INTERSEMIÓTICAS GERADAS POR TRADUÇÃO AUTOMÁTICA	12
3. CORPORA MULTIMODAIS	16
3.1 Coleta, etiquetagem e anotação	20
4. METODOLOGIA	25
4.1 Descrição da fonte de dados	25
4.2 Procedimentos metodológicos e de análise	26
5. ANÁLISE DOS RESULTADOS E DISCUSSÃO	28
5.1 Nimbus Capture	29
5.2 Evernote	32
5.3 UAM ImageTool.....	34
6. TABULAÇÃO DOS ACHADOS	38
7. CONSIDERAÇÕES FINAIS	40
8. REFERÊNCIAS BIBLIOGRÁFICAS	42
Anexo A	44

1. INTRODUÇÃO

A linguagem, no contexto da globalização, encontra-se sob a influência de novas tecnologias, “cuja interferência é diretamente visível na reorganização das práticas sociais e dos gêneros discursivos” (VIEIRA; SILVESTRE, 2015, p. 15). Ainda segundo as autoras (ibid.), “tal fato enseja relevantes mudanças, assinaladas pelo surgimento de textos multimodais, marcados pela presença de múltiplas semioses em sua composição”. Assim, a informação possui um papel chave, através da internet (*web-based*), na expansão de documentos multimodais, os quais dependem crucialmente da tecnologia (BATEMAN, 2008). Tais documentos, originalmente criados de forma a relacionar semanticamente elementos linguísticos e visuais (PIRES, 2017), como por exemplo, texto e imagem, criam “novos significados” (BATEMAN, 2008) ou uma “multiplicação de significados” (LEMKE, 1998) entre os diversos modos presentes.

Os avanços tecnológicos testemunhados nas últimas décadas forneceram ao linguista melhores ferramentas para gravar, armazenar e consultar múltiplas formas de registros digitais, o que fomentou a base para o recente aumento no interesse pela linguística de corpus multimodal (KNIGHT, 2011, p. 392). Este é definido por Foster e Oberlander como “uma coleção anotada de conteúdo coordenado em canais de comunicação” (2007, p. 307-308). Apesar disso, contrariamente aos corpora monomodais, “que possuem uma longa história de uso em linguística, a construção e o uso de corpora multimodais ainda estão em sua infância relativa, com a maioria das pesquisas associadas a este campo datadas em apenas uma década” (KNIGHT, 2011, pág. 392, tradução nossa¹).

Neste contexto, pesquisas e comparações sistemáticas de várias ferramentas de anotação multimodais disponíveis em épocas específicas nos últimos quinze anos surgiram em formato impresso e em vários seminários de conferência (DUNCAN, 2013, p. 1017), sendo anotação assim entendida como a “inserção de cabeçalhos informativos (BERBER SARDINHA, 2004, p. 145) (vide seção 3.1). Algumas das pesquisas mais citadas sobre o discurso multimodal continuam a ser realizadas em

¹ Do original: “Unlike monomodal corpora, which have a long history of use in linguistics, the construction and use of multimodal corpora is still in its relative infancy, with the majority of research associated with this field spanning back only a decade.” (KNIGHT, 2011, pág. 392)

grande parte com o uso de tecnologias de edição de texto, mídia e gráficos de múltiplos propósitos, devido à facilidade de uso, capacidade de aprendizado, confiabilidade, baixa probabilidade de obsolescência de software e capacidade de compartilhamento e usabilidade a longo prazo dos corpora desenvolvidos a partir de tais ferramentas (ROHLFING et al., 2006, p. 102).

Além disso, o desenvolvimento de ferramentas e softwares projetados especificamente para anotação de corpora multimodais envolvendo áudio, vídeo e outros tipos de dados aumentou na última década (DUNCAN, 2013, p. 1016-1017). Duncan (2013, p. 1017, tradução nossa²) salienta ainda que a profusão de tais ferramentas é promissora na medida em que “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 demorados deste trabalho”.

No entanto, a pesquisa de corpus multimodal (CM) enfrenta grandes problemas, como a natureza morosa do processo de anotação, o que implica em tamanhos relativamente pequenos de CM, e a falta de padrões de anotação existentes, tais como ferramentas, formatos e esquemas (ABUCZKI; GHAZALEH, 2013, p. 87). Como também discutido em Knight (2011, p. 395, tradução nossa³), apesar da existência de “vários esquemas distintos para marcação, codificação e a anotação de dados multimodais, até o presente estudo nenhuma abordagem padrão é utilizada em todos os corpora multimodais”.

Ademais, como salientado pelos achados em Duncan (2013, p. 1016-1017), a relativamente extensa variedade de ferramentas disponíveis, com capacidades aparentemente sobrepostas, torna o processo de escolha por parte de pesquisadores difícil e demorada, somado ao fato de que muitas ferramentas de anotação multimodais têm vidas curtas, tendo sido desenvolvidas em plataformas que se tornaram obsoletas ou por pesquisadores que se deslocam para outros projetos, deixando-os sem suporte.

² 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)

³ Do original: “various different schemes exist to mark up, code and annotate multimodal data, and as yet no standard approach is used across all multimodal corpora.” (KNIGHT, 2011, p. 395)

Há, ainda, a questão de que esta gama de ferramentas para tratamento de corpora multimodais está voltada, em sua maioria, para investigações do modo de fala, geralmente baseadas em transcrições textuais, ou seja, pesquisas baseadas em corpus e compreensão de outros modos como características prosódicas e movimentos corporais como posturas, expressões faciais e movimentos de cabeça, mão e braço (CHEN; ADOLPHS, 2017). Tais investigações, cujo objeto de estudo é composto por documentos de áudio-vídeo, serão nesta investigação referidos como documentos multimodais dinâmicos.

Já as pesquisas em corpora multimodais que exploram a relação texto-imagem, ou seja, documentos multimodais estáticos, permanecem pouco exploradas. Isso é salientado pelo fato de que, apesar dos recursos fornecidos por ferramentas e softwares para anotação textual, exploração e descrição, esses métodos são inadequados para a detecção e descrição de comportamentos não-verbais, como descrito em Chen e Adolphs (2017). Para fins desta investigação, entende-se por estáticos os documentos formados por texto-imagem de natureza 2D e sem animações (movimento).

Segundo Knight (2011, p. 392), a utilidade da pesquisa e dos métodos baseados em corpus está se tornando popular em uma série de diferentes disciplinas acadêmicas e campos de pesquisa, muito além da linguística. Apesar disso, conforme os estudos de Duncan (2013, p. 1017, tradução nossa⁴), um obstáculo para os pesquisadores que buscam por ferramentas de anotação apropriadas para seu tipo de investigação é o tempo despendido na procura e testagem de tais programas, pois “podem não ser capazes de dizer, sem investir muito tempo de aprendizado, se pacotes específicos de ferramentas possuem os recursos que eles precisam ou irão combinar com seu estilo de pesquisa”.

Com isso, este trabalho tem como objetivo geral testar e combinar ferramentas e softwares de anotação multimodais, dado que o uso de tais programas pode “melhorar significativamente a eficiência e a confiabilidade de investigações multimodais” (CHEN; ADOLPHS, 2017, tradução nossa), no contexto de corpora de documentos multimodais estáticos. O projeto possui, ainda, o objetivo específico de

⁴ 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)

testar tais ferramentas a partir da identificação de incompatibilidades intersemióticas geradas por tradução automática, ou seja, alteração ou perda parcial do significado originalmente pensado de um texto multimodal causado pelo processo de tradução, a partir de uma perspectiva multidisciplinar envolvendo língua e tecnologia (PIRES, 2017). Este objetivo possui, ainda, o fim de integrar técnicas de corpus monomodais existentes para a análise multimodal, vertente presente no grupo de estudos COMPLETT⁵ – Corpus Multilíngue para Pesquisas em Línguas Estrangeiras, Tradução e Terminologia, inserido na linha de pesquisa intitulada “Tradução, Tecnologia, e Multimodalidade”.

Para tanto, o presente estudo propõe-se a testar ferramentas automáticas e semiautomáticas de coleta, anotação e etiquetagem que possam auxiliar e tornar menos dispendiosos os processos para a construção de corpora de documentos multimodais estáticos, de natureza texto-imagem, visto que os processos semi-automatizados de anotação de dados facilitam a “velocidade com que os corpora multimodais são desenvolvidos e analisados” (KNIGHT, 2011, p. 407, tradução nossa), pois geralmente considera-se que “a parte mais trabalhosa para adquirir um corpus multimodal é a anotação dos dados, em particular para a modalidade visual” (FANELLI et al., 2010, p. 70, apud KNIGHT, 2011, p. 397, tradução nossa).

Na seção seguinte, define-se os conceitos-chave de incompatibilidades intersemióticas e sua relação com o uso de tradutores automáticos.

2. INCOMPATIBILIDADES INTERSEMIÓTICAS GERADAS POR TRADUÇÃO AUTOMÁTICA

Segundo Pires (2017), uma incompatibilidade intersemiótica em tradução automática (TA) é resultado de uma nova configuração semântica entre os modos verbais e visuais em relação aos mesmos modos encontrados no documento original, ou seja, o resultado gerado através de uma tradução automática, que relaciona

⁵ COMPLETT é um grupo de pesquisas acadêmicas formado por docentes e discentes do Instituto de Letras (IL) 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, bem como as línguas de sinais, para desenvolver corpora multimodais que possibilitem analisar as estruturas e construções realizadas por tecnologias de tradução como também conhecer a utilidade e importância desses dados na formação de tradutores e intérpretes que atuam nessas áreas”. Para maiores informações, vide <dgp.cnpq.br/dgp/espelholinha/4492045805651960582297>.

semanticamente imagem e texto (ibid., p. 73-74), sendo “modo” aqui descrito como todo aspecto visual passível de apresentar informação (BATEMAN, 2008, p.1). Tais modos são concebidos multimodalmente, já que as interações semióticas, verbal e visual, recebem um papel definido e igual a desempenhar (KRESS; VAN LEEUWEN, 2006, p. 113). Com isso, parte-se do “pressuposto de que há relações de significado entre elementos linguísticos e visuais possíveis de serem investigados no contexto onde uma tradução automática foi gerada” (PIRES, 2017, p. 19).

A multimodalidade relaciona-se com o fenômeno intersemiótico na análise de textos compostos, visto que qualquer texto cujos significados são percebidos através de mais de um código semiótico é multimodal (KRESS; VAN LEEUWEN, 2006, p. 177). A multimodalidade em si é definida por Kress e Van Leeuwen (2006) como uma abordagem de caráter interdisciplinar baseada em teoria comunicativa e semiótica social que caminha em direção a uma teoria, e segundo Jewitt (2009), preocupa-se com o entendimento e representação isoladamente e interconectadamente de formas comunicacionais, como, por exemplo, imagens.

Bateman (2009, p.1) caracteriza documentos multimodais como uma variedade de modos visualmente interpostos com o intuito de formar “uma coleção orquestrada de modos comunicativos”. Tais modos individuais contribuem para a criação de um único modo, a multimodalidade – o texto, a imagem, e assim por diante -, que consiste na combinação de múltiplos modos dentro de artefatos únicos (ibid., 2009, p. 1), o que torna possível uma multiplicação do conjunto de significados (LEMKE, 1998, p. 6).

Com isso, “a forma como aspectos visuais e verbais são combinados pode revelar significados em potencial para serem traduzidos” (PIRES, 2017, p.17). No entanto, ainda que o uso de tradutores automáticos em documentos multimodais seja frequente, por sua rapidez e facilidade de uso, tais ferramentas desenvolvidas até o presente momento são primordialmente programadas para reconhecer somente determinadas unidades linguísticas, conforme explicita Pires (2017, p. 99):

[...] a tradução automática i) não é desenvolvida para reconhecer padrões de ligações coesivas entre modo textual e visual para informar suas traduções (como acontece frequentemente em traduções humanas intersemióticas, tais como nos gêneros de quadrinhos, tradução audiovisual e localização) e ii) [que] tais modificações [modificações na tradução] emergem a partir de

modificações de sentido dentro do modo textual, ocasionadas pelo resultado de tradução automática, ou seja, de uma incompatibilidade (*mismatch*) dentro da estrutura textual que se refere semanticamente a um componente visual, ou parte dele.

Com isso, as relações intersemióticas, como texto-imagem, de um documento multimodal têm o significado originalmente pensado alterado ou parcialmente perdido no processo de tradução (BATEMAN, 2014), resultando em uma incompatibilidade intersemiótica gerada por uso de tradutor automático (PIRES, 2017). Neste 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 mostrado nas figuras 1 e 2, retirado da base de dados desenvolvida em Pires (2017).

No exemplo abaixo, é possível verificar a ocorrência de uma incompatibilidade intersemiótica gerada por uso de tradutor automático. No texto original, redigido em língua inglesa, a sigla “PM” refere-se a *Prime Minister* (Primeiro Ministro em português), a qual faz alusão a David Cameron, o homem que aparece na foto. Já em sua tradução automática, a sigla “PM” foi mantida. No entanto, na língua portuguesa, a sigla “PM” não é comumente utilizada para referir-se a um Primeiro Ministro, e sim à polícia ou a um policial militar, o que 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. Esta 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.

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



Fonte: BBC News. Acesso em <<http://www.bbc.com/news/uk-politics-36615028>>.

Figura 2 - Artigo de notícia “Brexit: David Cameron to quit after UK votes to leave EU”, submetido à tradução automática Google Tradutor.



Fonte: BBC News. Acesso em <<http://www.bbc.com/news/uk-politics-36615028>>.

Desta forma, parte do significado multimodal do texto original traduzido “modifica-se”, pois as máquinas de tradução não são desenvolvidas com o propósito de reconhecer as relações intersemióticas presentes em documentos multimodais (PIRES, 2017, p. 106), como por exemplo a relação semântica entre os modos visuais e verbais

de um documento texto-imagem. Assim, “pressupõe-se que parte do significado originalmente criado entre o componente linguístico e visual se modifica com a tradução” (*ibid*, pág. 17).

Na próxima seção, discute-se o conceito de corpora multimodais, a partir de uma comparação com os corpora monomodais.

3. CORPORA MULTIMODAIS

Um corpus, de maneira geral, pode ser definido, de acordo com Berber Sardinha (2004, p. 18), como:

Um conjunto de dados linguísticos sistematizados segundo determinados critérios suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de alguns de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.

Ainda segundo o autor (*ibid.*, 2004, p. 18-19), a definição supracitada seria a mais “completa”, pois menciona aspectos importantes referentes à:

- i. Origem: os dados devem ser autênticos;
- ii. Propósito: o corpus deve ter a finalidade de ser um objeto de estudo linguístico;
- iii. Composição: o conteúdo do corpus deve ser criteriosamente escolhido;
- iv. Formatação: os dados do corpus devem ser legíveis por computador;
- v. Representatividade: o corpus deve ser representativo de uma língua ou variedade linguística;
- vi. Extensão: o corpus deve ser vasto para ser representativo;

Além disso, os estudos de corpora compartilham certas características, como: a) são empíricos e analisam os padrões reais de uso em textos naturais, b) fazem uso extensivo de computadores na análise, empregando técnicas automáticas e interativas, e c) dependem de técnicas quantitativas e qualitativas (BERBER

SARDINHA, 2004, p. 38). Ademais, compartilham características quanto ao i) design e infraestrutura - metodologia usada para coletar, compilar, anotar e representar dados, ii) tamanho e escopo, iii) naturalidade, e iv) (re) usabilidade (KNIGHT, 2011, p. 394).

A linguística de corpus, historicamente, está condicionada à tecnologia, pois esta permite o armazenamento e exploração de corpora, e portanto, está relacionada à disponibilidade de ferramentas computacionais para análise de corpus (BERBER SARDINHA, 2004, p. 15). Os corpora multimodais compartilham as características anteriormente citadas com os corpora monomodais, a exceção do quesito “extensão”, conforme será discutido na seção seguinte.

Apesar de apresentarem pontos em comum, os corpora multimodais se diferem dos monomodais em alguns pontos cruciais, entre eles, a natureza dos documentos analisados. Um corpus monomodal lida com dois modos principais: o modo falado, composto de porções de fala transcritas, e o modo escrito, composto de textos escritos, impressos ou não (BERBER SARDINHA, 2004, p. 20), sendo “modo” definido neste contexto como todo aspecto visual passível de apresentar informação (BATEMAN, 2008, p.1). Já um corpus multimodal lida com documentos compostos por mais de um modo, ou seja, formados por documentos multimodais, artefatos que contêm uma variedade de modos visuais simultaneamente organizados “para preencher uma coleção orquestrada de objetivos comunicativos interligados entre si” (BATEMAN, 2008, p.1). Este é o caso, por exemplo, dos documentos procedimentais presentes na base de dados desenvolvida em Pires (2017), extraídos do site *Wikhow*, nos quais ambos texto e imagem são necessários para o entendimento do processo apresentado. Dentre os corpora multimodais, grande parte é formada por estudos que visam a linguagem corporal, como gestos de mão e olhar, compostos por estudos de vídeo, conforme exemplifica a tabela a seguir:

Figura 3 - “An index of multimodal corpora”.

Name, Reference(s), Language	Size, Composition and Additional Information
AMI Meeting Corpus. Ashby et al., 2005. Corpus of European English.	100 hours of recordings taken from 3 different meeting rooms. This corpus was created for the use 'of a consortium that is developing meeting browsing technology'.
CID (Corpus of Interactional Data). Bertrand et al., 2006; Blache et al., 2008. Corpus of French.	8 hours of dyadic conversations, comprised of 2 participants sat in close proximity of one another, each wearing a microphone headset. Participants were encouraged to chat informally, so with no directions on how to structure the talk.
CUBE-G corpus. Rehm et al., 2008. Corpus of Japanese and German.	Dyadic conversations involving Japanese and German speakers (this is a cross cultural corpus). One participant is an actor, whose contributions are scripted/ instructed.
Czech Audio-Visual Speech Corpus/Corpus for recognition with Impaired Conditions, Železný et al., 2006; Trojanová et al., 2008. Corpus of Czechoslovakian.	Developed to test and train the 'Czech audio-visual speech recognition system' (automatic speech recognition). The first corpus features 25 hours of audio-visual records from 65 speakers. The second has 20 hours of data across 50 speakers. In both each speaker was instructed to read 200 sentences, in laboratory conditions (50 common sentences; 150 were speaker specific).
D64 Corpus. Campbell, 2009. Corpus of British English.	4-5 people recorded over two 4 hour sessions across two days. Non-directed and spontaneous conversations in a domestic environment. Participants wore reflective sticky markers to track movement.
Fruits Cart Corpus. Aist et al., 2006. Corpus of American English.	104 videos of 13 participants (4-8 minutes each). Approximately 4000 utterances in total. Comprised of task-orientated dialogues in an academic setting. Designed to explore language comprehension, now used to analyse language production (NLP).
Göteborg Spoken Language Corpus. Allwood et al., 2000. Corpus of Swedish.	Small components of this 1.2 million word spoken language corpus have been aligned with video records. This corpus contains conversations from different social contexts with a range of different speakers talking spontaneously (i.e. non-directed or scripted).
IFADV Corpus, Van Son et al., 2008. Corpus of Dutch.	A free dialog video corpus composed of face-to-face interaction between close friends/ colleagues. This corpus is comprised of twenty 15 minute conversations (5 hours in total).
MIBL Corpus (Multimodal Instruction Based Learning), Wolf and Bugmann, 2006. Corpus of British English.	Human-to-human instruction dialogues, with one participant teaching a card game to the other (similar to map task activities, see the Map Task Corpus, Anderson et al., 1991 and the Danish DanPASS map task corpus, Grønnum, 2006). This corpus links speech to movement on the screens and is used to train service robots.

Mission Survival Corpus 1 (MSC 1), Mana et al., 2007/ Corpus of Italian.	A meeting corpus which includes a range of short meetings, with up to 6 participants in each. The topics and tasks covered in the meetings are controlled but not scripted.
MM4 Audio-Visual Corpus. McCowan et al., 2003. Corpus of European English	Features 29 short meetings between 4 people filmed in controlled, experimental conditions. The majority of the meetings were scripted and cover specific, predetermined topics and tasks.
NIST Meeting Room Phase II Corpus . Garofolo et al., 2004. Corpus of American English.	Part of the NIST MDCL (Meeting Data Collection Laboratory). This corpus contains 15 hours of recordings from 19 meetings; comprised of both scenario-driven meetings and 'real' meetings.
NMMC (Nottingham Multimodal Corpus). Knight et al., 2009. Corpus of British English.	250,000 words; 50% single speaker lectures, 50% dyadic academic supervisions. Sessions were video and audio recorded, transcribed and aligned using DRS (the Digital Replay System).
SaGA (Bielefeld Speech and Gesture Alignment Corpus). Lücking et al., 2010. Corpus of German.	This corpus contains 280 minutes of audio/video recorded data comprised of 25 direction giving and sight description dialogues based in a Virtual Reality environment. The data is considered to be 'naturalistic' as content is spontaneous, though controlled/prompted.
SK-P 2.0 SmartKom multimodal Corpus. Schiel et al., 2002. Corpus of German.	This corpus contains 96 different single users who were video/audio recorded across 172 sessions. Each user carried out specific, prompted tasks and was recorded in public spaces such as cinemas and restaurants. This corpus is effectively HCI based.
SmartWeb Video Corpus (SVC). Schiel and Mögele, 2008. Corpus of German.	99 recordings of human-human-machine dialogue, with 1 speaker interacting with a human person and a dialogue system (i.e. the main participant is using a Smartphone, which records their face and they are talking to the other participant).
UTEP ICT . Herrera et al., 2010. Corpus of American English.	Cross cultural corpus involving task-based conversations between groups of 4 participants who are stood in a room and free to move around. 200 minutes of data.
VACE Multimodal Meeting Corpus . Chen et al., 2005. Corpus of American English.	Comprised of meeting room 'planning sessions'. Spontaneous talk in controlled task-based environments. 5 participants in each scenario, with 5 scenarios recorded.

Fonte: Knight, 2011a, p. 6-7.

Os resultados encontrados em Pires (2017), cuja base de dados multimodal é utilizado nesta investigação como base para o teste de ferramentas, apresenta um estudo pioneiro, pois ademais de utilizar documentos multimodais de texto-imagem em sua composição, trata também das relações intersemióticas entre elas, além do fenômeno da

incompatibilidade entre tais modos gerada pelo uso da tradução automática. Além de Pires (2017), dentre outros poucos investigadores a tratarem a multimodalidade entre texto e imagem em corpora de documentos estáticos estão Célia Magalhães, Andrea Leitão e Daniela Fernandes (2017), com os estudos de *picture books*, e Tuomo Hiippala (2013), com o estudo de brochuras turísticas da cidade de Helsinque.

Na subseção a seguir, destacam-se as definições de coleta, etiquetagem e anotação em corpora, seguido de uma análise das diferenças de tais processos em corpora mono e multimodais.

3.1. Coleta, etiquetagem e anotação

As etapas de coleta, etiquetagem e anotação em um corpus multimodal estão presentes na construção e análise dos dados que constituem tal corpus. A coleta é composta pela compilação dos dados, ou documentos, que farão parte do corpus. É neste estágio em que muitos dos passos enumerados por Berber Sardinha (2004, p. 18-19) (vide seção 3) são estudados, como a origem dos dados, sua composição, formatação e, de certo modo, a extensão do corpus. A coleta da base de dados multimodais extraídas de páginas *web* desenvolvida em Pires (2017), por exemplo, utiliza o processo semi-automático 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. Para tal coleta, Pires (2017) utilizou os programas Skitch⁶ e a extensão Fireshot⁷.

A etiquetagem de um corpus “consiste na inserção de informações referentes a cada unidade texto”, como por exemplo, estruturas morfológicas, sintáticas, semânticas, discursivas (BERBER SARDINHA 2004, p. 150), as quais podem ser quantificadas. Em corpora monomodais de texto, a etiquetagem pode ser executada por computador, de maneira automática, o que permite o tratamento de grandes quantidades de texto de maneira rápida (ibid., 2004, p. 150). Já em corpora multimodais, seja de vídeo (documento dinâmico) ou imagem (documento estático), o processo de etiquetagem é, geralmente, feito de forma semi-automática ou manual, pois 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

⁶ Acesso em <<https://evernote.com/products/skitch>>.

⁷ Acesso em <<http://screenshot-program.com/>>.

suas características. Ademais, os etiquetadores de corpora monomodais de texto operam, em sua maioria, em formato XML ou TXT, para que seja possível a leitura do texto pela máquina, enquanto as imagens são, geralmente, de formato JPEG ou PNG, e, portanto, não podem ser codificadas da mesma maneira, ou “lidas” por estes mesmos etiquetadores automáticos.

Tendo-se em vista que “a criação de um etiquetador computacional novo é uma tarefa complexa que exige, entre outras coisas, grande conhecimento de programação na área de Processamento de Linguagem Natural” (BERBER SARDINHA 2004, p. 145), uma opção viável para pesquisadores trabalhando em corpora multimodais cujo conhecimento em programação é nulo ou limitado é realizar a etiquetagem de forma manual ou semi-automática. Para tanto, existem funções presentes em ferramentas de uso tanto especializado, ou seja, programas que foram desenvolvidos para o estudo em corpora, quanto ferramentas não especializadas, ou seja, que não foram programadas para servirem de auxílio aos estudos de corpora, que permitem diferentes formas de etiquetagem (vide seção 5).

Já a anotação de um corpus “compreende a inserção de cabeçalhos informativos nos arquivos” (BERBER SARDINHA, 2004, p. 145). Por cabeçalhos, o autor (ibid., p. 145) entende “trechos demarcados contendo informação não veiculada verbalmente no evento comunicativo”, e cita como exemplos a inserção de detalhes adicionais aos dados, como autoria e tipologia. Similarmente ao processo de etiquetagem, a anotação nos corpora de documentos multimodais estáticos de texto-imagem ainda não possui um sistema padrão, ou seja, cada pesquisador desenvolve um esquema próprio de anotação. Tal anotação pode ser feita a partir da inserção de caixas de texto ou formas geométricas, por meio do modo edição de imagem, presente em ferramentas não especializadas como, por exemplo, Nimbus Capture e Evernote (vide seção 5).

Em seus estudos, Knight (2011b, p. 394) descreve que, semelhante aos corpora monomodais atuais, os meios como conteúdos são coletados e registrados, assim como o tamanho dos corpora multimodais são “altamente dependentes das metas e objetivos” de cada pesquisa. Assim posto, Knight (ibid., 2011b) exalta a existência de uma gama de diferentes corpora multimodais e projetos de pesquisa relacionados, mas que independentemente do tipo de estudo, os corpora, de maneira geral, possuem características próprias em relação ao design e à infraestrutura, nomeadamente

metodologias próprias de coleta, compilação (coleção de textos), anotação e representação de dados.

Knight (2011b, p. 396-397) ressalta ainda que o tamanho de corpora multimodais não é “relativamente grande” em comparação com seus equivalentes monomodais. Por exemplo, corpora multimodais formados por milhões de documentos, sejam eles vídeos ou imagens, ainda não existem (ibid., p. 403). Tal fato deve-se a natureza dos dados que compõem os corpora multimodais, que diferem dos corpora monomodais nas seguintes questões, conforme aponta Knight (2011b, p. 396-397, tradução nossa⁸):

Obviamente, a vantagem de usar o discurso baseado em texto na compilação de corpora é que grandes quantidades de dados estão prontamente disponíveis, já podem ser lidas por máquina e / ou relativamente fáceis de serem obtidas, portanto o processo de montagem de tais bancos de dados é relativamente simples.

Porém, como ressaltado nos achados de Rohlfing et al. (2006, p. 99), uma gama de ferramentas tornou-se disponível para anotação de dados de corpora de documentos multimodais dinâmicos, compostos por áudio-vídeo. Tais ferramentas podem tornar menos dispendioso o processo de construção de corpora multimodais, como ressalta Knight (2011b, p. 406), visto que os corpora de multi-contexto dependem de a) avanços tecnológicos, b) constante aprimoramento de sistemas que possibilitem a captura e estruturação da linguagem natural em uso e c) softwares que promovam a manipulação de diferentes conjuntos de dados multimodais.

Apesar disso, “nenhuma ferramenta de anotação multimodal é isenta de desvantagens em relação a certos objetivos analíticos” (DUNCAN, 2013, p. 1020). Por exemplo, Rohlfing et al. (2006, p. 99) salientam o desafio posto aos pesquisadores para determinar se certa ferramenta é adequada para seu conjunto de dados, questão de pesquisa ou computador disponível, e adiciona:

Para decidir sobre utilidade e usabilidade, é necessário saber sobre a facilidade de uso, pontos fortes / fracos para as finalidades de anotação específicas e o tipo de dados ou análise para a qual a ferramenta foi projetada

⁸ Do original: “Obviously, the advantage of using text-based discourse in the compilation of corpora is that large quantities of data are readily available, already machine-readable and/or relatively easy to get hold of, so the process of assembling such databases is relatively straightforward.” (KNIGHT, 2011b, p. 396-397)

- conhecimento que geralmente só é obtido após se tornar um especialista no uso de uma ferramenta em particular.⁹

Há, ainda, o fato de que grande parte dos programas de etiquetagem e anotação multimodais são serviços pagos, como explicitado em Knight (2011b, p. 408), o que pode inibir a acessibilidade de tais ferramentas para certos usuários. Além disso, “informações sobre limitações [...] geralmente não são divulgadas pelos desenvolvedores de ferramentas, mas teriam valor para os usuários de ferramentas em potencial” (ROHLFING et al., 2006, p. 100). Ainda segundo Rohlfling (ibid., p. 101, tradução nossa¹⁰):

Entre as ferramentas que suportam a análise de dados de discurso multimodal, os editores de mídia e texto, embora considerados de baixa tecnologia, ainda podem ter a base de usuários mais ampla, embora sejam geralmente considerados de baixa tecnologia. Isso é verdade, apesar da crescente disponibilidade de programas de computador que integram recursos de visualização, anotação e análise de dados digitalizados de muitos tipos [...]. Muitas dessas interfaces de softwares são altamente configuráveis pelo usuário e oferecem capacidades analíticas muito além daquelas oferecidas por editores de mídia e texto. [...] Por editores de texto, queremos dizer softwares de processamento de texto e também calculadoras de planilhas.

Por fim, até o presente momento, não há uma convenção que determine um esquema de anotação ou uma forma padrão de como integrar as pesquisas em um banco de dados que cubra todos os elementos multimodais do discurso (ABUCZKI; GHAZALEH, 2013, p. 94).

Em relação à anotação e etiquetagem em corpora multimodais compostos por documentos estáticos de texto-imagem, como o proposto em Pires (2017), base de dados do presente estudo, há ainda escassos estudos, principalmente devido a sua abordagem distinta. O sistema de anotação e etiquetagem presentes no corpus de Hiippala (2013), por exemplo, utiliza as habilidades experientes de seu desenvolvedor, o qual possui foco em estudos de programação e tratamento de linguagem natural, o que

⁹ Tradução nossa, do original: “to decide about usefulness and usability, it is necessary to know about the ease of use, strengths/weaknesses for specific annotation purposes, and the type of data or analysis the tool is designed for – knowledge that is usually gained only after becoming an expert in the use of a particular tool.” (ROHLFING et al., 2006, p. 99)

¹⁰ 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, though they are generally regarded as “low-tech”. This is true despite the increasing availability of computer programs that integrate visualization, annotation, and analysis capabilities for digitized data of many types [...]. Many of these software interfaces are highly user-configurable and offer analytic capabilities well beyond those afforded by media and text editors. [...] By text editors, we mean word processing software and also spreadsheet calculators.” (ROHLFING et al., 2006, p. 100)

torna sua abordagem de difícil entendimento para linguistas com pouco ou razoável entendimento de programação, tornando sua replicação ou adaptação de técnicas para outros corpora demorada devido ao tempo necessário para a compreensão dos mecanismos. O estudo de Célia Magalhães, Andrea Leitão e Daniela Fernandes (2017), que lida com livros ilustrados infantis, ou seja, textos intermodais cuja narrativa é representada pelas linguagens visual e verbal, apesar de lidar com a multimodalidade, enfoca nos estudos da linguagem verbal. Para tanto, a análise foi realizada através da transcrição de sentenças, de forma automática, pelo software UAM CorpusTool (O'DONNELL, 2016), sendo a contabilização dos dados feita com o auxílio do software Microsoft Excel (2016). Neste contexto, texto e imagem foram, portanto, analisados separadamente.

Na abordagem de Pires (2017), a análise dos documentos multimodais estáticos, assim como o processo de anotação, foi feita em um documento único, em que texto e imagem não são separados, conforme exemplificado na figura 4, retirada da base de dados presente em Pires (2017).

Figura 4 –Exemplo de anotação (PIRES, 2017)



Fonte: Pires (2017).

Nela, vê-se um exemplo esquemático de anotação em um documento texto-imagem. Na cor vermelha, Pires ressalta uma “anotação de imagem”, feita a partir da inserção de formas geométricas, indicando uma incompatibilidade intersemiótica gerada por tradução automática. Na cor azul, o pesquisador marca uma anotação verbal, feita a partir da inserção de uma caixa de texto, salientando a natureza da incompatibilidade, ou seja, uma classificação dada a esta incompatibilidade.

Na próxima seção, apresenta-se a metodologia utilizada para fins deste estudo, nomeadamente a descrição das fontes de dados e os procedimentos metodológicos e de análise.

4. METODOLOGIA

4.1. Descrição da fonte de dados

O percurso metodológico do presente trabalho possui caráter qualitativo, pois estuda as particularidades de cada ferramenta eleita para teste, e exploratório, tendo-se em vista que o teste de ferramentas para coleta, anotação e etiquetagem de corpora de documentos multimodais estáticos possui ainda escassa pesquisa investigativa (PIRES, 2017). Para tanto, a base de dados utilizada para o teste dos programas encontrados, a fim de verificar se tais processos podem ou não tornar mais ágil o processo de construção de tais corpora, é aquela presente em Pires (*ibid.*), cujo trabalho tem como resultado uma base de dados multimodal¹¹ originalmente composta por 15 pares de documentos multimodais extraídos de *web pages*¹², traduzidos automaticamente do inglês para o português por meio da ferramenta online *Google Translate*, para estudo das relações texto-imagem nas quais o texto tenha sido resultado de erro de tradução automática (PIRES, 2017).

Pires (2017) salienta a importância do material escolhido para a análise de incompatibilidades em componentes multimodais, já que os documentos presentes no corpus, textos e artigos de cunho jornalístico, educativo e procedimental, por conterem recursos semióticos diversos, têm o potencial de apresentarem novas configurações intersemióticas ao serem traduzidos automaticamente (PIRES, 2017, p. 103).

¹¹ Base de dados disponível em <<https://www.dropbox.com/s/g7ky2traywqxbek/corpus.zip?dl=0>> (PIRES, pág. 120, 2017).

¹² Todos os direitos autorais foram mantidos, conforme explicitado em Pires (2017).

A escolha dos dados retirados da *World Wide Web* baseia-se no fato de que muitos de seus conteúdos utilizam-se de elementos visuais e verbais de forma a criar uma unidade semântica multimodal (PIRES, 2017, p. 102), como artigos de notícias e tutoriais. Nesse caso, teve-se como critério a escolha de documentos em que texto e imagem foram “instanciados visualmente e intencionalmente co-presentes dentro de uma composição conjunta estática”, fazendo com que texto e imagem atuem como uma única unidade de composição (BATEMAN, 2014, p. 28). Já a escolha da ferramenta *Google Translate* é justificada por sua representação de “um amplo meio de traduzir automaticamente conteúdo de *websites*” (PIRES, 2017, p. 102), além de sua compatibilidade com o navegador escolhido para a pesquisa, o *Google Chrome*.

4.2. Procedimentos metodológicos e de análise

O presente trabalho toma como base os procedimentos metodológicos e de análise presentes na pesquisa de Rohlfig et al. (2006) intitulada comparação de ferramentas de anotação multimodal¹³, um *workshop* organizado durante o Segundo Congresso da Sociedade Internacional para os Estudos de Gesto¹⁴, em junho de 2005 na cidade de Lyon. Neste *workshop*, foram apresentadas informações e demonstrações de algumas das ferramentas mais utilizadas para anotação e etiquetagem de corpora de documentos multimodais dinâmicos, mais especificamente aqueles compostos por análises de vídeos e transcrição de falas (áudio).

Nesse evento, cujo público alvo era composto por desenvolvedores, usuários especialistas e pesquisadores interessados em usar as ferramentas apresentadas, Rohlfig et al. (2006, p. 100) desenvolveram duas atividades principais, a primeira para que os usuários apontassem pontos positivos das ferramentas, nomeada *freestyle assignment*, e a segunda para descobrir pontos negativos das mesmas, nomeada *compulsory assignment*. Estas atividades tiveram a duração total de duas horas.

Nessa primeira atividade, intitulada *freestyle assignment*, os usuários testaram as funções de anotação e análise de dados a partir de um conjunto de dados de escolha própria, não disponibilizado pelo *workshop*. Já na segunda atividade, intitulada *compulsory assignment*, os usuários também testaram os sistemas de anotação, porém

¹³ Tradução nossa, do original “Comparison of multimodal annotation tools”

¹⁴ Tradução nossa, do original “Second Congress of the International Society for Gesture Studies”

em um conjunto de dados comum, disponibilizado pelo *workshop*. Este conjunto de dados era formado por documentos de áudio-vídeo cujo conteúdo mostrava a interação entre duas ou mais pessoas, em temáticas diversas. Segundo Rohlffing et al. (2006, p. 100), o intuito desta atividade era o de realizar uma comparação de usabilidade e funcionalidade entre as ferramentas, como por exemplo descobrir quais ferramentas eram capazes de mostrar o vídeo sem relatório de erros, quais não apresentavam suporte para o formato do vídeo, entre outras questões.

Para tanto, os pesquisadores foram guiados por quatro perguntas em relação às ferramentas testadas (Rohlffing et al., 2006, p. 101, tradução nossa), a seguir:

- a) Quão difícil é, e quanto tempo leva para aprender a usar as ferramentas?
- b) Com que rapidez os dados podem ser anotados? (Velocidade de execução)
- c) Carga mental (*mental load*): o usuário tem que pensar com cuidado e ter muitas informações em mente ao usar esta ferramenta?
- d) Com que frequência ocorrem erros e quão graves eles são?

Na pesquisa de Rohlffing et al. (2006), tais perguntas foram respondidas com a ajuda de usuários, profissionais e amadores, sobre as ferramentas testadas. Vale ressaltar que, apesar de tal participação, a investigação não se trata de um estudo de usuário. As quatro perguntas foram feitas para guiar os testes em relação ao que os autores (ibid., 2006) chamaram de usabilidade, seção destinada ao estudo das funções e limitações de cada ferramenta escolhida em relação ao seu uso em anotação e etiquetagem em documentos multimodais dinâmicos. Ademais, os autores (ibid.) também realizaram uma tabulação dos dados relacionados a outros quesitos das ferramentas, como informações gerais, modos de controle e informações técnicas (vide anexo A).

No presente estudo, optou-se por seguir estas perguntas de usabilidade, ainda que de forma adaptada. Nesta investigação, os testes foram realizados por somente um usuário, voltados para o contexto de anotação e etiquetagem em

documentos multimodais estáticos de texto-imagem. Levaram-se em conta os tópicos supracitados, em que i) a dificuldade de uso é 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, ii) o tempo para fazer anotações ou etiquetas faz parte do quão prolongado é para adicionar, excluir ou alterar tais processos, se estes são feitos individualmente ou se é possível reutilizar o mesmo esquema em diversos documentos, iii) a carga mental é o grau de dificuldade para que se possa utilizar a ferramenta em teste, como comandos, conhecimentos prévios de programação, entre outros, e iv) em que a frequência de erros leva em consideração problemas técnicos como relatórios de erros, congelamento de tela, fechamento involuntário do programa, entre outros. Ademais, a presente investigação toma igualmente como base a tabulação¹⁵ apresentada como produto final do *workshop* de Rohlfing et al. (ibid., p. 118-121), porém com adaptações para a testagem de ferramentas que possam agilizar o processo de coleta, etiquetagem e anotação em documentos multimodais texto-imagem cujo conteúdo apresente incompatibilidade(s) intersemiótica(s) gerada por uso de tradução automática, conforme descrito em Pires (2017).

Na próxima seção, seguem os resultados obtidos a partir da testagem das ferramentas escolhidas, a partir de uma adaptação dos achados em Rohlfing et al. (2006), conforme descrito na subseção 4.2.

5. ANÁLISE DOS RESULTADOS E DISCUSSÃO

Nesta seção, discute-se as vantagens e limitações das ferramentas elegidas para testagem nos quesitos coleta, etiquetagem e anotação em corpus multimodal de texto-imagem. Para a escolha prévia das ferramentas, levou-se em conta a gratuidade total ou parcial das funções de análise. A lista de programas estudados é esta:

1. Nimbus Capture (<https://nimbusweb.me>)
2. Evernote (<https://evernote.com/>)
3. UAM ImageTool (<http://www.wagsoft.com/ImageTool/>)

¹⁵ Para visualização da tabela, vide anexo A.

A avaliação de cada ferramenta foi baseada nos estudos de Rohlfin et al. (2006), apresentados na seção anterior, em que leva-se em conta quatro etapas principais em relação a usabilidade e funcionalidade.

5.1. Nimbus Capture

O programa Nimbus Capture é uma extensão de navegador¹⁶ cuja principal função é a captura de telas, mas que se destaca por oferecer, também, esquemas de anotação e etiquetagem de documentos multimodais estáticos (vide seção 3.1). Neste estudo, optou-se por esta ferramenta não especializada, ou seja, que não foi programada para fins específicos de investigação em corpora, pois, “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, 2006. p. 101).

5.1.2. Usabilidade

A extensão é de uso fácil, pois o usuário conta com interfaces simples e intuitivas, ou seja, que não necessitam de um estudo prévio, o que torna o tempo para aprender a usá-la mínimo ou nulo, ou seja, a “carga mental” (vide seção anterior) do usuário, ao utilizá-la, é ínfima. Ademais, erros de mau funcionamento ou “congelamento” de tela ao utilizar esta extensão são raros, tornando-a uma ferramenta de alta confiabilidade. O programa Nimbus Capture, além disso, oferece uma versão gratuita, o que o torna acessível a um público diverso.

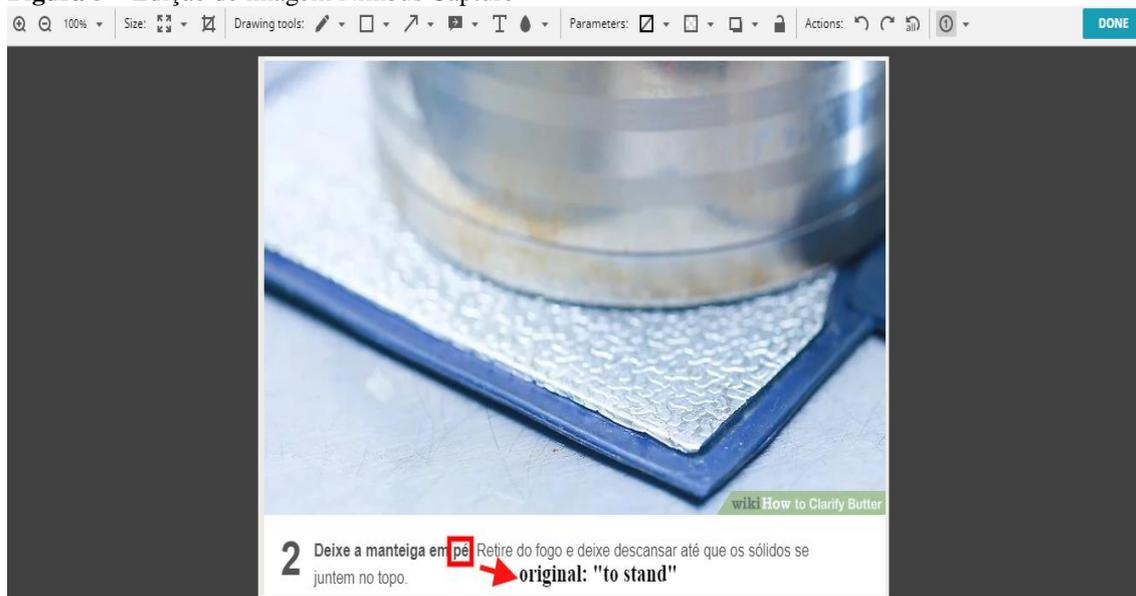
Uma das maiores vantagens desta para com a coleta de arquivos é a opção de salvar as capturas em diversos formatos, como PDF, JPEG e PNG, no computador do usuário, sendo possível nomeá-las, fato que torna menos dispendioso o armazenamento dos documentos de maneira organizada, para possível busca posterior e formatação do corpus. Além disso, esta extensão não requer um grande espaço de memória, sendo possível e fácil sua instalação em qualquer máquina conectada ao serviço *online*.

A extensão Nimbus Capture, ademais da captura de telas, oferece também opções de edição de imagem. Esta função permite a anotação dos documentos texto-

¹⁶ As extensões de navegador são definidas, segundo a página Google Chrome Extensions, como “pequenos programas de software que personalizam a experiência de navegação”.

imagem como a adição de formas e textos, além da possibilidade de borrar partes selecionadas, caso queira-se dar enfoque a somente certas informações do documento ou excluir informações adicionais. Após a captura, a qual o usuário pode escolher entre tela “cheia”, ou seja, capturar a página como um todo, ou somente parte do documento, a ferramenta Nimbus Capture permite fazer anotações no documento gerado, desde marcações, como setas e retângulos, à inserção de texto, como exemplificado na figura 6, em que foram feitas marcações na cor vermelha, para destacar a incompatibilidade intersemiótica gerada por tradução automática, e edição de texto, indicado pela seta.

Figura 5 – Edição de imagem Nimbus Capture



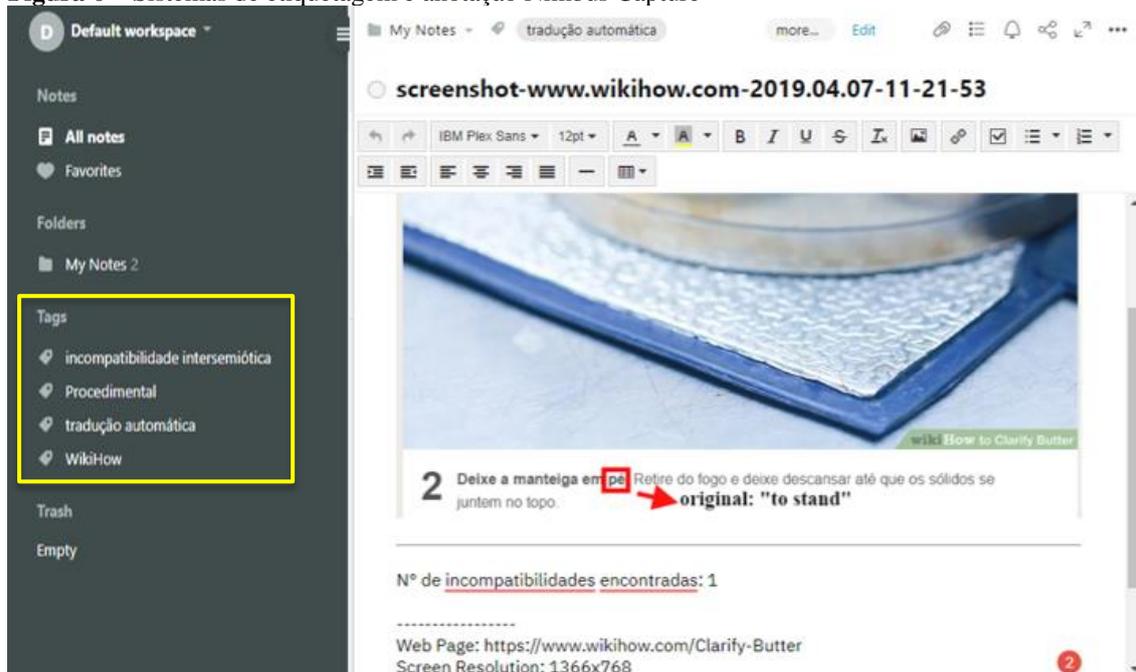
Fonte: Próprio autor.

Ademais das funções de coleta e anotação em documentos multimodais texto-imagem retirados de páginas *online*, a ferramenta Nimbus Capture também oferece um sistema de etiquetagem. Tal sistema é oferecido através da criação de uma conta, totalmente gratuita, pelo site do desenvolvedor, na qual os documentos são salvos no sistema nuvem (*cloud*), ou seja, de forma *online*, sem sobrecarregar o sistema do usuário. Uma vez criado, o cadastro permite ao usuário salvar documentos, sejam capturados com o uso da extensão, sejam previamente salvos no computador, e organizá-los em pastas, as quais é possível (re)nomear a qualquer momento. Ademais das anotações que podem ser feitas com a ferramenta de edição de imagem, a conta também permite fazer anotações na pasta, sem interferência na captura de tela (vide figura 6).

Além disso, ao enviar o documento para a conta, o sistema da Nimbus Capture, como nomenclatura padrão, permite ao usuário saber o endereço do site de onde a captura foi tirada, ademais da data, e caso o usuário queira, informações do sistema, como resolução da tela e a plataforma usada pelo usuário.

As etiquetas, referidas como *tags*, podem ser adicionadas, excluídas ou modificadas a qualquer momento, exemplificadas na figura 6 pela marcação em amarelo. O processo é otimizado pela interface, que grava as etiquetas já cadastradas, de forma que o usuário não precisa escrevê-las todas as vezes em que precisar marcar um arquivo. As *tags* também facilitam a busca e agrupamento de documentos, uma vez que o sistema permite reagrupar dados, conforme etiquetagem, em diversas categorias. Por fim, não há limitação de *tags* por arquivo. Para visualização completa dos achados, vide seção 6.

Figura 6 – Sistemas de etiquetagem e anotação Nimbus Capture



Fonte: Próprio autor.

5.1.3 Limitações

O software Nimbus Capture possui a vantagem de oferecer funções que permitem a coleta, anotação e etiquetagem de documentos estáticos extraídos de páginas *web*. No entanto, por não ser uma ferramenta especializada, ou seja, programada para ser utilizada em estudos de corpora, não possui certas funções de análises comuns a estas, como é o caso das funções estatísticas, ou seja, a extração de padrões ou número

de ocorrências de, por exemplo, um tipo específico de incompatibilidade intersemiótica. Para tanto, o usuário do Nimbus Capture deve realizar esta etapa de maneira semi-automática, pois o programa permite somente a classificação por etiquetas (*tags*).

5.2. Evernote

Evernote é um software gratuito que possibilita a seus usuários capturar, salvar, modificar e compartilhar arquivos. Entre outras funções, este programa também permite a sincronização dos documentos em diversos dispositivos, a partir de uma conta, que pode ou não ser vinculada ao endereço de e-mail do usuário.

5.2.2. Usabilidade

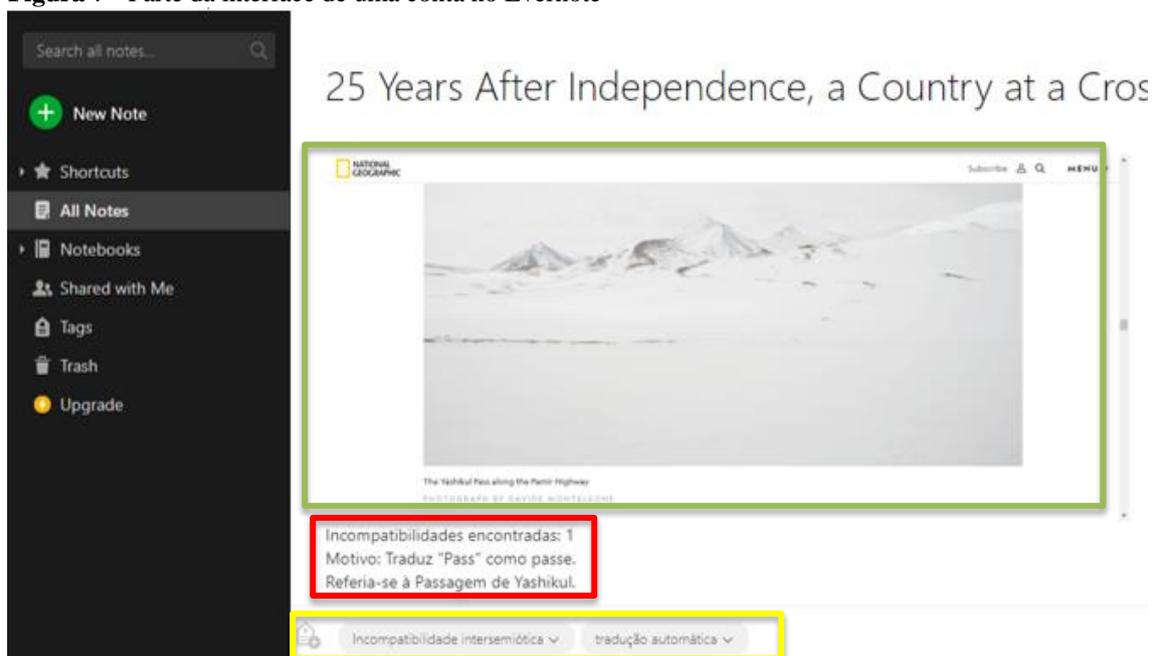
A ferramenta Evernote, seja em sua versão desktop ou aplicativo, possui uma interface simples e intuitiva, sendo de uso fácil mesmo para aqueles que não possuem grande conhecimento em utilitários online, o que significa uma baixa “carga mental” (vide seção anterior). Sua proposta é permitir que seus 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*.

As capturas de tela funcionam de maneira similar a extensão Nimbus Capture. Basta selecionar a página que se queira copiar e selecionar o campo, seja apenas uma parte do conteúdo ou o artigo como um todo. Após a captura, o Evernote permite ao usuário fazer modificações no arquivo, os quais podem ser convertidos nos formatos JPEG, PNG, PDF, entre outros. Tais modificações podem ser adicionadas antes ou após o compartilhamento dos dados com outros usuários da ferramenta. Em relação às ferramentas de edição, estão presentes o sistema de cores, que permitem sublinhar e/ou marcar qualquer ponto da imagem ou texto, a inserção de caixa(s) de texto(s) e opção de borrar informações.

Um diferencial do software Evernote é a opção de inserir etiquetas nos documentos salvos, por meio de uma conta de acesso gratuito. Como exemplificado na imagem abaixo, é possível fazer anotações no arquivo, acrescentar notas em um arquivo ligado ao documento e por fim, adicionar etiquetas (*tags*), as quais ajudam o usuário a localizar, agrupar ou até mesmo quantificar os arquivos, ainda que de forma semi-automática. Outra característica positiva da ferramenta é que erros técnicos, como congelamento de tela ou fechamento repentino, não ocorrem comumente.

Na figura 7, destacam-se as funções de coleta, etiquetagem e anotação de documentos multimodais estáticos extraídos de páginas *web*, feitas com o uso da ferramenta Evernote. Na cor verde, sinaliza-se um exemplo de coleta, realizada com uso da função captura de tela. A etiquetagem, sinalada pela cor amarela, é feita a partir da inserção manual de *tags*. Já a anotação, marcada pelo uso da cor vermelha, foi realizada por meio da inserção de informações adicionais no formato bloco de notas, ou seja, abaixo do documento coletado, de forma similar a uma legenda.

Figura 7 - Parte da interface de uma conta no Evernote



Fonte: Próprio autor.

5.2.3. Limitações

A ferramenta Evernote, apesar de fornecer ferramentas que podem auxiliar o pesquisador na etiquetagem de documentos multimodais estáticos, não é uma ferramenta especializada, ou seja, não foi desenvolvida com este intuito. Similarmente ao software Nimbus Capture, em relação ao processo de etiquetagem, ainda que o programa permita ao usuário escolher quais e com quantas etiquetas marcar um documento, não há uma divisão de projetos. 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 usuário utilize a ferramenta para realizar investigações distintas, pois as informações destas seriam armazenadas pelo Evernote em uma mesma pasta, mesclando os conteúdos.

Além disso, o Evernote não possui um sistema que permita a quantificação automática do sistema de etiquetagem, o que implica em um processo manual por parte do usuário caso este queira extrair dados estatísticos ou verificar a existência de padrões. Por fim, o sistema de etiquetagem, similar ao encontrado na extensão Nimbus Capture, é feito individualmente, ou seja, o esquema de anotação ou etiquetagem não pode ser reutilizado em outros arquivos. Para visualização completa dos achados, vide seção 6.

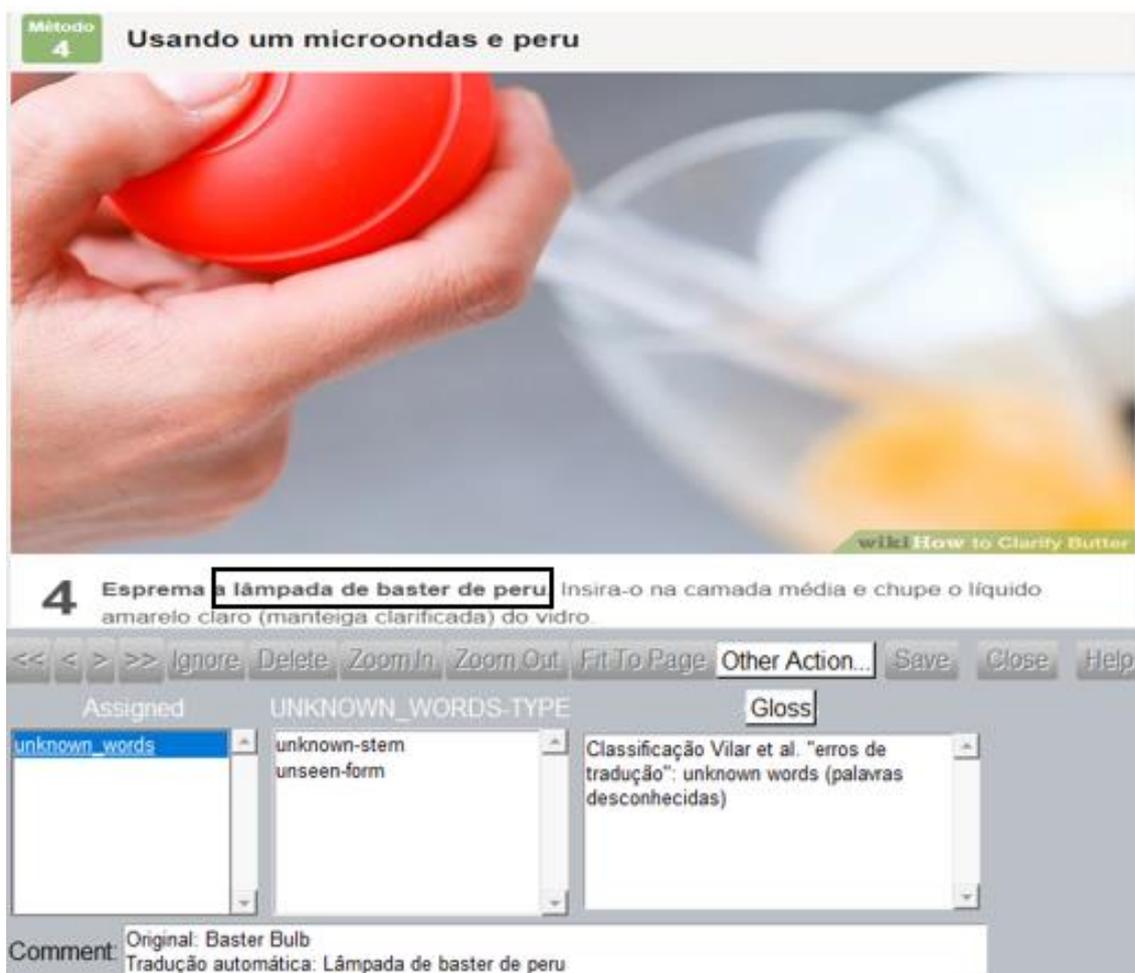
5.3. UAM ImageTool

O UAM ImageTool é uma ferramenta especializada em anotação de imagens, voltada para o desenvolvimento de corpora. O programa permite a etiquetagem ou anotação de esquemas de parte ou todo de uma imagem, ademais de oferecer a possibilidade de “quantificar os tipos de categorias analisadas” (PIRES, 2017, p. 106).

5.3.2. Usabilidade

Diferentemente das demais ferramentas apresentadas, no UAM ImageTool as anotações e etiquetas são armazenadas em arquivos no formato XML, o que torna sua exportação ou compartilhamento para outros programas possível e simplificada, ou seja, menos dispendiosa. Os esquemas de anotações e etiquetagem também são feitos de maneira distinta, pois o programa permite a “classificação cruzada e a subespecificação de recursos” (O’Donnell, 2008, p.15). Esta função implica que é possível fazer correlações e “ligações” entre os documentos, estejam em uma mesma pasta ou não, chamadas de “camadas” (*layers*). Por exemplo, caso seja feita qualquer tipo de alteração no esquema de anotação de uma imagem, como excluir uma classificação, o sistema do UAM ImageTool irá automaticamente atualizar todos os arquivos do corpus que utilizam a mesma anotação, não sendo preciso alterar manualmente cada uma das imagens. Com isso, faz-se possível, também, a anotação de múltiplas imagens reutilizando o mesmo esquema de anotação, o que torna o tempo despendido para tais tarefas menor se comparado ao sistema individual das demais ferramentas analisadas neste estudo.

Figura 8 – Exemplo de anotação e etiquetagem no programa UAM ImageTool

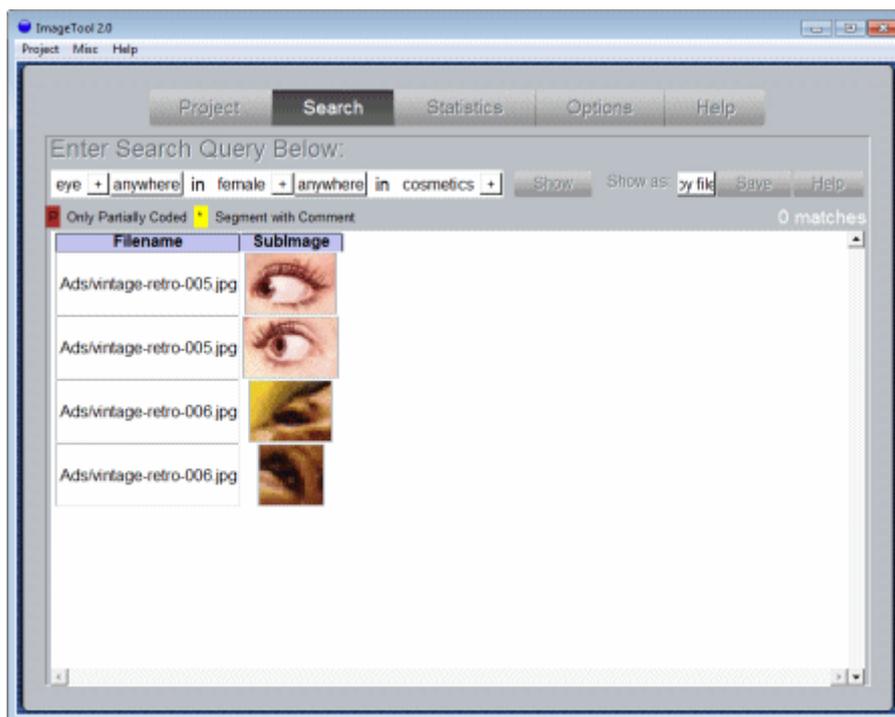


Fonte: Próprio autor.

A anotação de imagens, foco da ferramenta, pode ainda ser feita no que O'Donnell (2008) chama de “níveis”, o que significa que cada imagem pode ser dividida de acordo com o foco da pesquisa em andamento. Como exemplo, O'Donnell cita uma pesquisa realizada a partir do UAM ImageTool, presente em seu website¹⁷, em que a ferramenta de pesquisa do programa separa a imagem em diferentes níveis de anotação, demonstrando que “consultas complexas são possíveis, usando recursos em diferentes níveis de anotação” (ibid., 2008). Abaixo, mostra-se o exemplo prático dado pelo autor (ibid.), em que o programa agrupa partes de imagens conforme as delimitações de pesquisa escolhidas. Neste caso, o trabalho analisava o olhar feminino em propagandas de cosméticos. Além disso, o UAM ImageTool utiliza um sistema de cores para que o usuário possa facilmente identificar quais documentos já foram completamente analisados e etiquetados.

¹⁷Acesso em <<http://www.wagsoft.com/ImageTool/>>.

Figura 9 – O olhar feminino em propagandas de cosméticos.



Fonte: Michael O'Donnell, UAM Image Tools web site.

Por fim, um dos maiores diferenciais da ferramenta em comparação as demais apresentadas é que ela oferece recursos que podem ser utilizados para estudos estatísticos. Segundo seu criador “o UAM Tool suporta uma série de análises estatísticas de corpora, revelando padrões que o olho pode perder” (O'Donnell, UAM ImageTool main page). Tais estatísticas comparativas podem ser feitas entre subconjuntos de imagens, como por exemplo, uma comparação entre as incompatibilidades intersemióticas em diferentes tipos de páginas *web*, como aquelas de cunho procedimental e jornalístico. Tais dados são gerados de forma automática pelo programa, em formato de tabelas e gráficos.

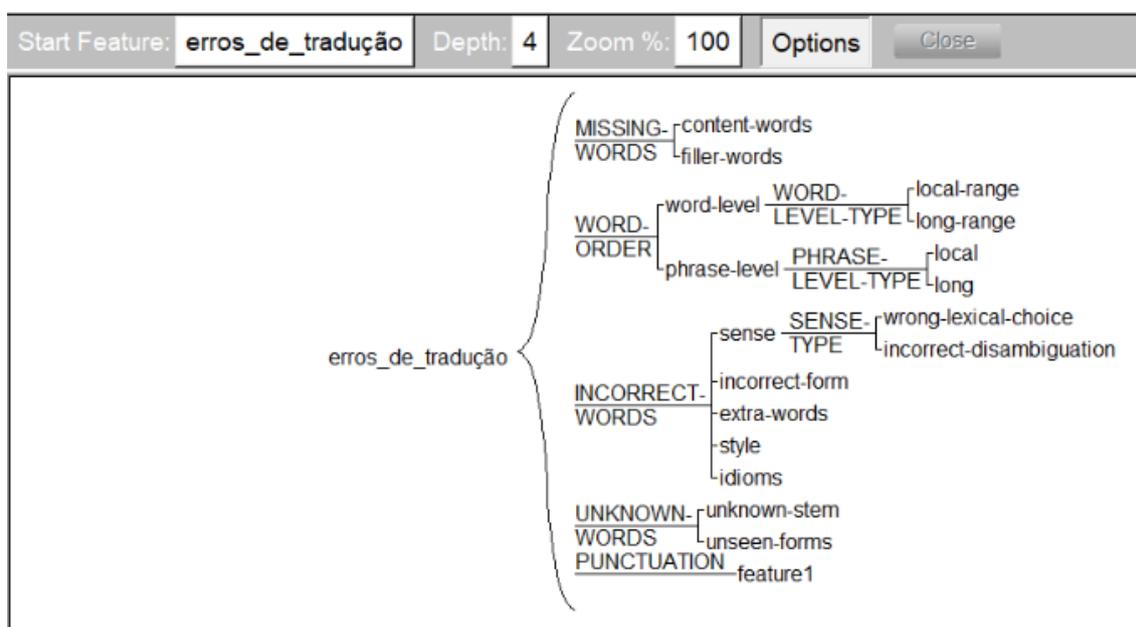
5.3.3. Limitações

Ainda que O'Donnell (2008, p. 16) descreva ao UAM ImageTool como uma ferramenta criada para ser fácil de ser utilizada mesmo por linguistas cujo foco não é a programação, sua usabilidade requer um aprendizado prévio. O desenvolvedor oferece suporte ao usuário por meio de artigos ilustrados, oferecidos nos idiomas inglês, japonês e chinês, em que mostra desde como adicionar imagens para o tratamento na ferramenta a como extrair dados estatísticos a partir das etiquetas ou dos esquemas de anotação, ou por meio de uma lista com as principais funções do programa.

Se comparado às demais ferramentas testadas neste estudo, o UAM ImageTool possui a vantagem de ser o único programa de uso especializado, ou seja, foi projetado com o fim de servir a estudos de Linguística de Corpus. Apesar disso, o tempo despendido para aprender a utilizar o programa desenvolvido por O'Donnell é maior, devido à complexidade de funções da ferramenta, que, por exemplo, os programas Nimbus Capture ou Evernote. No entanto, o programa se assemelha a outros programas voltados para a criação e anotação de corpus, como o próprio UAM CorpusTools ou ainda o AntConc (Anthony, L.), facilitando seu entendimento para pesquisadores que já tenham trabalhado com ferramentas similares, ainda que voltadas para o tratamento de corpora monomodal de texto.

Outra limitação da ferramenta é a ausência da opção “desfazer” (do inglês *undo*), caso o usuário queira retornar a um determinado ponto inicial. Sem esta ferramenta, ações equivocadas devem, necessariamente, ser deletadas manualmente. Um outro aspecto é que no sistema de anotação, não é possível fornecer a mesma nomeação para dois sistemas. Neste caso, por exemplo, não é possível replicar na íntegra o sistema de classificação de erros de tradução de Vilar et al (2005), utilizado por Pires (2017) para a classificação das incompatibilidades intersemióticas encontradas em seu banco de dados, já que a classificação *local* e *long range* está presente em duas ramificações de classificação distintas. Para visualização completa dos achados, vide seção 6.

Figura 10 – Tentativa de réplica da classificação de erros de tradução de Vilar et al (2005) no sistema UAM ImageTool.



Fonte: Próprio autor.

6. 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
Descrições técnicas			
Plataformas suportadas	Google Chrome, Opera, Firefox, Android, Microsoft Edge, Safari	MAC, IOS, Windows e Android	Windows e MacOSX
Limite para o tamanho o tamanho de arquivo?	Sim	Sim	Não
Importa de:	Desktop	Desktop, Skitch	Desktop
Exporta anotações?	Não	Sim	Sim
Futuro suporte para o software	Sim. Atualizações são periódicas pelo site do desenvolvedor	Sim. Atualizações são periódicas pelo site do desenvolvedor	Não encontrado. Última atualização em 2011.
Custo	Gratuito	Gratuito	Gratuito
Modos de controle			
Suporte para múltiplas imagens?	Sim	Sim	Sim
Tipos de anotação definidos pelo usuário?	Sim	Sim	Sim
Pode-se alterar as definições de anotação depois que os dados são anotados?	Sim	Sim	Sim
As anotações podem ser definidas hierarquicamente?	Sim	Sim	Sim

Usabilidade			
Tamanho da base de usuários (informações retiradas dos sites oficiais de cada ferramenta)	1,493,318	225,000,000	n/a
Oferece suporte ao usuário?	Sim, por meio de tutoriais em vídeo e imagem, artigos e manual fornecidos em língua inglesa, sendo alguns artigos redigidos em russo. Opção de pedir suporte ao desenvolvedor.	Sim, por meio de tutoriais em vídeo e imagem fornecidos em mais de 15 idiomas. Opção de pedir suporte ao desenvolvedor ou demais usuários.	Sim, por meio de artigos, fornecidos nos idiomas inglês, japonês e chinês. Opção de entrar em contato com o desenvolvedor.
"Carga mental" (Mental load)	Mínima	Mínima	Média
Velocidade de execução	Rápida	Rápida	Rápida
Frequência de erros	Baixa	Baixa	Baixa
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 a localização 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: próprio autor.

7. CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo principal a exploração de métodos que visam tornar menos dispendioso o processo de etiquetagem de incompatibilidades intersemióticas geradas pelo uso de tradutor automático em corpora multimodais compostos por documentos multimodais estáticos. Para tanto, utilizou-se os fundamentos de Rohlfing et al. (2006) como base para os quesitos de análise, e os achados da base de dados de Pires (2017) como documentos utilizados para a realização dos testes de coleta, etiquetagem e anotação.

Ao todo, foram testados três programas, de usos diversos. A partir desta análise, constatou-se que as ferramentas não especializadas, ou seja, recursos que não foram programados com o fim de servirem como ferramentas de auxílio para estudos de corpora, possuem um maior potencial para agilizar os processos de coleta, anotação e etiquetagem que os programas especializados. Tal fato é baseado principalmente no tempo despendido para que o usuário possa aprender a utilizar a ferramenta, definido por Rohlfing et al (2006) como “carga mental”. Os programas não especializados, de maneira geral, são mais simples e intuitivos que os softwares voltados para o público de linguistas ou linguistas computacionais. Com isso, ferramentas como extensões de navegador e programas de compartilhamento e edição de imagens, tais como Nimbus Capture e Evernote, não requerem conhecimentos prévios de computação e/ou programação para que o usuário possa começar a utilizá-lo, fator que influencia diretamente na rapidez com que os processos são realizados.

Ademais, os programas não especializados testados oferecem recursos que tornam possível os três processos testados, nomeadamente a coleta, a anotação e a etiquetagem em documentos *web* multimodais de natureza estática, com o uso de uma única ferramenta. Tal fato contribui para a rapidez com que os processos podem ser desenvolvidos pelo pesquisador.

Apesar disso, nota-se que os programas não especializados possuem limitações de análise importantes em relação às ferramentas desenvolvidas para o uso em corpus. Por exemplo, as ferramentas Nimbus e Evernote, usadas para o processo de etiquetagem, não oferecem análises estatísticas automáticas, diferentemente do programa UAM ImageTool. Tendo-se em vista que as pesquisas realizadas com o

auxílio de corpora podem ser de caráter tanto qualitativo quanto quantitativo, tal recurso mostra-se imprescindível.

A tabulação de limitações de cada ferramenta corrobora com os estudos de Duncan (2013, p. 1020), em que os achados apontam que nenhuma ferramenta é isenta de desvantagens em relação a certos objetivos analíticos. Tendo-se em vista tais limitações, constatou-se, também, que a combinação de ferramentas não especializadas com o uso de ferramentas especializadas possui um potencial ainda maior de tornar as etapas de coleta, etiquetagem e anotação menos dispendiosas. Tal afirmação leva em consideração a facilidade de anotação e coleta oferecidas pelas primeiras, e as funções semi-automática de etiquetagem e estatística das últimas.

Faz-se importante salientar que similarmente aos achados de Rohlfig et al. (2006, p. 122), o objetivo desta investigação não foi o de apontar uma única ferramenta como uma melhor opção.

Por fim, espera-se que os achados neste trabalho possam ser de utilidade para futuras investigações do grupo de pesquisa COMPLETT na Universidade de Brasília, com destaque para a expansão da base de dados desenvolvida em Pires (2017), trabalhado no segmento de estudo em Corpora Multimodais de documentos estáticos, voltado para a investigação das relações texto-imagem e incompatibilidades intersemióticas geradas por uso de tradutor automático.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- ABUCZKI, A.; GHAZALEH, E.B. **An overview of multimodal corpora, annotation tools and schemes.** In: *Argumentum* 9, p. 86-98. Debreceni Egyetemi Kiadó, 2013.
- BALDRY, A.; THUBAULT, P. **Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook.** London; Oakville, CT: Equinox, 2006.
- BATEMAN, J. A. **Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents.** New York: Palgrave MacMillan, 2008.
- BATEMAN, J. A. **Text and Image: A Critical Introduction to the Visual/Verbal Divide.** New York: Routledge, 2014.
- BERBER SARDINHA, T. **Linguística de Corpus.** São Paulo, Editora Manole Ltda. Barueri, 2004.
- CHEN, Y.; ADOLPHS, S. **Multimodal Corpus-based Investigations of the Gesture-Speech Relationship.** Conference: The 9th International Corpus Linguistics Conference. University of Birmingham, Birmingham, 2017.
- DUNCAN, SUSAN. **Multimodal annotation tools.** In book: *Body – Language – Communication*, Chapter: 66. Publisher: de Gruyter, Editors: MUeller, Cienki, Fricke, Ladewig, McNeill, Tellendorf, pp.1015-1022, 2013.
- FOSTER, M.; OBERLANDER, J. **Corpus-based generation of head and eyebrow motion for an embodied conversational agent.** In: *Language Resources and Evaluation*. V. 3, chapter 41, pp 305-323. December, 2008.
- HIIPPALA, T. **Modelling the structure of a multimodal artefact.** Thesis—Helsinki: University of Helsinki, 2013.
- JEWITT, C. **The Routledge Handbook of Multimodal Analysis.** London: Routledge, 2009.
- KNIGHT, Dawn. **Multimodality and Active Listenership.** *Corpus and Discourse*, Continuum, 1st edition, 2011a.
- KNIGHT, Dawn. **The future of multimodal corpora.** *The University of Nottingham, UK. RBLA*, v. 11, n. 2, p. 391-415, 2011b.
- KNIGHT, D.; ADOLPHS, S. **The Nottingham Multi-Modal Corpus: a demonstration.** In: 6th Language Resources and Evaluation Conference (LREC), Marrakesh, Morocco, 28-30 May 2008.

KRESS, G.; VAN LEEUWEN, T. **Reading Images: The Grammar of Visual Design**. 2nd. ed. London: Routledge, 2006.

LEMKE, J.L. **Multiplying Meaning: visual and verbal semiotics in Scientific text**. In: J. R. Martin e R. Vell. Eds.. Reading Science. London: Routledge, p. 87-113, 1998.

LIU, Y.; O'HALLORAN, K. L. **Intersemiotic Texture: analyzing cohesive devices between language and images**. Social Semiotics, v. 19, n. 4, p. 367–388, 2009.

MAGALHÃES, C.; LEITÃO, A.; FERNANDES DE ASSIS, D. **A função interpessoal e o posicionamento dialógico de tradutores em traduções de picturebooks**. Em: Domínios de Linguagem, vol. 11, n. 5. Uberlândia, 2017.

O'DONNELL, M. **Demonstration of the UAM CorpusTool for text and image annotation**. Proceedings of the ACL-08: HLT Demo Session (Companion Volume), pages 13-16, 2008.

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**. Tese de doutorado—Brasília: UnB, 2017.

ROHLFING, K.; LOEHR, D.; DUNCAN, S., et al. **Comparison of multimodal annotation tools: Workshop report**. *Gesprächsforschung*. 2006;7.

VIEIRA, J.; SILVESTRE, C. **Introdução à multimodalidade: contribuições da gramática sistêmico-funcional, análise de discurso crítica, semiótica social**. Brasília, 2015.

VILAR, D. et al. **Error analysis of statistical machine translation output**. In: PROCEEDINGS OF LREC. Genoa: 2006.

ANEXO A

	Media & Text Editors	Anvil	ELAN
Home page	N/A	http://www.dfki.de/~kipp/anvil/	http://www.mpi.nl/tools/elan.html
Version used at the workshop		4.5	2.4.1
Main idea of tool	Video is viewed with video editor; transcriptions typed in text editor with annotation conventions describing non-verbal behavior	Synchronized musicalscore annotation ("partitur") and video	Synchronized multimodal annotation of audio and/or video data
Technical description			
Supported platforms	N/A	Theoretically runs on any Java compliant machine, but only known to run well on Windows. Known not to run well on Mac. Said to run under Linux	Java-based, ships on Win, Mac, Linux. Runs well on Windows and Mac. Media playback performance on Mac is slightly less but still satisfactory. Video playback performance is problematic on Linux
Supported video formats/codecs	Any used by video editor of choice	Any supported by Java Media Framework v 2.1.1	Depends on the media framework of the underlying operating system. On Windows any format that is supported by Windows Media Player (DirectX) or QuickTime, on Mac OS any format supported by QuickTime, on Linux any format supported by Java Media Framework 2.1.1
Practical limit on file sizes or durations?	No	Probably not suited for material over 30 minutes	No (at least not known)

Import from:	N/A	Praat	Showbox, CHAT and Transcriber
Export annotations?	N/A	Besides txt, Anvil supports export as tables (separated with tabs e.g.) for statistic software like SPSS and Statistica. This feature is in the Project Tool and a valuable analysis option	Yes: tab-delimited text, "interlinear" text, Shoebox, CHAT, SMIL, QuickTime text track
Annotations saved in XML?	N/A	Yes	Yes
Future support for software?	Always	Will be maintained for some time to come...	Financial support for new releases for at least several years. Firm support by the institute
Open source software?	N/A	No	Yes
Cost	Cost of media and text editors	Free	Free
Control modes			
Can view time-aligned annotations with video?	No	Yes	Yes
Handles multiple videos?	Yes	No	Yes, up to 4 videos
Can play video at variable speeds?	Yes	Yes	Yes

Can move fram-by-frame	Yes	Yes	Yes. Can move step-by-step, with a minimum step size of 1 milisecond
Can hear audio at slow motion?	Yes (and at frame-by-frame speed)	No. But if audio annotations are imported from Praat, less of an issue	Yes
Search function	Visual and word-processor searches over annotated text	Can search for annotations across multiple files	Complex searches based on temporal and/or structural relations in a single file and simple text search across multiple files
Analysis functions?			No. Export to tab-demilited text file that can be imported into Excel and similar applications.
Can code non-verbal behavior in absence of speech?	Yes, although temporal extent must be additionally annotated somehow	Yes	Yes
User-defined annotation types?	Yes	Yes. Done in XML editor outside tool, requires some learning curve	Yes. User can define multiple annotation types based on 4 predefined "stereotypes", within the tool itself and while working on a file

Can change annotation definitions after data is annotated?	Yes	You CAN add tracks even for already started annotations. You can change the specification file of the currently loaded annotation and then, reload the file and see if it works. You can also add tags for "ValueSet" attributes. Anvil was developed with the idea of "incremental annotation scheme development" in mind!	New annotation types can always be added. Limited support for changing types of existing annotations
Can annotations be defined hierarchically?	Yes	Yes	Yes
Usability (end-evaluation discussed at the workshop)			
User-base size	Still quite large	The users on the web list are only a small part of the "real" user base. The mailing list comprises 792 members from 40 different countries and from 318 research institutions (for some institutes you have many users, e.g., MIT, U Edinburgh, U Tokyo, U Chicago, MITRE, and so on)	Unknown; registration is not mandatory. At least several hundreds on mailing lists or otherwise known
Positive features discussed at workshop	Ease of learning/training, ease of use, widespread availability, no limit on amount (min./hrs.) of data that can be processed and analyzed. Robust: system crashes, upgrade incompatibility, inadequate user support, etc. are not issues	See tool description in Rohlfing et al. (2006)	Multiple views on the data, multiple undo/redo, use of templates, support for multiple character sets, versatile search options
Other positive features	Not constrained to tabular report formats: annotated transcripts are useful "visualizations" in support of certain analyses; utility for analysis of extended connected discourse sequences	Displays waveform, pitch, and intensity of speech signal.	Displays waveform, up to 4 videos, print function, preliminary 2D video annotation, several automated annotation options

Negative features discussed at workshop	Cumbersome database maintenance and query	See tool description in Rohlfing et al. (2006)	No built-in help system, more export formats could be added, certain inflexibility of tier structure
Other negative features	No support for inclusion of instrumental analyses of audio or video data, or instrumental measures of nonverbal behavior		
	EXMERaLDA	TASX	MacViSTA
Home page	http://www.rrz.uni-hamburg.de/exmeralda/index-em.html	http://medien.informatik.fhfulda.de/tasxforce/TASX-annotator	http://vislab.cs.vt.edu/~rtr/
Version used at the workshop	1.3.2		
Main idea of tool	Designed for verbal transcription with non-verbal (e.g., video) seen as an additional "comment". Interface may be less intuitive if verbal and non-verbal behavior are equally important or if the latter is paramount, as timeline maps transcription typography, not absolute duration	Synchronized musical score annotation ("partitur) and video	Synchronized musical score annotation ("partitur) and video
Technical description			
Supported platforms	Java-based, known to work well on Windows, Linux, and MAC OS 10.x. Sometimes problems with certain media formats on MAC	According to TASX web site, it is Java-based, written and tested on Win98. Also tested (but some features do not work as well) on Win2000, NT, Solaris, Linux, and Mac OS9	Mac OS X 10.2.8 or later. Mac G5 recommended

Supported video formats/ codecs	Any supported by Java Media Framework v 2.1.1	Any format that is supported by Java Media Framework v 2.1.1	Quicktime
Practical limit on file sizes or durations?	No	Tested with video files up to 4 hours. The number of visible elements on a tier is the main limiting factor	It plays 320MB movie file (40min) with no problem; textgrid from Praat slows down the movie and it drops frames
Import from:	Praat, TSX, ELAN, TEI, Simple EXMERaLDA (text file). Also: syncWriter and HIAT-DOS (both	Praat, Anvil, Transcriber and ESPS Waves+ (using built-in XSLT Processor)	Praat
Export annotations?	Yes: TASX, ELAN, Praat, AIF, plain text segments lists, CHAT, TEI, built-in support for "free" conversion with XSL-Stylesheets	HTML, Praat, ESPS Waves+, STM (export via integrated XSLT engine)	Yes, though CocoaMySQL
Annotations saved in XML?	Yes	Binary (more efficient loading) and XML	Yes
Future support for software?	At least three more years of funding ongoing development, support through e-mail and website	Follow up software Eclipse Annotator is launched at LREC 2006	Uncertain
Open source software?	No, but I am willing to share code with whoever is interested	Yes	
Cost	Free	Free	Free

Control modes			
Can view time-aligned annotations with video?	Yes	Yes	Yes
Handles multiple videos?	No	Yes	Yes, but if they are "genlocked", i.e., video synchronization must be done first, outside of MacVista
Can play video at variable speeds?	No	Yes	Yes
Can move fram-by-frame	Yes, if the video code supports this	Yes (if video format permits)	Yes
Can hear audio at slow motion?	No. But if audio annotations are imported from Praat, less of a issue	No	Yes, in video primary mode if NOT played frame by frame. Audio is available in frame by frame only in audio primary mode
Search function	Simple search functionality in the Partitur-Editor itself, more elaborate search functionality in a separate tool (ZECKE, see website)	Rudimentary	No
Analysis functions?	Word, utterance etc. segmentation, analysis through XSL-Stylesheets	Yes, using external programs (e.g., Praat)	No

Can code non-verbal behavior in absence of speech?	Yes, but coding only non-verbal behaviour (i.e., without at least sporadic verbal behaviour) will not be very intuitive	Yes	Yes
User-defined annotation types?	User defined annotation categories (i.e., names for pre-defined types)	Yes. Done inside tool.	Yes. Done inside tool.
Can change annotation definitions after data is annotated?	Can easily add new definitions, tiers or speakers	Yes	Yes, color categorization for tags can be easily reassigned. New tiers can also be added later
Can annotations be defined hierarchically?	No	Not supported (though TASX format allows to)	No
Usability (end-evaluation discussed at the workshop)			
User-base size	Several hundred users in Germany (mostly discourse/conversation analysis and language acquisition research), some users outside Germany	Small (unfortunately)	Regularly used by 1, test-used by 3 in McNeill lab

Positive features discussed at workshop	See tool description in Rohlfig et al. (2006)	See tool description in Rohlfig et al. (2006)	See tool description in Rohlfig et al. (2006)
Other positive features	Supports several transcription systems (HIAT, GAT, DIDA, CHAT, IPA) through a number of parameterised functions		
Negative features discussed at workshop	See tool description in Rohlfig et al. (2006)	See tool description in Rohlfig et al. (2006)	See tool description in Rohlfig et al. (2006)
Other negative features		User interface is crummy, hard to learn. Despite the available (powerful) extensability features, only very few people committed to the software. Only load a single file, not suited for working with a complete corpus	Under development - many bugs still being fixed

Fonte: Adaptado de Rohlfig et al., 2006, p. 118-121.