

Corpus de legendas de *animes* (CorLeAni)

Anime subtitles corpus (CorLeAni)

Daniel Gomes da Cunha  *^{1,2} e **Janailton Mick Vitor da Silva**  †³

¹Centro de Ensino Unificado de Brasília, Brasília, DF, Brasil.

²Faculdade de Tecnologia e Ciências Sociais Aplicadas, Brasília, DF, Brasil.

³Instituto Federal de Brasília - Campus Ceilândia, Brasília, DF, Brasil.

Resumo

Reportamos, neste artigo, o processo de compilação de um *corpus* formado por legendas de *animes* em português brasileiro, aqui denominado Corpus de Legendas de Animes (CorLeAni). O CorLeAni, atualmente com cerca de 1 milhão de palavras, foi compilado seguindo o suporte teórico-metodológico dos Estudos da Tradução Baseados em Corpus. A compilação deu-se ao longo de dois Projetos de Pesquisa de Iniciação Científica de Ensino Médio (PIBIC-EM), 2019-2020 e 2020-2021, com bolsa do Instituto Federal Goiano - Campus Campos Belos. Reconhecendo que as *fansubs* apresentam um grande potencial para discussões e estudos no âmbito audiovisual, disponibilizamos o CorLeAni de forma gratuita, podendo ser utilizado em futuras pesquisas com focos linguístico e tradutório.

Palavras-chave: *Anime.* *Corpus.* *Fansubber.* *Fansubs.* *Legendagem.*

Abstract

We report, in this article, the process of compiling a corpus consisting of *anime* subtitles in Brazilian Portuguese, hereby named Anime Subtitles Corpus (CorLeAni). The corpus, currently with almost 1 million words, has been compiled following the theoretical-methodological support of Corpus-Based Translation Studies. The compilation took place over two Projects of Scientific Initiation of High School (PIBIC-EM), 2019-2020 and 2020-2021, with a grant from the Instituto Federal Goiano-Campus Campos Belos. Recognizing that fansubs have great potential for discussions and studies in the audiovisual field, we have made CorLeAni available for free so it can be used in future linguistic and translation research.

Keywords: *Anime.* *Corpus.* *Fansubber.* *Fansubs.* *Subtitling.*

1 Introdução

Em Díaz Cintas (2013), o autor destaca que vivemos na era da audiovisualização, onde testemunhamos várias manifestações tradutórias em mídias diversas, como a legendagem, dublagem, *voice-over*, narração, audiodescrição, entre outras. Em prosseguimento as suas reflexões, Díaz Cintas e Neves (2015) reconhecem, ademais, a Tradução Audiovisual, área que abrange essas e outras manifestações tradutórias, como uma das áreas que mais crescem dentro dos Estudos da Tradução, principalmente por se associar com tecnologias que estão em rápida e constante evolução.

Nesse contexto da Tradução Audiovisual, sobretudo da legendagem, compreendemos as legendas como sendo textos escritos, advindos de textos orais, de sons e textos multissemióticos, bem como imagens, símbolos, inscrições imagéticas e outros (DÍAZ-CINTAS; REMAEL, 2007). Partindo dessa compreensão, neste artigo, reportamos o processo de compilação de um *corpus* utilizando legendas produzidas para *animes*, aqui denominado Corpus de Legendas de Animes (CorLeAni). Tendo em vista a necessidade de ampliação de análises linguísticas e de fenômenos tradutórios dentro dos Estudos da Tradução Baseados em Corpus (ETBC), desenvolvemos dois Projetos de Pesquisa de Iniciação Científica de Ensino Médio (PIBIC-EM), 2019-2020 e 2020-2021, com bolsa do Instituto Federal Goiano – Campus Campos Belos. Sendo desenvolvidas pelo primeiro autor e orientadas pelo segundo autor deste artigo, a primeira pesquisa (2019-2020) objetivou o início da compilação do *corpus*, enquanto a segunda (2020-2021) buscou a expansão do CorLeAni já criado anteriormente, a partir da inserção de novos textos e a consolidação da metodologia empregada nesse processo. Em ambos



Linguagem e Tecnologia
DOI: 10.35699/1983-3652.2022.38695

Seção:
Artigos

Autor Correspondente:
Daniel Gomes da Cunha

Editor de seção:
Daniervelin Pereira
Editor de layout:
Leonardo Araújo

Recebido em:
9 de março de 2022
Aceito em:
4 de abril de 2022
Publicado em:
27 de maio de 2022

Essa obra tem a licença
“CC BY 4.0”.



*Email: dng20@hotmail.com

†Email: janailtonm@gmail.com

os casos, fomos motivados particularmente pela crescente demanda para se entender o processo de criação de legendas de *animes*, bem como o modo no qual a língua portuguesa e as outras linguagens são empregadas nessas legendas. Para tanto, foi feita uma compilação criteriosa do *corpus*, conforme discutiremos mais adiante. Ademais, buscamos disponibilizar¹ o CorLeAni em sua versão atual para a comunidade acadêmica, a fim de que possa ser utilizado em futuras pesquisas e estudos voltados ao uso da língua portuguesa traduzida no contexto dos *animes*.

Feita a exposição dos percursos realizados pela pesquisa, para este artigo trazemos os resultados de ambos os projetos PIBIC-EM. Em primeiro lugar, apresentaremos os conceitos teóricos que iluminaram a compreensão do tema, seguido do passo a passo metodológico empreendido na compilação do CorLeAni. Por fim, lançaremos mão dos resultados gerais quantitativos obtidos por meio da inserção dos arquivos de legendas no *WordSmith Tools*®, versão 7. No processo de compilação do *corpus*, salientamos que utilizamos, além de programas pagos, alguns programas gratuitos, como *Aegisub* e *Subtitle Edit*, conforme apresentamos na Seção 3 deste texto.

2 Fundamentando as pesquisas: conceitos sobre legendagem, *fansubs*, *animes* e *corpus*

Díaz-Cintas e Remael (2007) entendem a legendagem como um conceito que pode ser entendido como uma prática tradutória, em que é apresentado um texto escrito geralmente na parte inferior da tela e visa relatar um diálogo original entre locutores. Além disso, é possível traduzir, nessas legendas, elementos discursivos que aparecem na imagem, tais como letras, inscrições, cartazes e informações presentes na trilha sonora, como é o caso de músicas e *voices off*.

Nos dias de hoje, dispomos das chamadas *fansubs*, que são as legendas produzidas por fãs e que também são conhecidas pelo termo ‘legendagem amadora’ (BOGUCKI, 2009). Díaz-Cintas e Sánchez (2006, p. 1) entendem as *fansubs* como uma versão legendada e traduzida por fãs de um *anime* japonês, e defendem que “não seria exagero afirmar que as *fansubs* são a manifestação mais importante da tradução de legendas feita por fãs”². Trata-se de uma prática tradutória que tem o intuito de fornecer materiais legendados gratuitamente e difundi-los entre públicos de diferentes comunidades linguísticas (BOGUCKI, 2009).

De acordo com Ferrer Simó (2005), há, nas *fansubs*, algumas características que as tornam específicas, tal como as notas e os comentários do tradutor. São ferramentas que geralmente explicam algumas tradições, celebrações e lugares. Além disso, há o uso de fontes de texto para diferenciar os atores e o uso de legendas karaokê nas músicas de abertura e encerramento, bem como outras características.

As *fansubs* foram originalmente pensadas para produção de legendas inclusive como forma de divulgar a cultura *pop* japonesa (DÍAZ-CINTAS; SÁNCHEZ, 2006). É importante chamar a atenção para o termo ‘cultura *pop* japonesa’, que engloba mensagens e imagens criadas no Japão, servindo como poderosos assimiladores de sentidos daquele contexto cultural. Atualmente, aspectos da cultura *pop* japonesa se encontram espalhados pelo mundo a partir das novas tecnologias de informação e novas mídias (SATO, 2005 apud PEREIRA, 2017). Destacam-se a sua presença a partir dos mangás (histórias em quadrinhos japoneses), dos *animes* e dos demais produtos comerciais que propagam tais características culturais como parte de si.

O caso do Brasil traz especificidades que chamam a atenção. Isso porque “[n]o Brasil, no entanto, muito antes d[e o] mundo ‘descobrir’ os mangás, estes já eram fartamente lidos pela comunidade dos descendentes de japoneses.” (LUYTEN, 2014, p. 8). Por se preocuparem com a manutenção do idioma e com o aprendizado de novos termos, os japoneses realizavam importações de mangás (e, posteriormente, de *animes* e filmes), para que eles e seus filhos continuassem tendo contato com a cultura. Assume-se, portanto, que tivemos uma grande contribuição das comunidades nipônicas na

¹ Conforme site a seguir: <http://corleani.epizy.com>

² “It would be no exaggeration to state that *fansubs* are nowadays the most important manifestation of fan translation [...]” (DÍAZ-CINTAS; SÁNCHEZ, 2006, p. 1, tradução nossa). Neste artigo, optamos por realizar nossa própria tradução de trechos dos artigos em inglês para o português brasileiro. Especialmente para esse texto de Díaz-Cintas e Sánchez (2006), ressaltamos a publicação recente de uma versão traduzida para o português brasileiro, conforme Díaz Cintas, Muñoz Sánchez e Moura (2022).

difusão dessa cultura.

O *pop* japonês se misturou com outros elementos de diversas origens, e, apesar de atrair um público muito específico, já não é mais algo tão estranho à cultura urbana brasileira (NETO, 2017). Para demonstrar o quanto incluído o *pop* japonês está na cultura urbana brasileira, Neto (2017) afirma que ocorre, por todo o Brasil, uma quantidade relevante de eventos de *animes*, também chamados de ‘animencontros’. O autor ressalta que esses eventos não se limitam aos *animes*, abordando outros elementos da cultura *pop*, oriundos inclusive de outras origens geográficas. De todo modo, quem normalmente frequenta esses ambientes são pessoas que compartilham um gosto por *animes* e por mangás.

Com os interesses de entender não somente a legendagem, mas também as legendas de *animes* que resultam desse processo, assumimos o conceito de corpus enquanto uma coleção de textos em formato eletrônico e que são passíveis de passarem por análises automáticas ou semiautomáticas (BAKER, 1996). Tais textos podem servir como pano de fundo para serem realizadas avaliações sobre uma determinada língua, com o auxílio da Linguística de Corpus (LC). Respaldando-nos em McEnergy e Wilson (2001), sublinhamos que a LC é a área da linguística que se ocupa em estudar o lado empírico da língua, ou seja, aquilo que é, de fato, produzido por falantes, a partir da observação de dados naturais que estão contidos em um *corpus*. Berber Sardinha (2004) orienta que a compilação de um *corpus* precisa seguir alguns pré-requisitos básicos, como, por exemplo, a questão da autenticidade, escolha criteriosa e representatividade dos textos. Desse modo, a junção dos conhecimentos da LC, dos ETBC e da tradução audiovisual para o estudo dos *animes* foi extremamente frutífera para o trabalho realizado durante os dois projetos de iniciação científica, conforme prosseguimos mais detalhadamente no tópico abaixo.

3 Passos metodológicos para criação do CorLeAni

Os pressupostos metodológicos que orientaram o desenvolvimento das pesquisas PIBIC-EM 2019-2020 e 2020-2021 seguiram o paradigma interpretativista, a metodologia descritiva e a tipologia quanti-qualitativa (MOREIRA; CALEFFE, 2008). Além disso, afiliamo-nos aos conhecimentos teórico-metodológicos dos ETBC (BAKER, 1995, 1996; CAMARGO, 2007), especialmente por meio da utilização das ferramentas *WordList* e *Concord* do programa *WordSmith Tools*®, versão 7.0, para compilar o CorLeAni com legendas, em língua portuguesa traduzida, feitas por fãs brasileiros (*fansubbers*).

3.1 Seleção dos *animes* e obtenção das legendas *fansubs*

Com base no que foi desenvolvido na pesquisa PIBIC-EM 2019-2020, decidimos manter e empregar, como critério de seleção dos *animes* em ambos os projetos, as 25 maiores notas de avaliações, que variam em um espectro que vai de 0 a 10. As notas foram dadas por fãs aos *animes* da temporada de primavera de 2020 e retiradas no site *MyAnimeList*³. Em sequência, utilizamos as informações do site *InfoAnime*⁴ para buscar os grupos de *fansubbers* que haviam legendado os *animes* que foram criteriosamente selecionados. Nos dois sites apropriamo-nos dos seguintes dados: a avaliação dos fãs, no site *MyAnimeList*; os nomes dos grupos *fansubbers*, no site *InfoAnime*; e também a quantidade de episódios legendados pelos *fansubbers*. A partir dos dados coletados, conseguimos elaborar a Tabela 1 e a Tabela 2, buscando relacionar algumas dessas informações sobre os *animes* que compõem o CorLeAni.

Compartilhamos aqui, ademais, os sites oficiais⁵ em que se encontravam divulgados os *animes* legendados para o público, com ênfase nos *animes* cujos arquivos de mídia poderiam ser baixados. Outro caso que nos possibilitou o download envolveu dois grupos de *fansubbers*, *Seoltang*⁶ e *Illumi-*

³ Endereço do site: <https://myanimelist.net/>. Acesso em: 14 Abr. 2022.

⁴ Endereço do site: <https://www.infoanime.com.br/>. Acesso em: 14 Abr. 2022.

⁵ Aenianos: <https://www.aenianos.org/>. Acesso em: 20 fev. 2020. ECN: <http://ecn.fansubs.com.br/>. Acesso em: 20 fev. 2020. Eternal Animes: <https://eternalanimes.org/>. Acesso em: 20 fev. 2020. Infinite: <https://infinitefansub.com/>. Acesso em: 20 fev. 2020. Kyoshiro fansub: <https://www.kyoshirofansub.com/>. Acesso em: 20 fev. 2020. Kirinashi: <https://kirinashi.fansubs.com.br/>. Acesso em: 20 fev. 2020.

⁶ Endereço do site: <https://seoltangs.blogspot.com/>. Acesso em: 12 fev. 2021.

Tabela 1. Tabela de *animes* que compõem o CorLeAni (PIBIC-EM 2019-2020).

Classificação	Pontuação	Animes	Episódios	Fansubbers
1.	9.08	Shingeki no Kyojin parte 2 temp 3	10	Kyoshiro fansub
2.	8.95	Kimetsu no Yaiba	26	Aenianos
3.	8.33	Fruits Basket	25	Aenianos
4.	8.08	Carole & Tuesday	24	Infinite
5.	7.79	Kono Oto Tomare!	13	Infinite/Punch
6.	7.68	Sarazanmai	11	Aenianos
7.	7.61	Isekai Quartet	12	Eternal Animes
8.	7.52	One Punch Man	12	Eternal Animes
9.	7.44	Sewayaki Kitsune no Senko-san	12	ECN
10.	7.42	Bokutachi wa Benkyou ga Dekinai	13	Infinite
11.	6.96	Midara na Ao-chan wa Benkyou ga Dekinai	12	Infinite
12.	6.76	Kenza no Mago	12	Eternal Animes

Fonte: Elaboração nossa.

Tabela 2. Tabela de *animes* que compõem o CorLeAni (PIBIC-EM 2020-2021).

Classificação	Pontuação	Nome	Episódios	Fansubbers
1.	8.80	Kaguya-sama wa Kokurasetai?: Tensai-tachi no Renai Zunousen	12	Boteco/Fênix Sub
2.	8.19	Honzuki no Gekokujou: Shisho ni Naru Tame ni wa Shudan wo...	12	Boteco
3.	8.04	Kakushigoto	12	Moshi Moshi
4.	7.77	Kami no Tou	13	Kyoshiro/TDM
5.	7.63	Fugou Keiji: Balance:Unlimited (Detective milionario-traduzido)	11	Seoltang
6.	7.52	Otome Game no Hametsu Flag shika Nai Akuyaku Reijou ni...	12	Kawaii Otome
7.	7.37	Shokugeki no Souma: Gou no Sara	13	Boteco
8.	7.36	Appare-Ranman!	13	Elite Fansub
9.	7.09	Princess Connect! Re:Dive	13	Illuminati/Shakai
10.	6.41	Komatta Jiisan	13	Isekai

Fonte: Elaboração nossa.

*nati*⁷. Eles oferecem a possibilidade de fazer o download das legendas do *anime* em si, em vez de somente disponibilizarem um *link* para baixar o episódio.

Verificamos ainda a possibilidade de extrair as legendas disponibilizadas nos episódios utilizando os programas de legendagem *Subtitle Edit*⁸ (versão 3.5.10) e/ou *Aegisub*⁹ (versão 3.2.2). Observamos, logo de início, a existência de *animes* que foram traduzidos pelos *fansubbers*, mesmo estando eles veiculados oficialmente na *Crunchyroll*¹⁰.

Como descrito nos tópicos abaixo, realizamos duas ações que nos ajudaram a obter o nosso material para compilação do *corpus*.

3.1.1 Contato com o grupo de *fansubbers*

Como percurso para coleta de dados, fizemos levantamento prévio dos contatos dos *fansubbers*, no intuito de criar um canal de comunicação que nos permitisse solicitar as legendas a cada grupo diretamente, especialmente através do Facebook e de *e-mails*. O procedimento parecia o mais adequado no primeiro momento, pois otimizaria o tempo da pesquisa. No entanto, a abordagem não foi tão efetiva, uma vez que, ora os grupos não respondiam ao nosso contato, ora o material disponibilizado era insuficiente (falta de episódios ou a inclusão de séries que não atendiam ao nosso critério de seleção).

Dentre os cinco grupos de *fansubbers* que entraram nos critérios de seleção, conseguimos resultados satisfatórios apenas com um grupo, a *Eternal Animes*. Para esse caso, angariamos um total de 36 arquivos de legendas para três diferentes séries de *anime*. Todavia, para os outros 4 grupos, tivemos que recorrer ao que se tornou a melhor estratégia para obter os episódios: o *download* individual das obras, conforme descrito no subitem abaixo. No PIBIC-EM 2020-2021, procedemos apenas com o *download* direto das legendas, e não pelo contato com os *fansubbers*, dada a experiência adquirida com o insucesso observado no projeto anterior em relação a esse aspecto.

3.1.2 *Download* de arquivos de mídia

Utilizando-nos de meios disponibilizados nos *sites* oficiais dos *fansubbers*, realizamos o *download* dos arquivos de mídia dos *animes*. No geral, esses arquivos estavam disponibilizados no formato .MKV, no serviço de armazenamento de arquivos disponibilizado pela *Google* para seus usuários, o *Google Drive*. Em virtude de os arquivos baixados estarem no formato .MKV, realizamos um processo de extração perpassado no subitem a seguir.

3.1.3 Processo de extração de legendas

Podemos verificar, na imagem a seguir (Figura 1), a maneira na qual a legenda aparece no arquivo de vídeo, em uma cena do primeiro episódio do *anime* *Sewayaki Kitsune no Senko-san*, legendado pela *Infinite fansub*. Essa cena gira em torno do protagonista dessa comédia romântica, que retorna a sua residência após mais um dia exaustivo do trabalho em uma empresa exploradora, quando é surpreendido por um espírito de raposa divina.

Após realizarmos o *download* individual pelos meios disponibilizados pelos *fansubbers*, foi feita a extração das legendas utilizando o aplicativo *Subtitle Edit*. As imagens a seguir retratam a organização do material baixado no computador (Figura 2), o manuseio do programa *Subtitle Edit* e o passo a passo da extração das legendas nesse programa (Figuras 3 a 7). Por fim, temos o resultado do processo de extração (Figura 8). Para o caso das duas obras legendadas pela *Seoltang* e pela *Illuminati*, preferimos não realizar o processo de *download* e extração. Optamos, nesse caso, pelo *download* das legendas a partir dos *links* disponibilizados por estes grupos, o que otimizou nosso tempo.

⁷ Endereço do site: <https://illuminatisubsbr.wordpress.com/>. Acesso em: 12 fev. 2021.

⁸ Endereço oficial do site: <http://www.nikse.dk.subtitleedit>. Acesso em: 20 fev. 2020.

⁹ Endereço oficial do site: <http://www.aegisub.org/>. Acesso em: 20 fev. 2020.

¹⁰ O *Crunchyroll* consiste em um *site* de *streaming* dedicado a *animes* e à cultura asiática de um modo geral, que compra direitos de transmissão de *animes*, contribuindo para o financiamento de algumas obras, ajudando, assim, a fomentar o crescimento desse mercado. Endereço oficial do site: <https://www.crunchyroll.com/pt-br>. Acesso em: 20 fev. 2020.



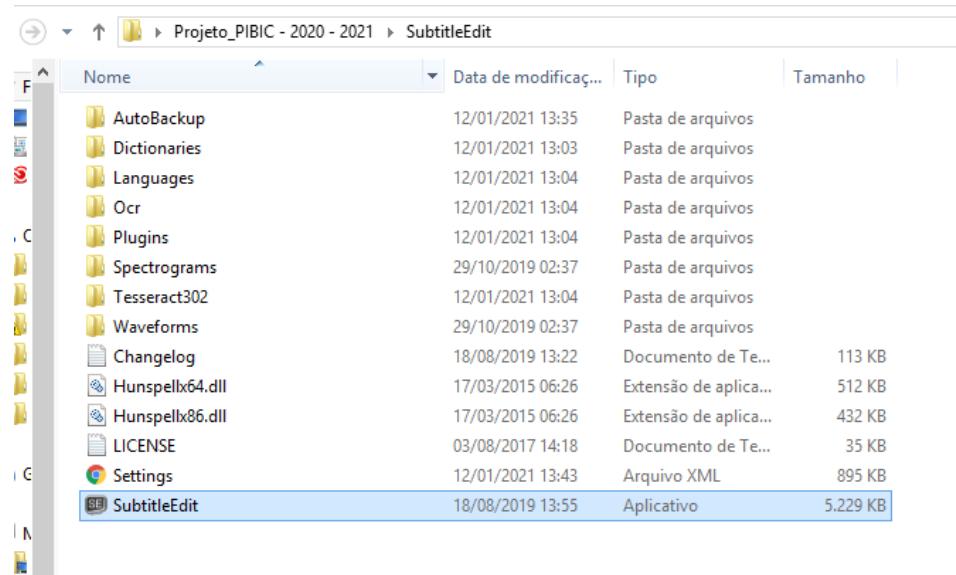
Figura 1. Cena de um anime baixado: *Sewayaki Kitsune no Senko-san*.

Fonte: Elaboração nossa.

Projeto_PIBIC - 2020 - 2021 > Obras > [Boteco] Shokugeki no Souma - Go no Sara > Vídeos				
	Nome	Data de modificação	Tipo	Tamanho
	[Boteco] Shokugeki no Souma - Go no S...	11/01/2021 21:34	Arquivo MKV	694.436 KB
	[Boteco] Shokugeki no Souma - Go no S...	17/07/2020 16:32	Arquivo MKV	730.640 KB
	[Boteco] Shokugeki no Souma - Go no S...	20/07/2020 14:10	Arquivo MKV	619.345 KB
	[Boteco] Shokugeki no Souma - Go no S...	09/08/2020 13:35	Arquivo MKV	523.027 KB
	[Boteco] Shokugeki no Souma - Go no S...	09/08/2020 13:56	Arquivo MKV	644.383 KB
	[Boteco] Shokugeki no Souma - Go no S...	10/08/2020 09:16	Arquivo MKV	665.076 KB
	[Boteco] Shokugeki no Souma - Go no S...	20/08/2020 20:07	Arquivo MKV	655.146 KB
	[Boteco] Shokugeki no Souma - Go no S...	10/09/2020 20:08	Arquivo MKV	799.926 KB
	[Boteco] Shokugeki no Souma - Go no S...	10/09/2020 20:16	Arquivo MKV	734.979 KB
	[Boteco] Shokugeki no Souma - Go no S...	10/09/2020 20:21	Arquivo MKV	737.834 KB
	[Boteco] Shokugeki no Souma - Go no S...	17/09/2020 14:11	Arquivo MKV	839.098 KB
	[Boteco] Shokugeki no Souma - Go no S...	21/09/2020 10:30	Arquivo MKV	721.622 KB
	[Boteco] Shokugeki no Souma - Go no S...	20/10/2020 22:18	Arquivo MKV	858.815 KB

Figura 2. Lista de episódios do anime: 'Shokugeki no Souma - Go no Sara' após download.

Fonte: Elaboração nossa.



	Nome	Data de modificação	Tipo	Tamanho
AutoBackup	12/01/2021 13:35	Pasta de arquivos		
Dictionaries	12/01/2021 13:03	Pasta de arquivos		
Languages	12/01/2021 13:04	Pasta de arquivos		
Ocr	12/01/2021 13:04	Pasta de arquivos		
Plugins	12/01/2021 13:04	Pasta de arquivos		
Spectrograms	29/10/2019 02:37	Pasta de arquivos		
Tesseract302	12/01/2021 13:04	Pasta de arquivos		
Waveforms	29/10/2019 02:37	Pasta de arquivos		
Changelog	18/08/2019 13:22	Documento de Te...	113 KB	
Hunspellx64.dll	17/03/2015 06:26	Extensão de aplica...	512 KB	
Hunspellx86.dll	17/03/2015 06:26	Extensão de aplica...	432 KB	
LICENSE	03/08/2017 14:18	Documento de Te...	35 KB	
Settings	12/01/2021 13:43	Arquivo XML	895 KB	
SubtitleEdit	18/08/2019 13:55	Aplicativo	5.229 KB	

Figura 3. Pasta do *Subtitle Edit* nos arquivos pessoais da pesquisa.

Fonte: Elaboração nossa.

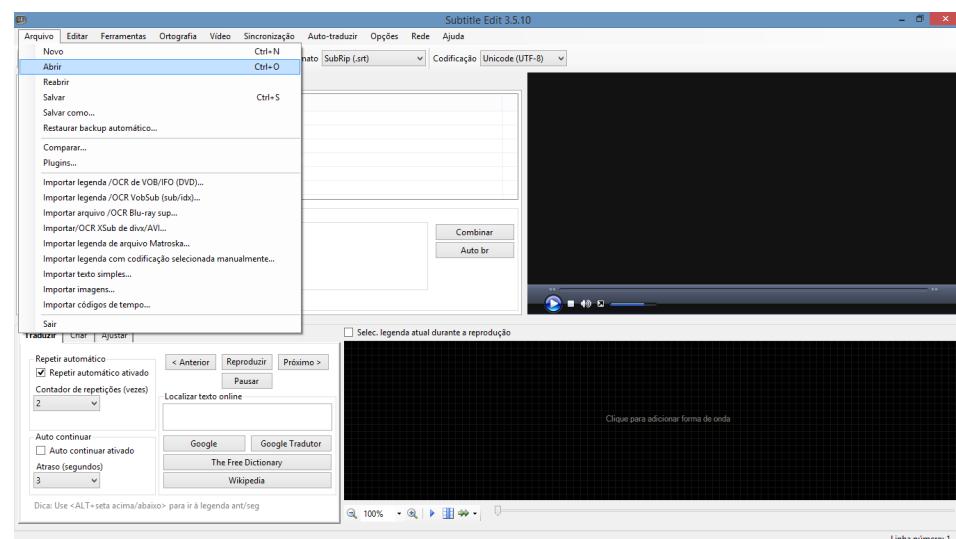


Figura 4. Interface inicial do programa *Subtitle Edit*.

Fonte: Elaboração nossa.

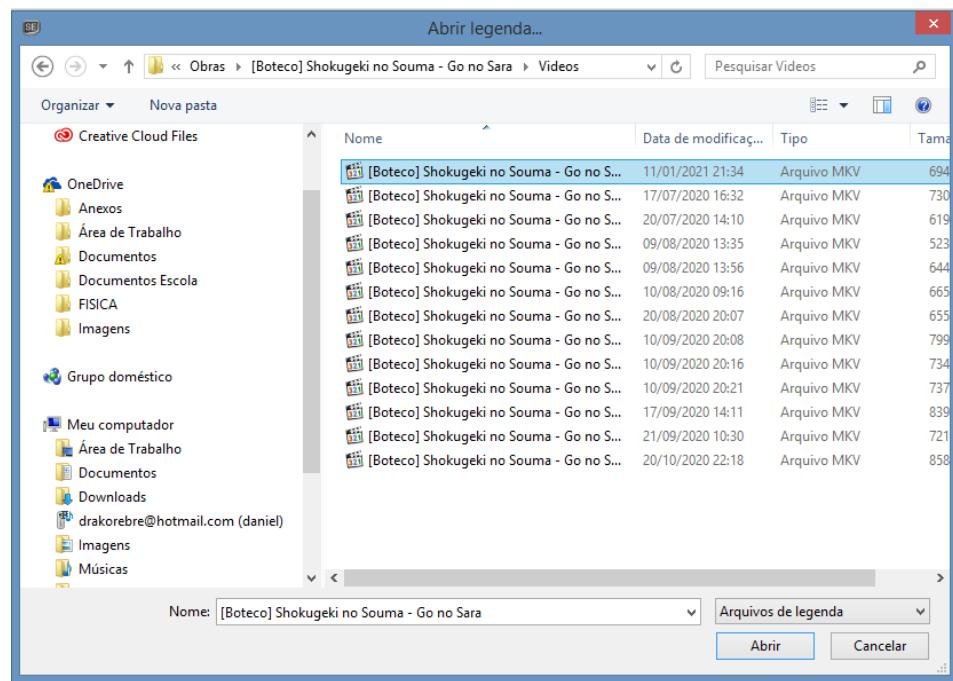


Figura 5. Selecionando um episódio dos *animes* para exemplificar a extração das legendas.
Fonte: Elaboração nossa.

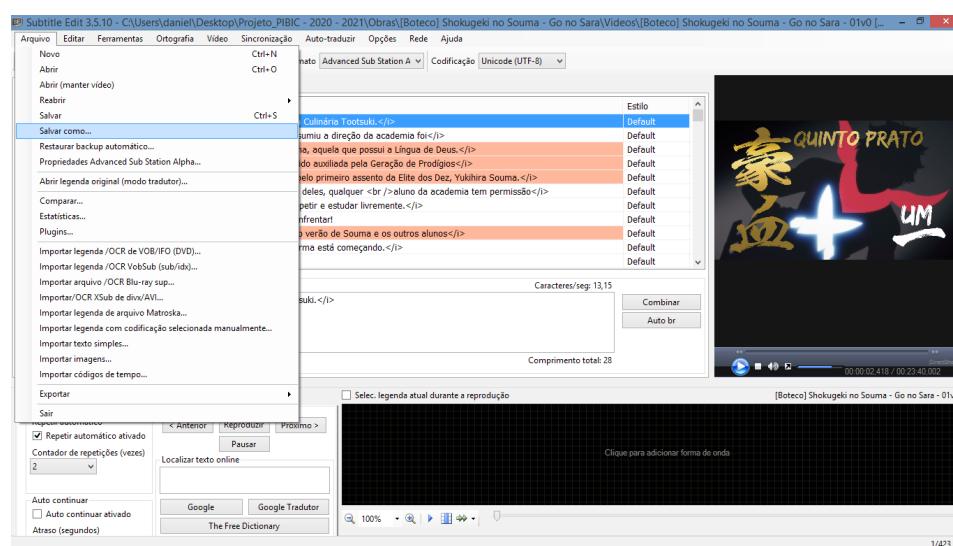


Figura 6. Realizando a extração da legenda de um episódio.
Fonte: Elaboração nossa.

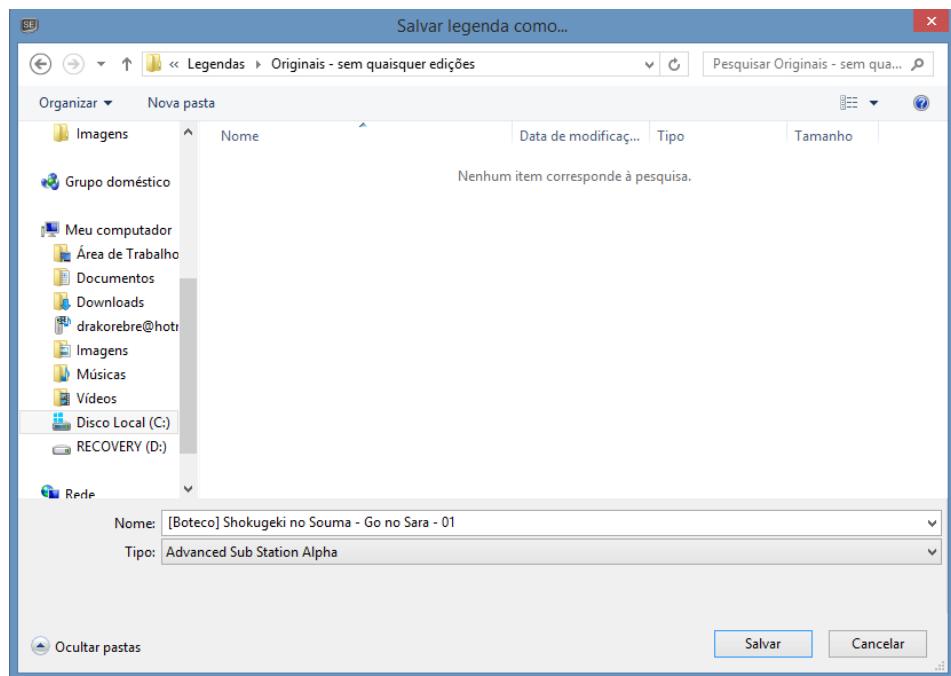


Figura 7. Salvando a legenda extraída em uma pasta específica dos arquivos da pesquisa.

Fonte: Elaboração nossa.

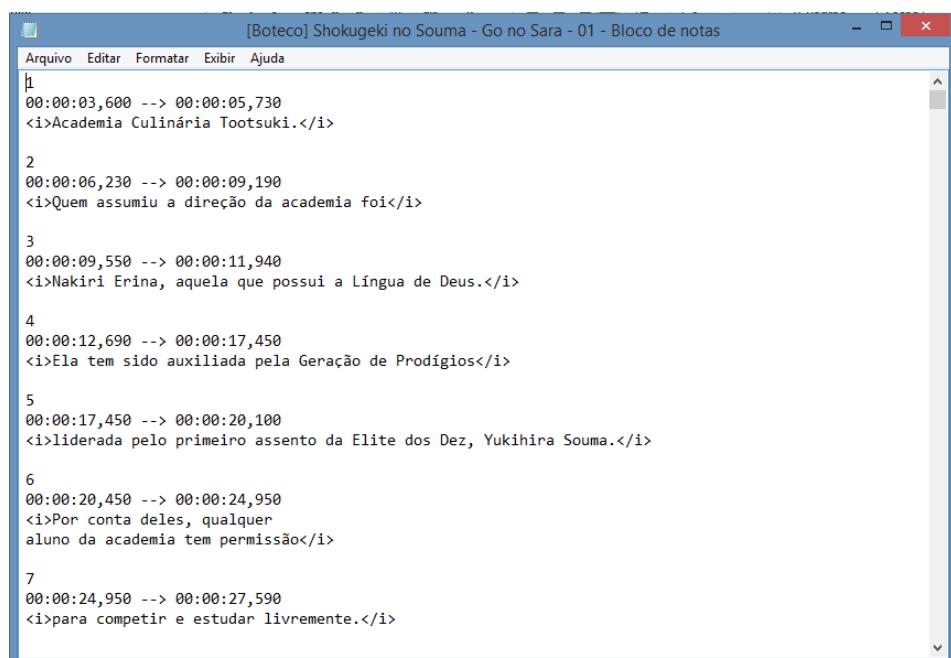


Figura 8. Trecho da legenda aberta no Bloco de Notas, após ser extraída no formato *SubRip*, para exemplificar o processo de extração.

Fonte: Elaboração nossa.

3.2 Etapa de elaboração de códigos para identificar as obras

Nesta etapa, com o material de estudo em mãos, seguimos com a nomeação dos arquivos de legendas (Figura 9). Para facilitar a identificação das obras, estabelecemos o padrão apresentado na Tabela 3.

Tabela 3. Critérios utilizados para nomear os episódios de *anime* nos arquivos das pesquisas.

Grupo	Nome da série	Temporada	Episódio
(Dois a três caracteres)	(Dois a três caracteres)	(Três caracteres)	(Três caracteres)
[MM]	YU	T01	E01
[TDM]	KnT	T01	E13

Fonte: Elaboração nossa.

[MM]YUT01E01 → Este código se refere ao grupo *fansubber* MM (Moshi Moshi), que produziu legendas do *anime* YU (*Yesterday wo Utatte*), para a temporada 01 do episódio 01 desta série.

[TDM]KnTT01E13 → Este código se refere ao grupo *fansubber* TDM, que produziu legendas do *anime* KnT (*Kami no Tou*), para a temporada 01 do episódio 13 desta série.

Nome	Data de modificação	Tipo	Tamanho
🕒 [TDM]KnTT01E01	26/04/2021 11:17	Arquivo ASS	26 KB
🕒 [TDM]KnTT01E02	26/04/2021 11:17	Arquivo ASS	32 KB
🕒 [TDM]KnTT01E03	26/04/2021 11:19	Arquivo ASS	38 KB
🕒 [TDM]KnTT01E04	26/04/2021 11:17	Arquivo ASS	30 KB
🕒 [TDM]KnTT01E05	26/04/2021 11:18	Arquivo ASS	25 KB
🕒 [TDM]KnTT01E06	26/04/2021 11:45	Arquivo ASS	40 KB
🕒 [TDM]KnTT01E07	26/04/2021 11:19	Arquivo ASS	40 KB
🕒 [TDM]KnTT01E08	26/04/2021 11:20	Arquivo ASS	33 KB
🕒 [TDM]KnTT01E09	26/04/2021 11:20	Arquivo ASS	32 KB
🕒 [TDM]KnTT01E10	26/04/2021 11:20	Arquivo ASS	32 KB
🕒 [TDM]KnTT01E11	26/04/2021 11:20	Arquivo ASS	29 KB
🕒 [TDM]KnTT01E12	26/04/2021 11:20	Arquivo ASS	23 KB
🕒 [TDM]KnTT01E13	26/04/2021 11:21	Arquivo ASS	34 KB

Figura 9. Lista dos episódios do *anime*: ‘Kami no Tou’, já nomeado.

Fonte: Elaboração nossa.

Concluída a identificação dos episódios conforme os critérios definidos, avançamos para a próxima etapa das pesquisas, apresentada na Seção 3.3. Nesta parte, estruturamos os arquivos de legendas antes de inclui-los no *WordSmith Tools*®, v. 7.

3.3 Preparação dos arquivos de legendas para inseri-las no WordList

Efetuada a nomeação dos arquivos, avaliamos e tratamos os arquivos de legendas com um olhar especial às peculiaridades de cada grupo e cada obra de *anime*, antes de ampliarmos o *corpus* em si. Decidimos, por exemplo, utilizar o programa *Aegisub*¹¹ para colocar em angulares (< >) as legendas com palavras escritas em caracteres estrangeiros, ou legendas com frases em outros idiomas, algo que geralmente aparece durante a abertura e o encerramento de um *anime*. Dessa forma, eles serão reconhecidos pelo *WordList*, ferramenta disponibilizada pelo *WordSmith Tools*®, e todos os caracteres que estão entre os angulares não servirão para contagem de palavras do *corpus*. O que fundamentou esta decisão foi principalmente a língua utilizada nos textos a serem imputados no *CorLeAni*: língua portuguesa. Em ambos os casos, de palavras escritas em caracteres estrangeiros e de legendas em outros idiomas que não o português brasileiro, são exemplificados a seguir, nas Figura 10 e Figura 11, respectivamente.

¹¹ Reppositório GitHub: <https://github.com/Aegisub/Aegisub/releases>. Acesso em: 08 jul. 2021.



Figura 10. Cena do anime *Kimetsu no Yaiba* com palavras escritas em caracteres estrangeiros ao português brasileiro.

Fonte: Elaboração nossa.



Figura 11. Cena do anime *Carole & Tuesday* com legendas em português e inglês.

Fonte: Elaboração nossa.

Especificamente no PIBIC-EM 2020-2021, essa decisão só foi aplicada ao grupo *fansubber* “Elite”, pois foi o único grupo que realizou a transcrição e tradução da abertura e do encerramento do anime intitulado: “*Appare-Ranman!*” (Figura 12).

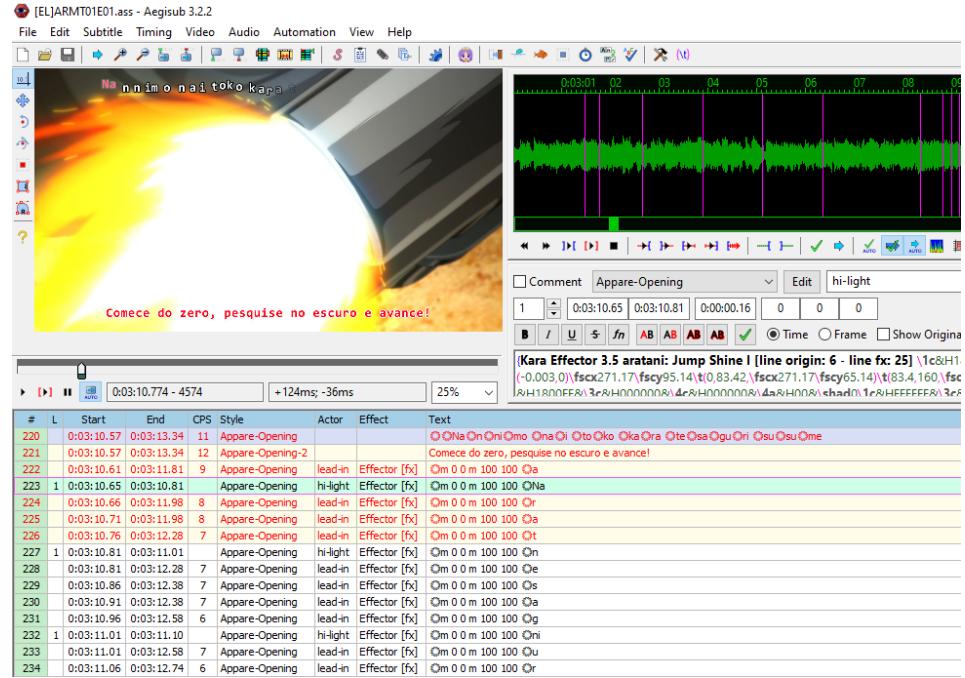


Figura 12. Transcrição da abertura do anime “*Appare-Ranman!*” em karaokê.

Fonte: Elaboração nossa.

Estabelecemos, também, um outro critério para definir a ordem em que as linhas de legendas aparecem nos arquivos. Percebemos que o padrão do software é organizar as linhas de legendas conforme elas são criadas, como é exemplificado na Tabela 5. Contudo, preferimos organizá-las pela ordem alfabética do nome dos estilos que foram preestabelecidos pelos *fansubbers*, conforme a Tabela 4.

Tabela 4. Representação de legendas na ordem alfabética dos estilos que foram preestabelecidos no arquivo de legenda.

Linha	Tempo inicial	Tempo final	Estilo	Texto
1	00:38	00:40	Legenda-Inferior	Olá.
2	00:46	00:51	Legenda-Inferior	Que isso, o prazer é todo meu.
3	00:40	00:43	Legenda-Superior	Olá! Prazer em conhecê-lo.

Fonte: Elaboração nossa.

Tabela 5. Representação de legendas na ordem em que foram inseridas no arquivo de legenda.

Linha	Tempo inicial	Tempo final	Estilo	Texto
1	00:38	00:40	Legenda-Inferior	Olá.
2	00:40	00:43	Legenda-Superior	Olá! Prazer em conhecê-lo.
3	00:46	00:51	Legenda-Inferior	Que isso, o prazer é todo meu.

Fonte: Elaboração nossa.

Dessa forma, as linhas que têm uma característica em comum ficam organizadas próximas umas

das outras, a depender do estilo que foi determinado pelo *fansubber*. A decisão foi tomada para facilitar e ajudar na análise e no processo de inserção de angulares nas legendas. É importante ressaltar que, quando olhamos no arquivo de legenda à primeira vista, as legendas podem parecer estar desorganizadas ou dessincronizadas, conforme observamos na Tabela 4. No entanto, se vincularmos esses arquivos de legendas a um arquivo de vídeo qualquer, ou ao próprio episódio do *anime*, perceberemos que elas ainda serão apresentadas para o espectador da mesma forma que o arquivo de legenda dos *fansubbers* originalmente apresentaria. O tempo, responsável por definir em que momentos as legendas devem aparecer ou sair da tela, não foi alterado, mas sim o critério que organiza as linhas de legenda no arquivo (nesse caso, o estilo das legendas).

Na Figura 13, apresentamos o programa de legendagem *Aegisub* (versão 3.2.2), que, entre tantas funções, organiza as legendas pela ordem alfabética dos estilos. Nas Figura 14 e Figura 15, mostramos a frase “Ei, você, espere aí!” antes (Figura 14, em um arquivo de legenda, sem alteração no conteúdo) e depois (Figura 15) de mudar o critério que define a ordem das legendas no arquivo.

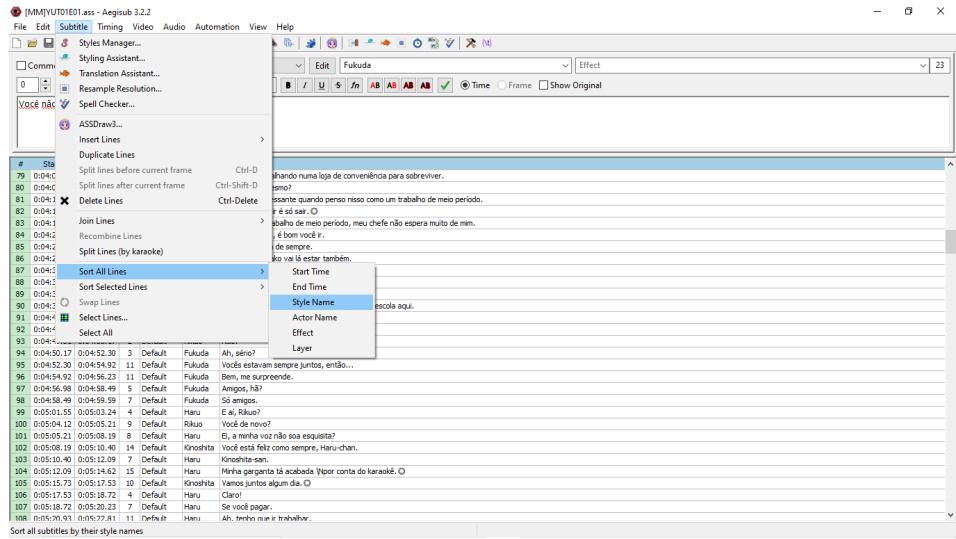


Figura 13. Organizando as legendas pelo nome dos estilos pré-definidos pelos *fansubbers*.

Fonte: Elaboração nossa.

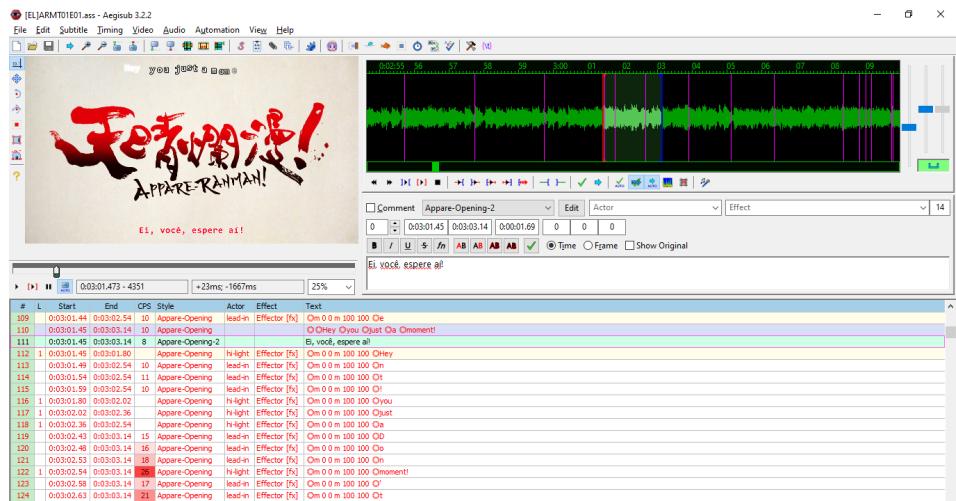


Figura 14. Arquivo de legenda após a extração, com legendas organizadas no formato padrão. A linha 110 está comentada e as outras apresentam karaokê e códigos de formatação.

Fonte: Elaboração nossa.

Observamos, a partir da análise, que mesmo quando o número da linha da frase tenha sido alterado no arquivo de legenda (Figura 15), as legendas continuam sendo exibidas para o telespectador no tempo definido pelos *fansubbers* (Figura 14). Outra coisa observada, na Figura 15, são as legendas

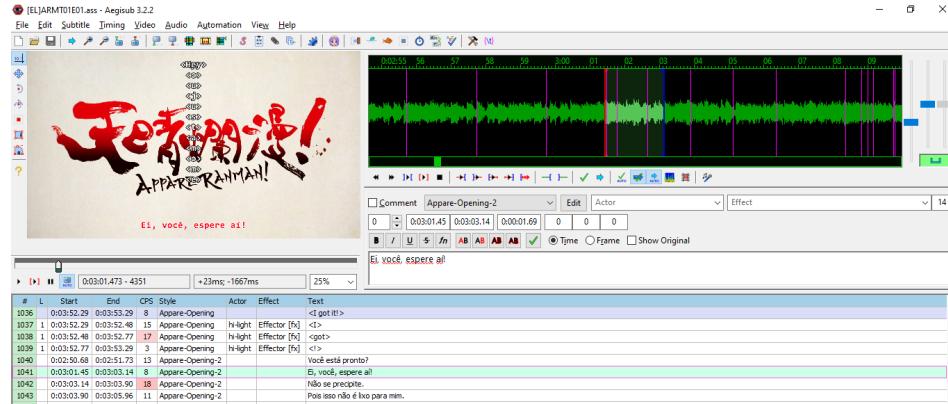


Figura 15. Legendas organizadas pela ordem alfabética dos estilos, com palavras estrangeiras e karaokê (linha 1036 a 1039) em angular.

Fonte: Elaboração nossa.

que aparecem entre os angulares (< >), códigos não assimilados pelo *Aegisub*. Analisando essas imagens, é possível inferir, ainda, que optamos por retirar os códigos de formatação dos arquivos de legenda, assunto abordado na Seção 3.3.1.

Uma outra decisão que tomamos se trata dos comentários feitos pelos *fansubbers*, característica bastante presente nesse tipo de legenda (DÍAZ-CINTAS; SÁNCHEZ, 2006). Neste caso, decidimos recorrer uma vez mais aos angulares para que essas ocorrências não sejam contabilizadas no *corpus*. Por comentários, nós entendemos legendas ou frases que não são exibidas para o telespectador, mas que servem como forma de comunicação entre os *fansubbers*. Em alguns casos, há a presença de *links de sites* (Figura 16) ou palavras escritas em caracteres estrangeiros, mas que são, em sua maioria, diálogos ou observações feitas entre ou para os próprios *fansubbers*.

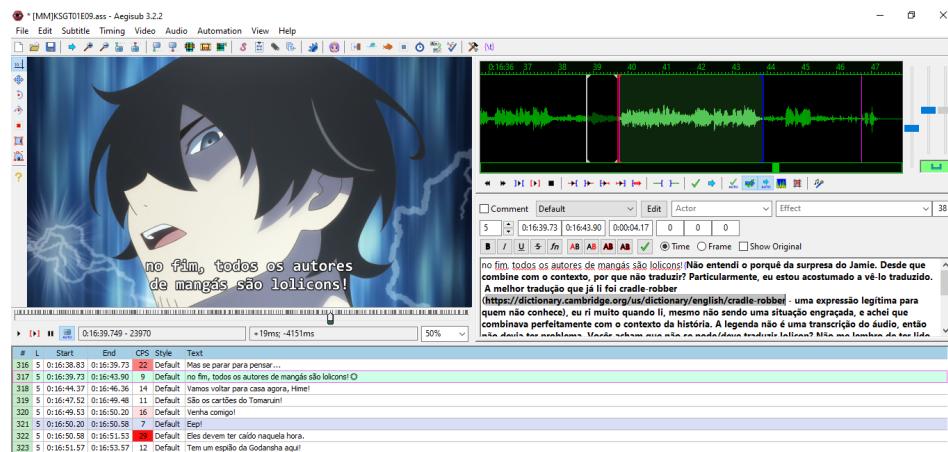


Figura 16. Exemplo de comentários (linhas 317 e 321) que não são exibidos para o espectador, ainda sem angular, que contém o *link* de um *site*.

Fonte: Elaboração nossa.

3.3.1 Exportação personalizada e uso da ferramenta conversão em lote

Como foi brevemente abordado, optamos por excluir os códigos de formatação presentes nos arquivos de legendas. No projeto PIBIC-EM 2019-2020, também tomamos essa decisão e avaliamos que, na pesquisa PIBIC-EM 2020-2021, essa escolha continuaria sendo válida, embora tenhamos encontrado uma quantidade consideravelmente menor de códigos de formatação. Contudo, ainda nos deparamos com outros códigos encontrados na primeira pesquisa, à exceção daqueles que eram, na verdade, palavras. Preferimos, portanto, utilizar um formato de texto personalizado, pois facilitaria a padronização dos cabeçalhos e tornaria mais prático o processo de colocar entre angulares o número das linhas e os tempos de início e fim de uma linha de legenda.

A ferramenta ‘conversão em lote’, do programa *Subtitle Edit*, também foi utilizada para conseguirmos retirar alguns códigos que o *Aegisub* destaca com um símbolo “”, como é exemplificado nas linhas 342 a 344 da Figura 17. Quanto aos códigos como os que são apresentados nas linhas 112 a 119 (onde aparece “m 0 0 m 100 100”), na Figura 14, tivemos que optar por realizar uma remoção manual. Todo esse processo de configurar um formato de texto personalizado, a partir da ferramenta *conversão em lote*, no *Subtitle Edit*, e o uso da própria ferramenta para remover alguns códigos de formatação específicos até a obtenção do arquivo final em bloco de notas, estão exemplificados nas Figura 17, Figura 18, Figura 19, Figura 20, Figura 21, Figura 22, Figura 23, Figura 24, Figura 25, Figura 26 e Figura 27.

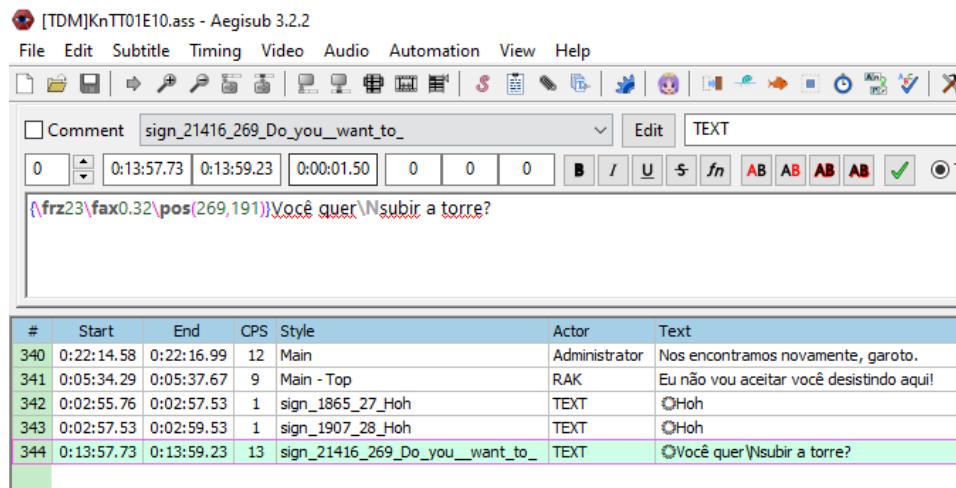


Figura 17. Código de efeitos nas linhas de legendas.

Fonte: Elaboração nossa.

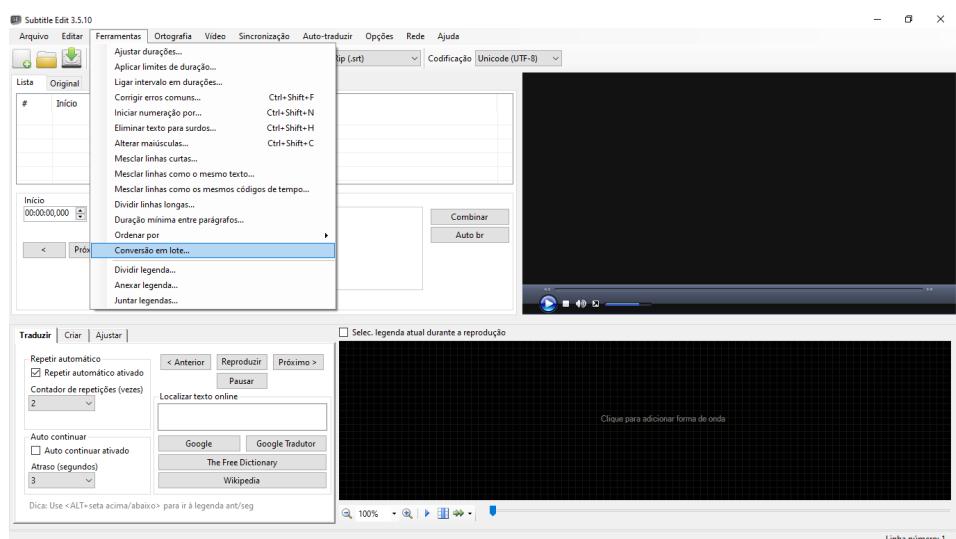


Figura 18. Abrindo a ferramenta “Conversão em lote” na interface do *Subtitle Edit*.

Fonte: Elaboração nossa.

Olhando para a linha 8 na Figura 26 (antes) e para essa mesma linha na Figura 27 (depois), conseguimos perceber que as tags de formatação foram removidas do arquivo de legenda. Destacamos que ambos os programas de legendagem supramencionados têm funções que não foram exploradas por estas pesquisas por fugirem aos interesses presentes, embora reconheçamos que algumas delas podem facilitar processos de padronização, como também de substituição de padrões. A ferramenta *conversão em lote*, por exemplo, poderia ser empregada para aplicar uma substituição múltipla de determinadas palavras ou códigos, tomando diversos arquivos de uma só vez, a depender do contexto. Porém, preferimos nos limitar às funções apresentadas, pois já nos trouxeram resultados satisfatórios.

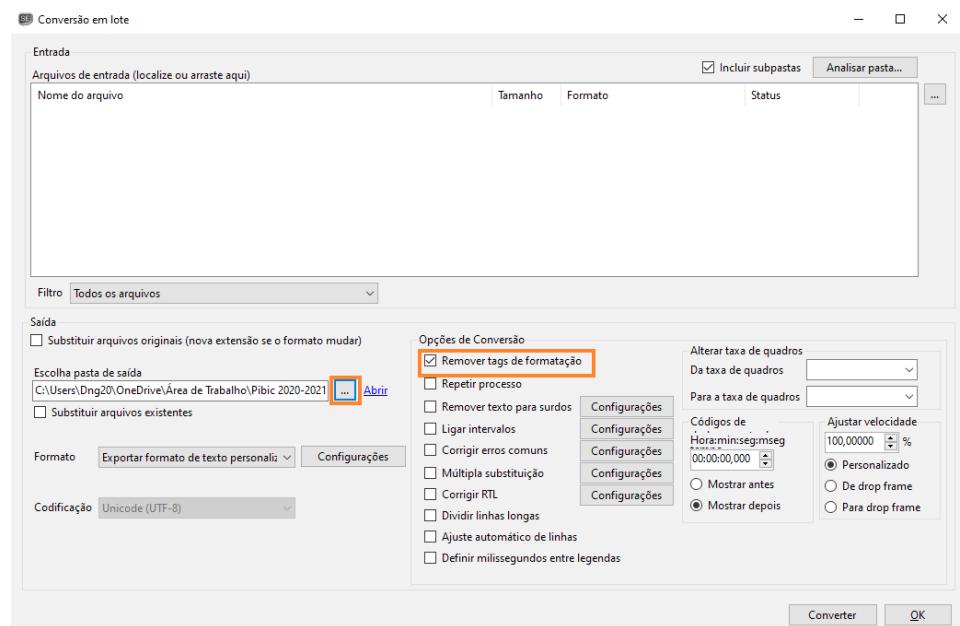


Figura 19. Selecionando a opção “Remover tags de formatação” e definindo a pasta onde os arquivos convertidos serão salvos.

Fonte: Elaboração nossa.

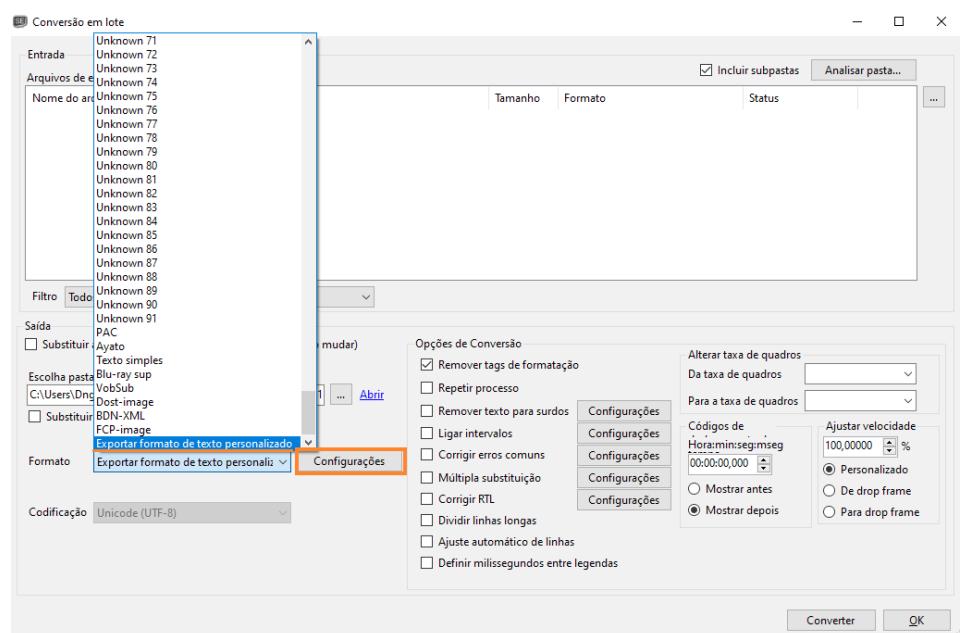


Figura 20. Selecionando a opção para exportar um formato de texto personalizado e abrindo as configurações para exportação de formato de texto personalizado.

Fonte: Elaboração nossa.

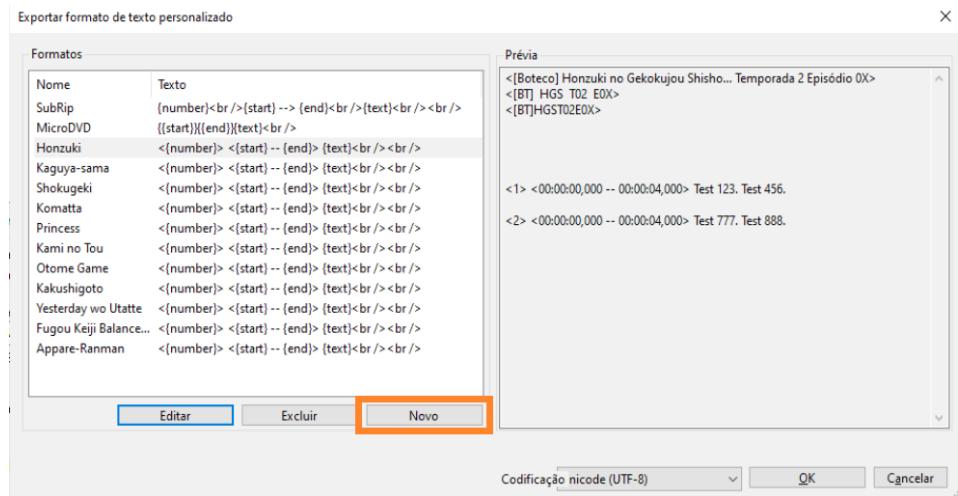


Figura 21. Criando um novo formato de texto personalizado.

Fonte: Elaboração nossa.

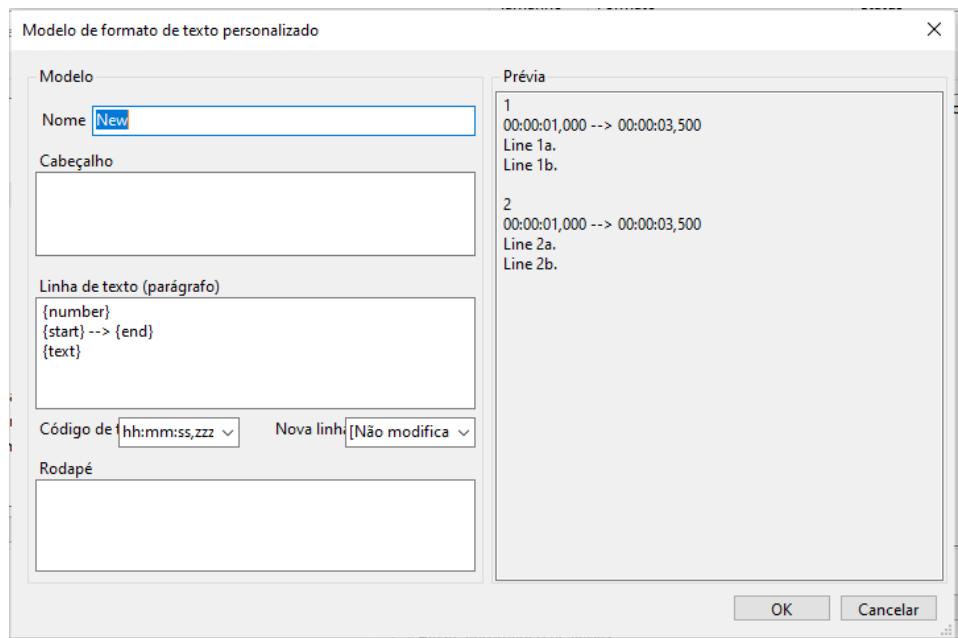


Figura 22. Interface para a configuração de um formato de texto personalizado.

Fonte: Elaboração nossa.

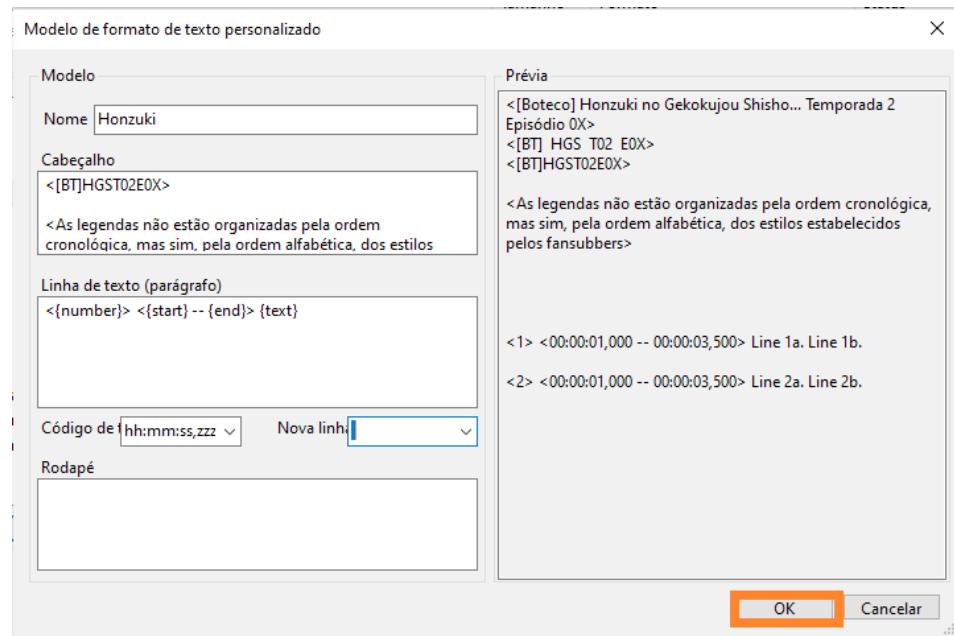


Figura 23. Novo formato de texto após ser configurado.

Fonte: Elaboração nossa.

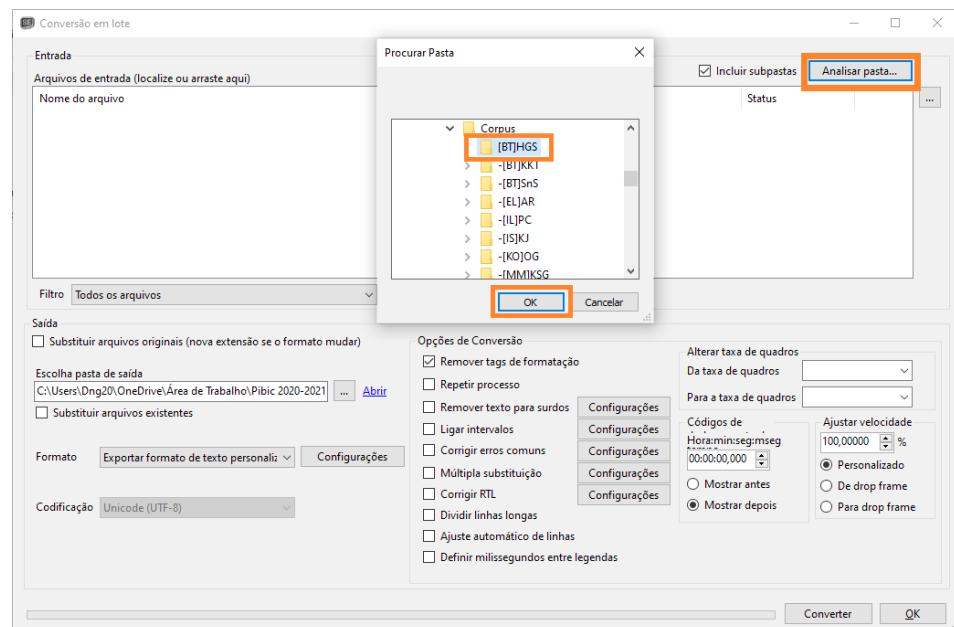


Figura 24. Selecionando pasta com as legendas a serem convertidas.

Fonte: Elaboração nossa.

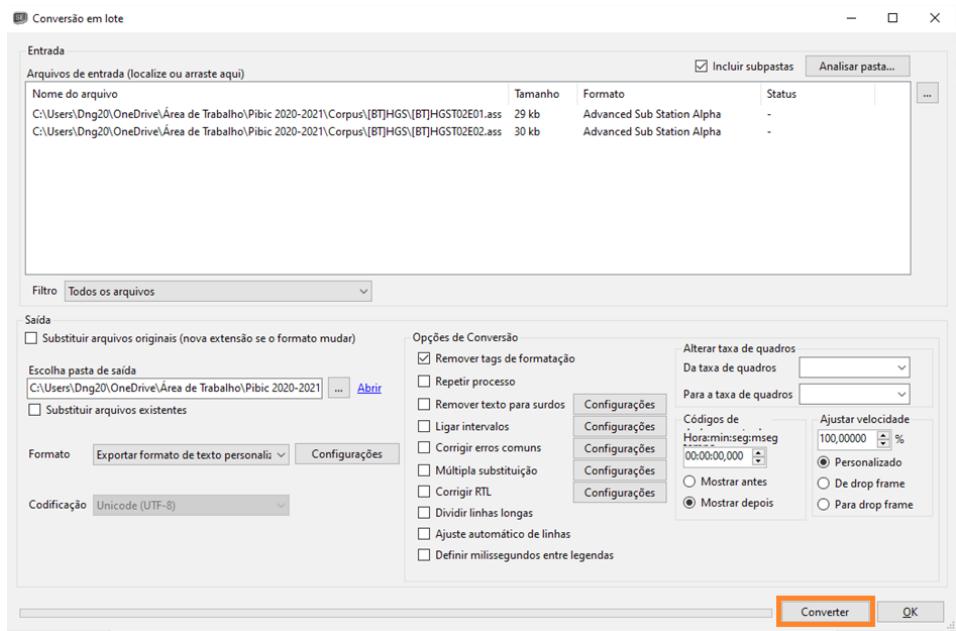


Figura 25. Legendas prontas para serem convertidas.

Fonte: Elaboração nossa.

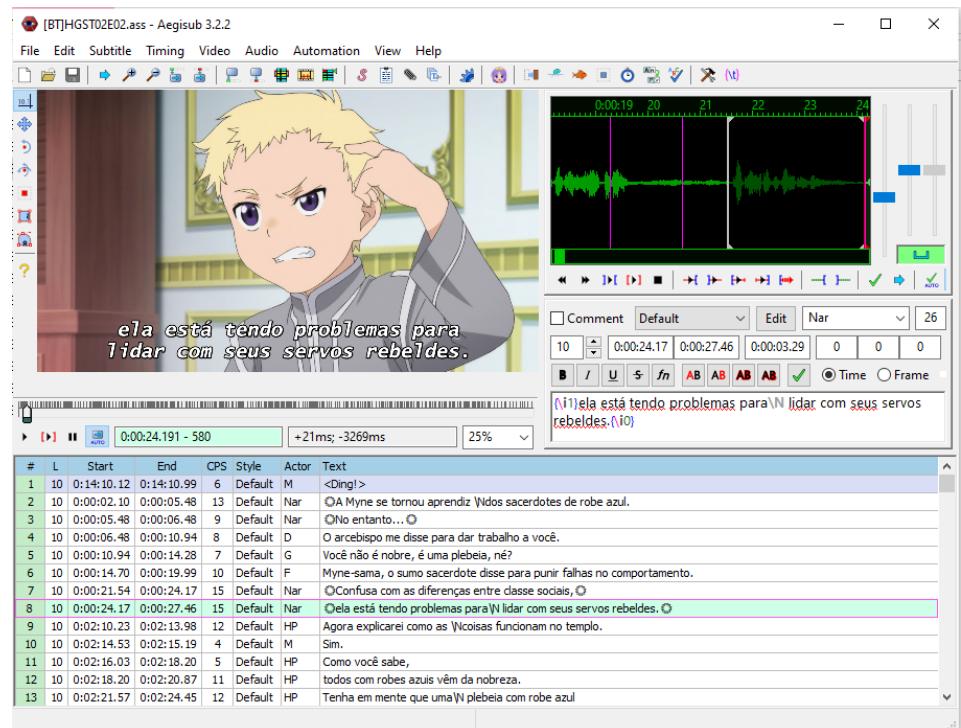


Figura 26. Arquivo de legenda antes da exportação.

Fonte: Elaboração nossa.

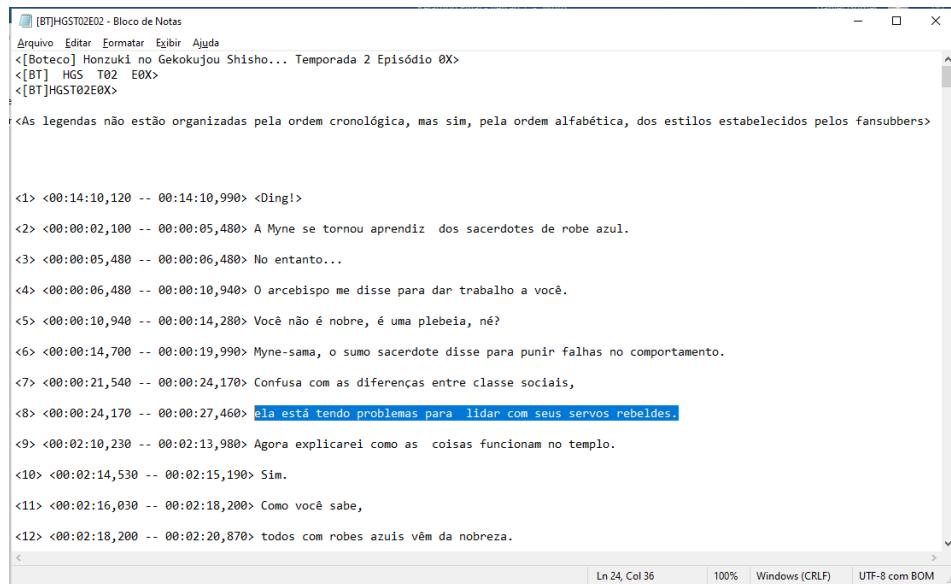


Figura 27. Arquivo de legenda após a exportação.

Fonte: Elaboração nossa.

3.4 Cabeçalhos dos arquivos de legendas

Utilizando o processo relatado no tópico anterior como vantagem, padronizamos a maioria dos cabeçalhos das legendas para cada obra, e, ao fim da conversão, os cabeçalhos já estavam em padrões pré-definidos. As Figura 28 e Figura 29 exemplificam esse processo no anime *Fruits Basket*, traduzido pelo grupo *Aenianos*.

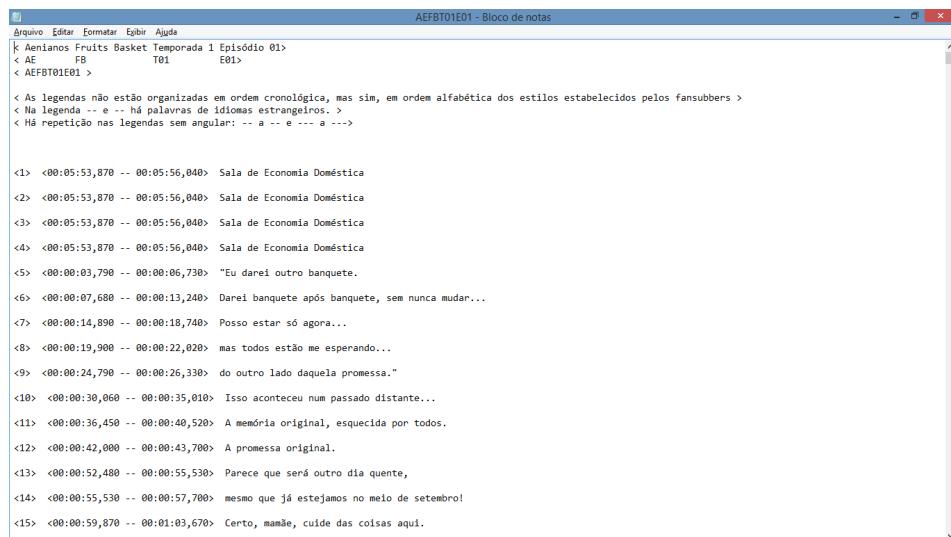


Figura 28. Cabeçalho pré-definido do exemplo do processo realizado na conversão em lote.

Fonte: Elaboração nossa.

No geral, colocamos nas etiquetas informações que dizem respeito à duplicação de legendas, e indicamos as legendas onde há palavras estrangeiras ou que fazem referência a uma, como demonstrado na Figura 30. Além disso, compõem os cabeçalhos os seguintes dados: o nome do grupo de *fansubbers*, o nome do *anime*, a temporada e o episódio em extenso, dos quais surge o código do episódio, também presente nos cabeçalhos e que pode ser visto nas Figura 28 e Figura 29.

Tendo percorrido todos os passos metodológicos relatados até então, os arquivos já se encontravam aptos a serem imputados na ferramenta *WordList* do *WordSmith Tools*® (versão 7). O procedimento seguinte consistiu em criarmos a lista de palavras e obtermos os dados estatísticos do *CorLeAni*.

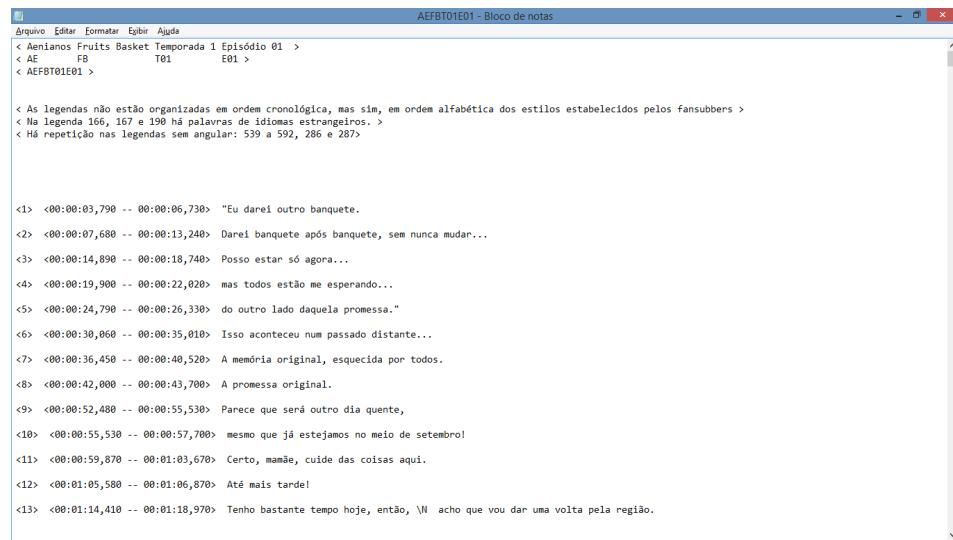


Figura 29. Cabeçalho do episódio 1 do anime “Fruits Basket”.

Fonte: Elaboração nossa.

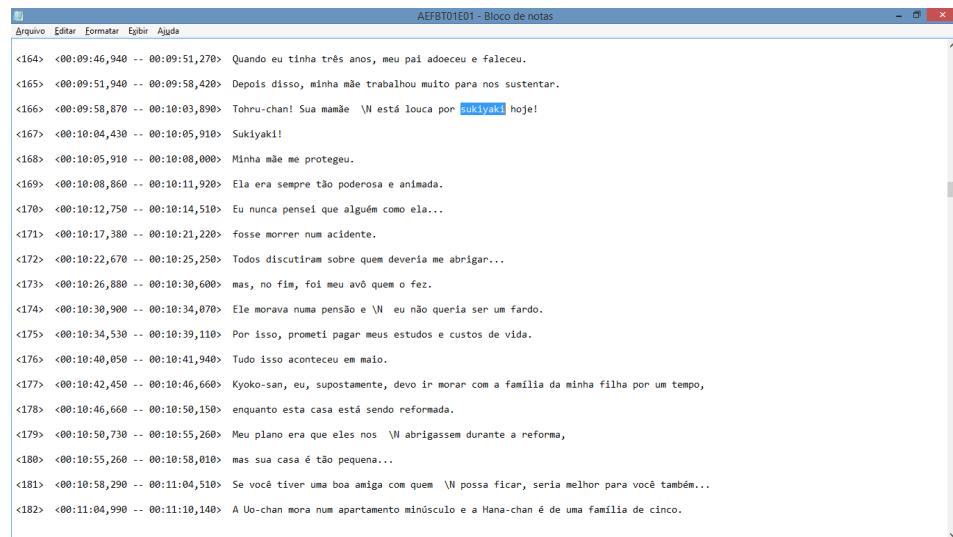


Figura 30. Palavra “sukiyaki” indicada no cabeçalho do episódio 1 do anime “Fruits Basket” como uma palavra estrangeira.

Fonte: Elaboração nossa.

3.5 Consolidando o CorLeAni com auxílio do *WordSmith Tools*

Nesta etapa da pesquisa, foi feito o uso do WordSmith Tools para compilação do corpus, conforme exemplificado de forma resumida, nas Figura 31, Figura 32 e Figura 33.

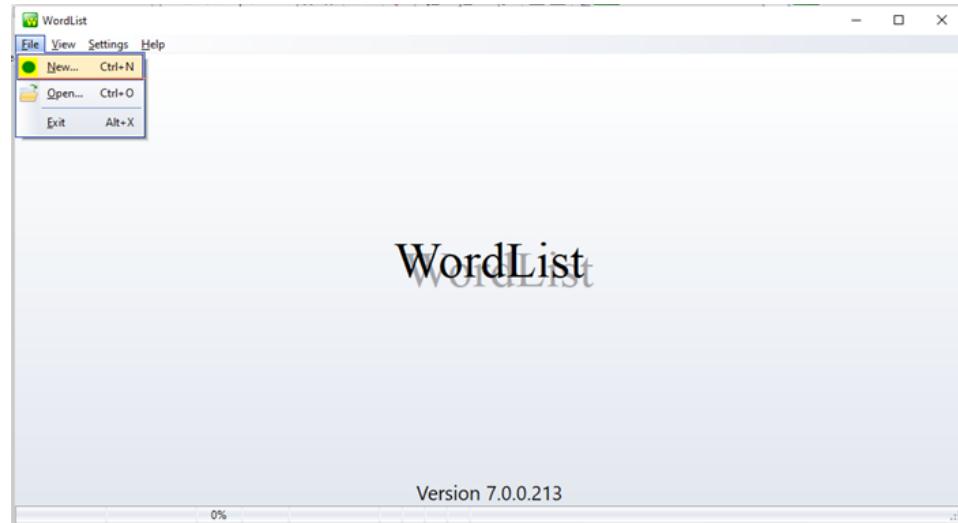


Figura 31. Interface do *WordList*.

Fonte: Elaboração nossa.

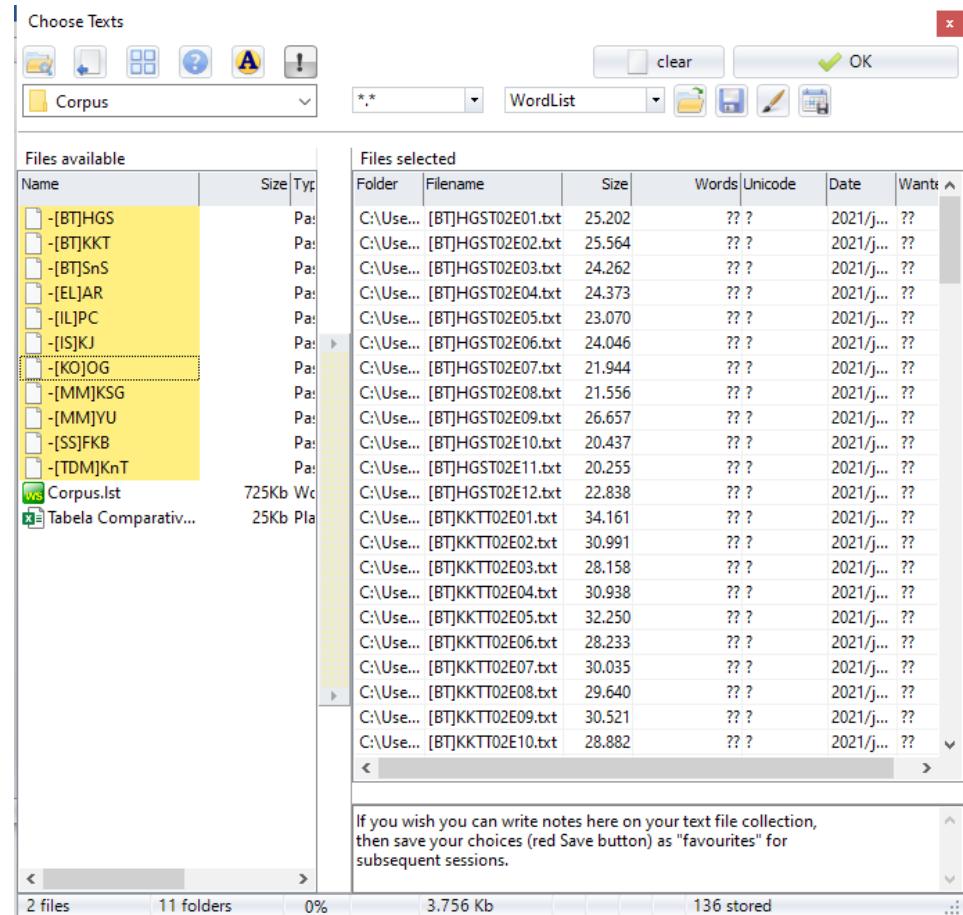


Figura 32. Indicando o diretório dos arquivos de legenda para gerar uma lista de palavras (exemplo: textos do PIBIC 2020-2021).

Fonte: Elaboração nossa.

N	Word	Freq.	%	Texts	%	Dispersion	Lemmas	Set
2.999	CASAS	6	6	4,84	0,58			
3.000	CAVALO	6	5	4,03	0,45			
3.001	CHAMADOS	6	6	4,84	0,58			
3.002	CHAMARIA	6	6	4,84	0,51			
3.003	CHASE	6	5	4,03	0,35			
3.004	CHEGARÁ	6	6	4,84	0,58			
3.005	CHICAGO	6	2	1,61	0,00			
3.006	CHINESA	6	3	2,42	0,00			
3.007	CHOQUE	6	5	4,03	0,45			
3.008	CIRCO	6	5	4,03	0,35			
3.009	CIRURGIA	6	3	2,42	0,45			
3.010	COMBINAÇÃO	6	5	4,03	0,35			
3.011	COMEM	6	6	4,84	0,51			
3.012	COMIDOS	6	2	1,61	0,17			
3.013	COMPETIDORES	6	4	3,23	0,30			
3.014	COMUNS	6	5	4,03	0,51			
3.015	CONCLUIDA	6	5	4,03	0,17			
3.016	CONDICÃO	6	5	4,03	0,67			
3.017	CONECTA	6	3	2,42	0,35			
3.018	CONFESSAR	6	5	4,03	0,58			
3.019	CONFIA	6	5	4,03	0,45			
3.020	CONHECIDA	6	6	4,84	0,67			
3.021	CONQUISTA	6	6	4,84	0,51			
3.022	CONSERTA	6	3	2,42	0,35			

Figura 33. Lista de palavras após ser criada.

Fonte: Elaboração nossa.

4 Apresentando e discutindo os resultados do CorLeAni

Concluídas as principais fases das pesquisas, descritas na Seção 3, voltamos os nossos olhares para os dados estatísticos obtidos. Para tanto, exibimos os gráficos nas Figura 34 e Figura 35, em que, no primeiro, comparamos os *types* e os *tokens*, e, no segundo, o *TTR* (*Type/Token Ratio*) e *STTR* (*Standardized Type/Token Ratio*).

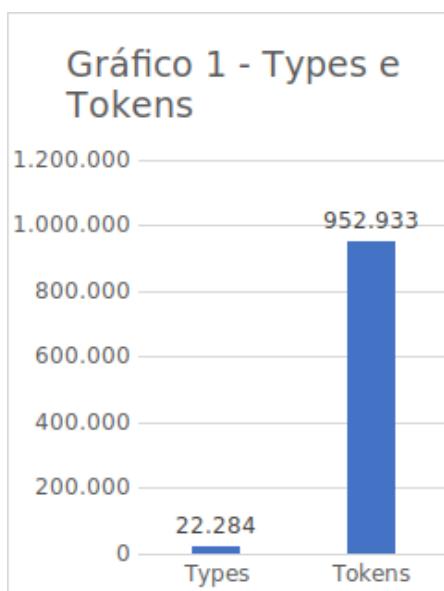


Figura 34. Gráfico 1 que mostra os resultados gerais do CorLeAni das pesquisas.

Fonte: Elaboração nossa.

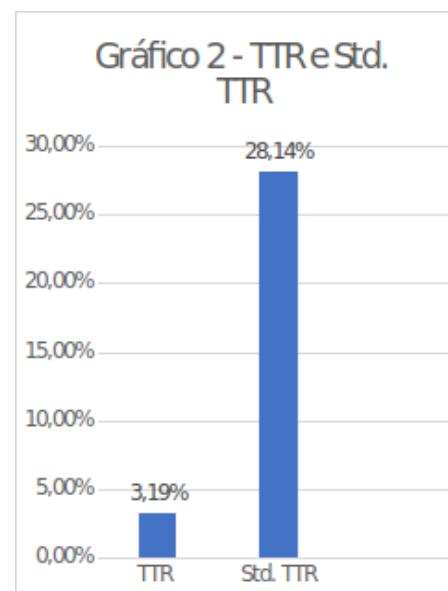


Figura 35. Gráfico 2 que mostra os resultados gerais do CorLeAni das pesquisas.

Fonte: Elaboração nossa.

Na Figura 34, observamos a quantidade de palavras diferentes no *corpus* (*types*) e palavras gerais (*tokens*), o que indica uma extensão considerável para o CorLeAni, isto é, quase 1 milhão de palavras. Na Figura 35, apresentamos outros dados em porcentagem, o *TTR* (*Type/Token ratio*) e *Standar-dized TTR*, que revelam a variação lexical presente do *corpus*. O primeiro é uma porcentagem de *Types* (palavras diferentes) em relação aos *Tokens* (todas as palavras). O segundo é uma média de porcentagens de *Types* em relação aos *Tokens*, que são feitas a cada 1.000 palavras pelo *WordSmith Tools*.

Um achado interessante que corrobora os dados estatísticos obtidos é a ocorrência de palavras repetidas no *corpus*, como podemos observar na Figura 36, conforme atestados pelos resultados obtidos por meio da ferramenta *Concord*.

Lançamos a hipótese de que essa repetição tenha aparecido nos arquivos de legenda por se tratar da apresentação, na tela, de legendas karaokês ou legendas que estão traduzindo uma parte do cenário, tais como placas, letreiros, entre outras. Trata-se de informações que, ao serem traduzidas, eventualmente demandam a criação de várias linhas de legenda. Para exemplificar tal questão da tradução intersemiótica, podemos pensar em uma cena de uma placa em movimento. Nesse caso, pode-se ter a criação de várias legendas sequenciais e a repetição da mesma frase para tornar a palavra mais visível, de forma que ela apareça na placa mesmo estando em movimento.

Outro achado interessante no *corpus* foi a presença de números e letras individuais utilizados para construir uma frase na vertical, ou, ainda, para realizar algum efeito quando da aparição do vídeo. As Figura 37 e Figura 38 retratam esses resultados.

Para compartilhar os arquivos do CorLeAni e divulgar os resultados das pesquisas, desenvolvemos um *site* que se encontra atualmente hospedado em um domínio grátis na web. Deixamos disponíveis, ademais, um *link* de acesso que redireciona a um serviço de armazenamento da Microsoft, o *OneDrive*, e ao serviço de armazenamento da *Google*, o *Google Drive*. O intuito é o de manter possível a realização do *download* de todo material compilado.

N	Concordance
321	... Os fofinhos... e os gentis. Estou tão feliz por poder passar meus dias com eles. Esta noite me
322	dias com eles. Esta noite me fez ficar agradecida por tudo outra vez. Muito obrigada. Me desculpem
323	por tudo outra vez. Muito obrigada. Me desculpem por não ser uma anfitriã melhor... Por favor, voltem
324	. Me desculpem por não ser uma anfitriã melhor... Por favor, voltem sempre... Não me use como muleta!
325	Por favor, voltem sempre... Não me use como muleta! Por favor, cuide-se bem. Você também, Tohru-san.
326	sua escola! E o Haru também! O quê?! Por que você e o Haru entraram na nossa escola
327	surpreender vocês! Tá falando sério? Até entendo por que fizeram vista grossa pro seu cabelo, mas
328	surpreso de N terem deixado o Haru entrar. Quê? Por quê? Honda-san, parece que você ainda não
329	xonadinhos de coração! Pare de dizer essa merda! Por quê? Você é contra a fofura? Olhinhos de
330	Agenda das Provas Finais Matemática Inglês Física POR FAVOR, NÃO RETIRE O LIXO ANTES DO
331	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
332	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
333	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
334	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
335	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
336	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
337	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
338	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
339	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
340	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
341	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
342	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
343	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
344	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
345	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
346	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
347	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
348	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
349	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
350	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
351	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
352	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
353	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
354	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
355	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
356	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
357	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO
358	SUA PRÓXIMA COLETA É NO DIA 14 DE MARÇO POR FAVOR, NÃO RETIRE O LIXO ANTES DO

Figura 36. Contexto da palavra “por” que aparece diversas vezes no arquivo de legenda, conforme *Concord*, ferramenta do *WordSmith Tools*© (versão 7).

Fonte: Elaboração nossa.

Concord									
N	Concordance	Set	Tag	Word	#Sents	Sent.	Para.	Para.	Head
		#Sents							
1	rogo para que me abras um caminho! Classe 2 1 0 0 Milhões Po	0	1.586	01...4	01...4	01...4	01...4	01...4	EAIQTE12.txt
2	rogo para que me abras um caminho! Classe 2 1 0 0 Milhões Po n	0	1.591	01...9	01...9	01...9	01...9	01...9	EAIQTE12.txt
3	N de olhos fechados. Com isso, são 51 vitórias e 50 derrotas para mim. Você tem que praticar mais. Vou	0	2.079	02...7	02...7	02...7	02...7	02...7	EAOPMTTE11.txt
4	, a soma do impulso ao redor do ponto C tem que ser 0. No ponto A, a força Na que o cilindro absorve é...	0	1.621	01...9	01...9	01...9	01...9	01...9	IFBBDT01E10.txt

Figura 37. Contexto do número 0 no *Concord*.

Fonte: Elaboração nossa.

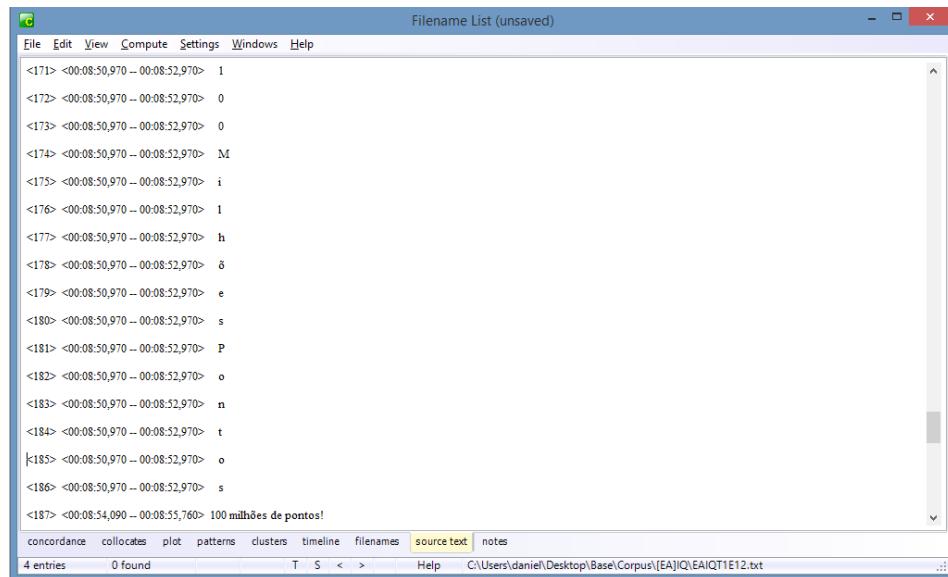


Figura 38. Número 0 e outras letras no arquivo de legenda.

Fonte: Elaboração nossa.

5 Considerações finais

Neste artigo, tivemos como objetivo reportar os resultados de dois projetos de iniciação científica que resultaram na compilação do CorLeAni. De acordo com Díaz-Cintas e Sánchez (2006, p. 1), “não seria exagero afirmar que as *fansubs* são a manifestação mais importante da tradução de legendas feita por fãs”¹². Como forma de alavancar esses estudos, reconhecemos o valor que as *fansubs* têm a agregar às discussões e aos estudos no âmbito audiovisual e *fandom*.

Até o fim do desenvolvimento das pesquisas, o CorLeAni contava com cerca de 1 milhão de palavras, número bastante representativo para servir a futuras investigações no âmbito da tradução e linguagens. Contudo, é importante apontarmos que, durante a execução dos projetos, enfrentamos diversos desafios. Um deles foi a quantidade de legendas selecionadas pelos critérios, um total de 181 arquivos de legendas (PIBIC 2019-2020) e 136 arquivos de legendas (PIBIC 2020-2021), para diferentes obras e legendadas por grupos distintos de *fansubbers*. Esse fato fez com que o tempo demandado na etapa de coleta de dados e edição das legendas fosse maior do que o esperado.

A grande quantidade de linhas de legendas com a mesma frase ou com apenas uma letra nos arquivos de legenda foi, para nós, um grande impasse. Contudo, nos foi possível refletir em torno de perguntas como “O que é uma *fansub*?”, “O que caracteriza uma *fansub*?”. As questões nos levaram a utilizar, como principal referencial teórico, o artigo de Díaz-Cintas e Sánchez (2006). A partir da leitura do texto, conseguimos chegar ao entendimento de que essas peculiaridades são parte das legendas dos fãs, e retirá-las ou desconsiderá-las da composição do *corpus* seria o mesmo que descaracterizá-las. Com isso em mente, preferimos criar etiquetas nos cabeçalhos das legendas, indicando as partes em que esses casos eram notados.

No entanto, para as informações que não são exibidas ao público, não conseguimos ter o mesmo entendimento das duplicações nas linhas de legendas, pois, em alguns momentos, foi possível constatar a presença de *links* de *sites* de pesquisa. Assim, resolvemos colocá-los entre angulares (<>) de forma a não serem contabilizados pelo *WordList*. Contudo, alguns exemplares permanecerem nos arquivos de legendas. Isso não impede, todavia, a possibilidade de serem utilizados em futuros estudos sobre *fansubs*.

Os códigos de formatação e a repetição de frases nas linhas dos arquivos de legenda foi algo que nos trouxe um grande impasse, acarretando uma demanda maior de tempo na etapa de edição de legendas, potencializada principalmente pela quantidade de arquivos a serem editados. Após um

¹² “It would be no exaggeration to state that *fansubs* are nowadays the most important manifestation of fan translation [...]” (DÍAZ-CINTAS; SÁNCHEZ, 2006, p. 1, tradução nossa).

estudo teórico e de possíveis ferramentas, tomamos as decisões que foram relatadas anteriormente neste artigo na Seção 3.3.1, que, para o caso dos códigos de formatação, foi utilizar a ferramenta “conversão em lote”.

Destacamos ainda que a quantidade de palavras (+255.287 *tokens*) inseridas na versão do *corpus* do PIBIC-EM 2020-2021 foi menor em comparação com o PIBIC-EM 2019-2020 (697.646 *tokens*). Atualmente, o CorLeAni em si tornou-se maior e mais representativo do que a sua versão anterior, no sentido da consolidação do *corpus* por meio da adição de mais palavras, advindas de legendas de novos *animes* de gêneros diversos (ação, comédia, romance, entre outros) e do lançamento de novas temporadas. Apesar de o conceito representatividade ser um ponto sobre o qual parece não haver muito consenso dentro da LC (BERBER SARDINHA, 2004), reiteramos que, para os objetivos das nossas pesquisas, o *corpus*, em sua atual versão, configura-se como representativo da população de *animes* que gostaríamos de retratar, ilustrados pelas seções das Seção 3 e Seção 4.

Como pesquisadores, as problemáticas que nos foram apresentadas em ambos os projetos nos fascinaram e nos motivaram a seguir com o nosso trabalho. Hoje, após dois anos de pesquisas, percebemos que ainda há muito para se estudar e explorar sobre temas relacionados às *fansubs*. Concluímos as investigações com um sentimento de satisfação, por termos contribuído com algo que tem a sua relevância nas discussões e estudos no âmbito audiovisual. Atribuímos uma ênfase também no ato de legendar, observando como as *fansubs* fazem uso do karaokê, de notas explicativas e que mantêm, às vezes, os honoríficos (sufixos como *-Kun* ou *-sensei*, utilizados no Japão para demonstrar uma distância social), além de outras particularidades.

Ademais, ponderamos que outros pesquisadores podem realizar o download dos arquivos de ambos os projetos e alterar os dados ou arquivos de legendas para atender aos seus objetivos e particularidades de cada pesquisa. A riqueza de dados disponível no *corpus* pode iluminar pesquisas que analisem variados aspectos linguísticos e tradutórios, como é o caso, por exemplo, do estilo das *fansubs*, as suas palavras específicas, os comentários dos tradutores como paratextos que não são exibidos para o espectador, mas que trazem indícios de autoria, entre outros aspectos.

A pandemia da COVID-19 foi algo que também impactou as pesquisas, particularmente o desenvolvimento do projeto PIBIC-EM 2020-2021, por impedir encontros e debates de forma presencial com o orientador e a comunidade acadêmica. Para resolver essa questão, utilizamos inicialmente o *Skype*, que permite a realização de videochamada e ligações por voz e, em seguida, a plataforma *Google Meet*. Em suma, conseguimos solucionar os problemas e dar sequência à execução dos projetos sem mais complicações.

Por fim, avaliamos que a experiência de desenvolver as pesquisas por meio do Programa de Iniciação Científica, via concessão de bolsas PIBIC-EM/IF Goiano, foi bastante enriquecedora em várias frentes, tanto para o aluno-bolsista, quanto para o orientador. Observamos de imediato a importância da pesquisa júnior já no âmbito da educação básica, profissional e tecnológica, no sentido de despertar a vocação científica, servindo como importante forma de incentivar talentos entre estudantes com excelência acadêmica. A partir dos dois projetos desenvolvidos, foi possível, especialmente por parte do aluno-bolsista, debruçar-se sobre uma nova área de conhecimento, os seus principais autores e as teses por eles defendidas; aprender sobre metodologia científica; defrontar-se com os problemas reais de se desenvolver pesquisa e ainda assim propor soluções que levaram, de forma organizada e criteriosa, ao resultado final almejado pelos projetos. Assim sendo, almejamos que o *corpus* possa ser de grande auxílio à comunidade acadêmica e iluminar futuras pesquisas.

6 Agradecimento

Agradecemos ao Programa de Iniciação Científica do Instituto Federal de Educação, Ciência e Tecnologia Goiano a concessão de bolsas PIBIC-EM/IF Goiano para realização desta pesquisa.

Referências

BAKER, M. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target. International Journal of Translation Studies*, v. 7, n. 2, p. 223–243, jan. 1995. DOI:

- 10.1075/target.7.2.03bak. Disponível em:
<<https://benjamins.com/online/target/articles/target.7.2.03bak>>. Acesso em: 23 mai. 2022.
- BAKER, M. Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, H. (Ed.). *Terminology, LSP and translation: studies in language engineering in honour of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1996. P. 177–186.
- BERBER SARDINHA, T. *Linguística de corpus*. Barueri: Editora Manole, 2004.
- BOGUCKI, L. Amateur subtitling on the internet. In: DÍAZ CINTAS, J.; ANDERMAN, G. (Org.). *Audiovisual translation: language transfer on screen*. Grã-Bretanha: Palgrave Macmillan, 2009. P. 49–57.
- CAMARGO, D. C. de. *Metodologia De Pesquisa Em Tradução E Linguística De Corpus*. São José do Rio Preto, SP: Laboratório Editorial do IBILCE/UNESP, 2007.
- DÍAZ CINTAS, J.; MUÑOZ SÁNCHEZ, P.; MOURA, W. H. C. Fansubs: Tradução Audiovisual em um Ambiente Amador. *Cadernos de Tradução*, v. 42, n. 01, p. 1–26, mai. 2022. DOI: 10.5007/2175-7968.2022.e80264. Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/80264>>. Acesso em: 24 mai. 2022.
- DÍAZ CINTAS, J.; NEVES, J. Taking stock of Audiovisual Translation. In: THE JOURNAL OF SPECIALISED TRANSLATION (Org.). *Audiovisual Translation: taking stock*. UK: Cambridge Scholars Publishing, 2015. P. 1–7.
- DÍAZ-CINTAS, J.; REMAEL, A. *Audiovisual translation: subtitling*. Manchester, UK ; Kinderhook, NY: St. Jerome Pub, 2007.
- DÍAZ-CINTAS, J.; SÁNCHEZ, P. M. Fansubs: Audiovisual Translation in an Amateur Environment. *The Journal of Specialised Translation*, n. 6, p. 37–52, 2006.
- FERRER SIMÓ, M. R. Fansubs y scanlations: la influencia del aficionado en los criterios profesionales. *Puentes*, v. 6, p. 27–44, 2005. Disponível em: <<http://wpd.ugr.es/~greti/revista-puentes/pub6/04-Maria-Rosario-Ferrer.pdf>>. Acesso em: 23 mai. 2022.
- LUYTEN, S. B. Mangá e animê – Ícones da Cultura Pop Japonesa. *Fundação Japão em São Paulo*, p. 1–11, 2014. Disponível em: <https://fjsp.org.br/estudos-japoneses/artigo/manga_anime_sonia_luyten/>. Acesso em: 23 mai. 2022.
- MCENERY, T.; WILSON, A. *Corpus linguistics: an introduction*. Edinburgh: Edinburgh Univ. Press, 2001. (Edinburgh textbooks in empirical linguistics).
- MOREIRA, H.; CALEFFE, L. G. *Metodologia de pesquisa para o professor pesquisador*. Rio de Janeiro: Lamparina, 2008.
- NETO, A. B. *Mangás e Animês: A cultura pop japonesa no Brasil*. 2017. Trabalho de Conclusão de Curso (Licenciatura em História) – Universidade Estadual de Montes Claros – UNIMONTES, São Francisco.
- PEREIRA, I. D. Cultura pop Japonesa no Brasil. *Temática*, v. 13, n. 8, p. 46–59, 2017. Disponível em: <<https://periodicos.ufpb.br/ojs2/index.php/tematica/article/view/35730>>. Acesso em: 23 mai. 2022.
- SATO, C. A. A cultura popular japonesa: animê. In: LUYTEN, S. M. B. (Org.). *Cultura Pop Japonesa: Mangá e Animê*. 1. ed. São Paulo: Hedra, 2005. P. 27–42.

Contribuições dos autores

Daniel Gomes da Cunha: Curadoria de dados, Análise formal, Investigação, Metodologia, Programas, Validação, Visualização, Escrita – rascunho original, Escrita – revisão e edição; **Janailton Mick Vitor da Silva:** Conceituação, Aquisição de financiamento, Metodologia, Administração de projetos, Recursos, Supervisão, Validação, Escrita – revisão e edição.