



UNIVERSIDADE ESTADUAL DO CEARÁ
CENTRO DE HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA APLICADA
MESTRADO ACADÊMICO EM LINGUÍSTICA APLICADA

ÍTALO ALVES PINTO DE ASSIS

**LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE): ANÁLISE BASEADA EM
CORPUS DA SEGMENTAÇÃO LINGUÍSTICA EM *AMOR ETERNO AMOR***



FORTALEZA – CEARÁ
2016

ÍTALO ALVES PINTO DE ASSIS

LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE): ANÁLISE BASEADA EM
CORPUS DA SEGMENTAÇÃO LINGUÍSTICA EM *AMOR ETERNO AMOR*

Dissertação apresentada ao Programa de Pós-graduação em Linguística Aplicada do Centro de Humanidades da Universidade Estadual do Ceará, como requisito parcial para a obtenção do grau de mestre em Linguística Aplicada. Área de concentração: Linguagem e interação.

Orientadora: Profa. Dra. Vera Lúcia Santiago Araújo.

FORTALEZA – CEARÁ

2016

Dados Internacionais de Catalogação na Publicação

Universidade Estadual do Ceará

Sistema de Bibliotecas

Assis, Ítalo Alves Pinto de.

Legendagem para surdos e ensurdecidos (LSE):
análise baseada em corpus da segmentação linguística
em Amor Eterno Amor [recurso eletrônico] / Ítalo
Alves Pinto de Assis. - 2016.

1 CD-ROM: il.; 4 ¾ pol.

CD-ROM contendo o arquivo no formato PDF do
trabalho acadêmico com 106 folhas, acondicionado em
caixa de DVD Slim (19 x 14 cm x 7 mm).

Dissertação (mestrado acadêmico) - Universidade
Estadual do Ceará, Centro de Humanidades, Programa
de Pós-Graduação em Linguística Aplicada, Fortaleza,
2016.

Área de concentração: Linguagem e interação.

Orientação: Prof.ª Ph.D. Vera Lúcia Santiago
Araújo.

1. Tradução Audiovisual Acessível. 2. Legendagem
para Surdos e Ensurdidos. 3. Segmentação
Linguística. 4. Linguística de Corpus. I. Título.

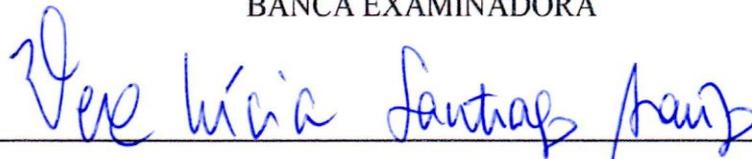
ÍTALO ALVES PINTO DE ASSIS

LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE): ANÁLISE BASEADA EM
CORPUS DA SEGMENTAÇÃO LINGUÍSTICA EM 'AMOR ETERNO AMOR'

Dissertação apresentada ao Programa de Pós-graduação em Linguística Aplicada do Centro de Humanidades da Universidade Estadual do Ceará, como requisito parcial para a obtenção do grau de mestre em Linguística Aplicada. Área de concentração: Linguagem e interação.

Aprovada em: 18/03/2016

BANCA EXAMINADORA



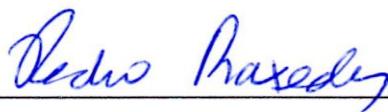
Profª. Dra. Vera Lúcia Santiago Araújo (Orientadora)

Universidade Estadual do Ceará – UECE



Profª. Dra. Ednusia Pinto de Carvalho

Universidade Federal do Ceará – UFC



Prof. Dr. Pedro Henrique Lima Praxedes Filho

Universidade Estadual do Ceará – UECE

Às duas pessoas mais importantes da minha vida: Francisca Ivone Pinto de Assis (mãe) e José Alves de Assis Neto (pai).

AGRADECIMENTOS

A Deus, por sua onipresença em minha vida, por sua destra que me sustenta e por suas misericórdias que se renovam todas as manhãs.

Aos meus pais, Ivone e Neto, por, mesmo em meios a dificuldades de todos os tipos, financeiras, principalmente, enxergarem na educação (que eles não tiveram) uma maneira de minimizar em minha vida as agruras pelas quais eles passaram, pelo apoio, amor e sustento ao longo da vida.

Ao meu irmão, Igor, pela companhia ao longo da vida, mesmo em meio as nossas brigas e discussões próprias de irmãos.

A minha namorada, Jayne, com quem dividi as dificuldades, indecisões e aflições dos últimos meses de escrita deste trabalho, pelo companheirismo de sempre, por ser a válvula de escape das tribulações da vida, pelo amor, carinho e afeto demonstrados, pelos risos e gaiatices compartilhados, por me fazer querer ser uma pessoa melhor todos os dias.

Aos amigos de diferentes círculos sociais: Jaciara, Luiz Ricardo, Expedito, Jael, Lucas e Jefferson, pela leveza das conversas e das risadas.

Aos meus irmãos da Igreja Batista Ágape em Parquelândia, aos quais não irei nominar por receio de esquecer alguém, mas que, na comunhão compartilhada, me ajudaram indiretamente nessa jornada.

A minha orientadora, Vera Santiago, muitas vezes mãe, pela dedicação e zelo pela pesquisa, por ser o pilar de apoio de tantos orientandos e de um laboratório com tantas pessoas, por ser minha tutora acadêmica desde a graduação e acreditar no meu potencial.

Aos amigos da turma de mestrado, em especial, Meysse, Jéssica, Hugo, Lindolfo, Paulo, Idália e Marcus, com os quais dividi essa jornada, com muito bom humor, em meio aos prazos apertados e cobranças da vida acadêmica.

A todos os membros do Grupo LEAD (esses sim são muitos e eu iria esquecer alguém!), com os quais aprendi muito do pouco que sei ao longo desses 4 anos de convívio.

Aos professores do PosLA, em especial, à Profa. Dra. Antônia Dilamar Araújo e à Profa. Dra. Maria Helenice Araújo Costa, pelos ensinamentos que ecoam para além dos muros da academia.

Aos Prof. Dr. Wilson Júnior de Araújo Carvalho e Profa. Dra. Marisa Ferreira Aderaldo, por suas contribuições por ocasião da Qualificação deste trabalho, mesmo que o tema proposto inicialmente tenha mudado bastante até esta versão final.

Aos Prof. Dr. Pedro Henrique Lima Praxedes Filho e Profa. Dra. Ednúsia Pinto de Carvalho, por suas contribuições por ocasião da Defesa deste trabalho que refletiram na versão final do mesmo.

À Coordenação do PosLA, em especial às secretárias Keiliane Dantas e Jamille Azevedo, pela prestatividade e por responder, com paciência, às minhas inúmeras dúvidas semanalmente.

À Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (CAPES), pelo auxílio financeiro para a realização desta pesquisa.

“Porque dEle e por Ele, e para Ele, são todas as coisas; glória, pois, a Ele eternamente. Amém.”

(Romanos 11: 36)

RESUMO

Segmentação em legendagem refere-se à divisão da fala traduzida em porções de texto ao longo da tela, podendo ser relacionada à semântica, à lexicogramática e à fonologia-fonética (segmentação linguística), ao ritmo da fala (segmentação retórica) e aos cortes de cena (segmentação visual) (ARAÚJO; ASSIS, 2014). Resultados de uma pesquisa de recepção realizada sobre Legendagem para Surdos e Ensurdidos (LSE) com 34 surdos de 4 regiões brasileiras sugeriram que uma segmentação apropriada pode garantir uma leitura confortável à comunidade surda mesmo quando essa legenda possui velocidades de 160 e 180 palavras por minuto (ppm) (NASCIMENTO; ARAÚJO, 2011; ARAÚJO, 2012). Pesquisas descritivas anteriores (ARAÚJO; ASSIS, 2014; GABRIEL, 2015) indicaram regularidades no que diz respeito aos problemas de segmentação linguística na LSE de telenovelas veiculada pela TV do Brasil: os sintagmas verbal e nominal são os mais problemáticos e os problemas, em geral, aparecem com mais frequência em legendas com 3 linhas e de velocidade alta (180ppm). Nesse contexto, a presente pesquisa teve por objetivo: i) descrever-classificar os problemas de segmentação linguística da LSE do tipo *closed caption pop-on* de 5 capítulos da telenovela *Amor Eterno Amor* e ii) propor a relegendagem da LSE de um trecho dessa telenovela, enfatizando o parâmetro da segmentação linguística no âmbito desse processo tradutório. Para o cumprimento do primeiro objetivo, uma metodologia baseada em *corpus* foi viabilizada com o uso de etiquetas discursivas próprias para esse tipo de análise, primeiramente desenvolvidas por Chaves (2012) e refinadas por Araújo e Assis (2014) e por Araújo (2015) no âmbito do Projeto CORSEL (*Corpus*, Segmentação e Legendagem). Os arquivos anotados foram, posteriormente, analisados com o programa de análise linguística *Wordsmith Tools 5.0* e os dados gerados interpretados quanti-qualitativamente. Para o cumprimento do segundo objetivo, o programa de legendagem *Subtitle Workshop 6.0B* e seus recursos de marcação, tradução e pré-visualização do vídeo legendando foram utilizados. Essa relegendagem foi feita a partir dos parâmetros técnicos, linguísticos e paralinguísticos abalizados pelo Grupo LEAD (Legendagem e Audiodescrição). Os resultados da pesquisa apontaram um maior número de problemas no sintagma verbal (43,8%) e no sintagma nominal (32,8%), assim como em legendas com velocidade alta (180ppm) e de 3 linhas, reforçando os resultados de pesquisas anteriores. A análise da atividade de relegendagem reforçou a ideia de se conceber a LSE holisticamente, sendo a mesma uma modalidade de Tradução Audiovisual (TAV) Acessível que engloba os parâmetros técnicos, linguísticos e paralinguísticos, sendo enfatizado o parâmetro da segmentação linguística nesse processo

tradutório, assim como abalizado por pesquisadores em TAV e por pesquisadores do Grupo LEAD.

Palavras-chave: Tradução Audiovisual Acessível. Legendagem para Surdos e Ensurdidos. Segmentação Linguística. Linguística de Corpus.

ABSTRACT

Segmentation in subtitling refers to the division of spoken speech in portions of text along the screen, being related to semantics, to lexicogrammar and phonology-phonetics (linguistic segmentation), to the flow of speech (rhetorical segmentation) and to scene cuts (visual segmentation) (Araújo and Assis, 2014). The results of a reception research in Subtitling for the deaf and hard-of-hearing (SDH) with 34 deaf people from 4 Brazilian regions suggested that an appropriate segmentation can ensure comfortable reading to the deaf community even when the subtitle rate is 160 and 180 words per minute (wpm) (Araújo and Nascimento, 2011; Araújo, 2012). Previous descriptive researches (Araújo and Assis, 2014; Gabriel, 2015) have indicated regularities concerning the so-called linguistic segmentation problems in the SDH of *telenovelas* aired by Brazilian TV networks: verb phrases and noun phrases are the ones with more problems, and, in general, these problems appear more often in 3-line subtitles and high-speed subtitles (180wpm). In this context, this study aimed to: i) describe-classify the linguistic segmentation problems in the pop-on closed-caption SDH of 5 chapters of the *telenovela Amor Eterno Amor* and ii) propose a resubtitling of the SDH of an excerpt from this *telenovela*, emphasizing the linguistic segmentation parameter in the scope of this translation process. To achieve the first aim, a corpus-based methodology was made feasible by means of discursive tags specific to this kind of analysis, firstly developed by Chaves (2012), and refined by Araújo and Assis (2014) and Araújo (2015) in the scope of the CORSEL Project (*Corpus, Segmentação e Legendagem*). The annotated files were later analyzed with the help of the linguistic analysis software Wordsmith Tools 5.0, and then the data generated were interpreted quanti-qualitatively. To achieve the second aim, the subtitling software Subtitle Workshop 6.0B and its resources for the cueing, translation and previsualization of the subtitled video were used. This resubtitling was carried out based on the technical, linguistic and paralinguistic parameters vouched for the research group called LEAD (*Legendagem e Audiodescrição*). The results of this study showed a greater amount of problems in verb phrases (43.8%) and in noun phrases (32.8%), as well as in three-line subtitles and high-speed subtitles (180wpm), reinforcing the results from previous studies. The resubtitling analysis reinforced the necessity of conceiving SDH holistically, since it is an Accessible Audiovisual Translation (AVT) modality that comprises technical, linguistic and paralinguistic parameters, with the emphasis on linguistic segmentation in this translation process, as it is vouched for AVT researchers and LEAD researchers.

Key-words: Accessible Audiovisual Translation. Subtitling for the deaf and hard-of-hearing. Linguistic Segmentation. Corpus Linguistics.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Exemplo de legenda <i>roll-up</i> no Programa do Jô.....	37
Figura 2 –	Exemplo de legenda <i>pop-on</i> em O Auto da Compadecida	38
Figura 3 –	Árvore sintática para explicar a Segmentação Linguística	39
Figura 4 –	Interface CCEXtractor	51
Figura 5 –	Arquivo de legenda em formato srt. aberto no Bloco de Notas	52
Figura 6 –	Interface inicial do Wordsmith Tools 5.0	56
Figura 7 –	Arquivo anotado aberto no Bloco de Notas	57
Figura 8 –	Interface do Subtitle Workshop 6.0B.....	59
Figura 9 –	Interface do <i>Subtitle Workshop</i> utilizando a ferramenta <i>TRANSLATOR</i> <i>MODE</i>	60
Figura 10 –	Tela do <i>Subtitle Workshop</i> na guia <i>SETTINGS > ADVANCED</i>	61
Figura 11 –	Tela do <i>Subtitle Workshop</i> evidenciando os erros da legendagem	62
Figura 12 –	Tela da <i>Wordlist</i> na guia <i>statistics</i>	64
Figura 13 –	Tela da <i>concord</i> após a busca pela etiqueta <PROSEGL>	67
Figura 14 –	Exemplo de identificação de falante desnecessária	86
Figura 15 –	Proposta de relegendagem sem a identificação do falante.....	86
Figura 16 –	Legenda [Gritos]	87
Figura 17 –	Legenda [Barulho de sapos].....	89
Figura 18 –	Legenda [Música de suspense].....	90
Figura 19 –	Tela do <i>Subtitle Workshop</i> com os intervalos da LSE de partida destacados...	91
Quadro 1 –	Formato das legendas na tela.....	40
Quadro 2 –	Etiquetas de problemas de segmentação de Chaves	43
Quadro 3 –	Estrutura geral dos sintagmas e sentenças complexas.....	45
Quadro 4 –	Etiquetas de características técnicas e de problemas segmentação linguística em legendagem.....	55
Quadro 5 –	Problema de segmentação linguística - quebra no sintagma verbal (SV_composto).....	68
Quadro 6 –	Problema de segmentação linguística - quebra no sintagma verbal (SV_verbo+SAdv).....	69
Quadro 7 –	Problema de segmentação linguística - quebra no sintagma verbal (SV_verbo+SAdv).....	70

Quadro 8 –	Problema de segmentação linguística - quebra no sintagma nominal (SN_especif+SN).....	71
Quadro 9 –	Problema de segmentação linguística - quebra no sintagma nominal (SN_subst+SP)	71
Quadro 10 –	Problema de segmentação linguística - quebra no sintagma nominal (SN_subst+SAdj).....	72
Quadro 11 –	Problema de segmentação linguística - quebra no sintagma nominal (SN_núcleo).....	73
Quadro 12 –	Problema de segmentação linguística - quebra no sintagma preposicional (SP_prep+SN).....	74
Quadro 13 –	Problema de segmentação linguística - quebra no sintagma preposicional (SP_prep+SV).....	75
Quadro 14 –	Problema de segmentação linguística - quebra no sintagma adjetival (SAdj_especif+SAdj).....	76
Quadro 15 –	Problema de segmentação linguística - quebra no sintagma adverbial (SAdv_adv+adv).....	77
Quadro 16 –	Problema de segmentação linguística - quebra em oração coordenada (COORD_conj+oração).....	78
Quadro 17 –	Problema de segmentação linguística - quebra em oração subordinada (SUBORD_conj/ pron_rel+oração).....	79
Quadro 18 –	Trecho relegendado da fala de Gracinha	88
Quadro 19 –	Trecho da LSE de partida com problema de segmentação retórica	92
Quadro 20 –	Proposta de relegendagem de LSE com problema de segmentação retórica .	92
Quadro 21 –	Exemplo de redução/condensação na proposta de relegendagem	93
Quadro 22 –	Exemplo de redução/condensação na proposta de relegendagem	94
Quadro 23 –	Exemplo de redução através do uso de sigla na proposta de relegendagem ...	94
Quadro 24 –	Exemplo de redução e ressegmentação na proposta de relegendagem	95
Quadro 25 –	Trecho da LSE original com problemas técnicos e linguísticos.....	95
Quadro 26 –	LSE original com problemas técnicos e linguísticos e a proposta de relegendagem.....	96
Quadro 27 –	LSE original com problemas técnicos e linguísticos e a proposta de relegendagem.....	97
Quadro 28 –	Relegendagem de sequência de legendas problemáticas.....	98

LISTA DE TABELAS

Tabela 1 –	Velocidade de leitura da legenda a 145ppm em cps.....	33
Tabela 2 –	Velocidade de leitura da legenda a 160ppm em cps.....	33
Tabela 3 –	Velocidade de leitura da legenda a 180ppm em cps.....	34
Tabela 4 –	VA e porcentagem dos tipos de PROSEGL em ‘Amor Eterno Amor’	66
Tabela 5 –	IFS e VA dos PROSEGL em ‘Amor Eterno Amor’	80
Tabela 6 –	IFS e VA dos PROSEGL em por gênero no projeto CORSEL.....	81
Tabela 7 –	Relação entre PROSEGL e número de linhas em	82
Tabela 8 –	Relação entre PROSEGL e velocidade da legenda em	82

LISTA DE ABREVIATURAS E SIGLAS

AEA	Amor Eterno Amor
CAPES	Coordenação de Aperfeiçoamento do Pessoal de Nível Superior
CORSEL	Corpus, Segmentação e Legendagem
Cpl	Caracteres por linha
Cps	Caracteres por segundo
IFS	Índice de Frequência Simples
LATAV	Laboratório de Tradução Audiovisual
LEAD	Legendagem e Audiodescrição
LSE	Legendagem para Surdos e Ensurdecidos
L	Linha de legenda
PcDVs	Pessoas com Deficiência Visual
Ppm	Palavras por minuto
PROSEGL	Problema de Segmentação Linguística
SAdj	Sintagma Adjetival
SAdv	Sintagma Adverbial
SN	Sintagma Nominal
SP	Sintagma Preposicional
SV	Sintagma Verbal
TAV	Tradução Audiovisual
TCR	<i>Time Code Reader</i>
UECE	Universidade Estadual do Ceará
UFMG	Universidade Federal de Minas Gerais
VA	Valor Absoluto

SUMÁRIO

1	INTRODUÇÃO	18
2	FUNDAMENTAÇÃO TEÓRICA.....	25
2.1.	LEGENDAGEM E LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE)	25
2.1.1	O <i>closed caption</i>.....	36
2.2	A SEGMENTAÇÃO NA LEGENDAGEM.....	38
2.3.	LEGENDAGEM E LINGUÍSTICA DE <i>CORPUS</i>	41
3	METODOLOGIA.....	48
3.1	TIPO DE PESQUISA.....	48
3.2	CONTEXTO DA PESQUISA.....	48
3.3	O <i>CORPUS</i>	49
3.3.1	Extração do <i>corpus</i> e armazenamento	50
3.3.2	Etiquetagem	53
3.4	A RELEGENDAGEM	57
3.4.1	<i>Subtitle Workshop 6.0B</i>	58
4	RESULTADOS DA ANÁLISE DOS DADOS	64
4.1	RESULTADOS: PANORAMA GERAL DO CORPUS E PROBLEMAS DE SEGMENTAÇÃO LINGUÍSTICA EM <i>AMOR ETERNO AMOR</i>	64
4.1.1	Panorama geral do corpus	64
4.1.2	Resultados da Etiquetagem.....	65
4.1.2.1	Sintagma Verbal (SV)	68
4.1.2.2	Sintagma Nominal (SN)	70
4.1.2.3	Sintagma Preposicional	73
4.1.2.4	Sintagma Adjetival (SAdj)	75
4.1.2.4	Sintagma Adverbial (SAdv)	77
4.1.2.5	Orações Subordinadas (SUBORD) e Coordenadas (COORD)	77
4.1.3	Índice de Frequência Simples e problemas de segmentação linguística no <i>corpus</i>	80

4.1.4	Relação entre problemas de segmentação e parâmetros técnicos	81
4.2	RELENDAGEM DE UM TRECHO DE <i>AMOR ETERNO AMOR</i>	83
4.2.1	Parâmetros Paralinguísticos	84
4.2.2	Relegendagem dos parâmetros técnicos e linguísticos	90
5	CONSIDERAÇÕES FINAIS.....	100
	REFERÊNCIAS.....	104

1 INTRODUÇÃO

A LSE – Legendagem para Surdos e Ensurdidos – tem se tornado uma das poucas saídas possíveis para que as pessoas que se encaixam no seu público alvo desfrutem da experiência de assistir a um produto audiovisual, como no caso de novelas, peças de teatro, telejornais, programas de TV, filmes, etc. Nesse contexto, a norma complementar 01/2006 aprovada pela portaria 310 do Ministério das Comunicações de 27 de junho de 2006 assevera, para um prazo de até 11 anos a partir da sua data de publicação, que a LSE, a Janela de Libras para pessoas surdas, além da Audiodescrição para Pessoas com Deficiência Visual - PcDVs, estejam presentes em toda a programação veiculada pelas estações transmissoras ou retransmissoras de radiodifusão.

Contudo, no caso específico da LSE, pesquisas de recepção realizadas pela UECE (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004, 2007, 2008; NASCIMENTO; ARAÚJO, 2011; ARAÚJO, 2012) com surdos cearenses e de todo o Brasil têm apontado para uma necessidade de ajustes no que tange os parâmetros utilizados na confecção da LSE do tipo *closed caption* – disponível na programação da TV brasileira, sugerindo que esses parâmetros não satisfazem a demanda da comunidade surda. Os resultados alcançados por uma pesquisa de recepção nacional com 34 surdos¹ de 4 regiões brasileiras (NASCIMENTO; ARAÚJO, 2011; ARAÚJO, 2012) sugeriram que legendas bem segmentadas promovem uma recepção eficaz do programa televisivo legendado mesmo que sejam rápidas (entre 160 e 180 palavras por minuto), o que nos² surpreendeu, pois acreditávamos, entre os pesquisadores do Grupo LEAD (Legendagem e Audiodescrição) com base em pesquisas anteriores (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004, 2007, 2008), que o principal entrave para uma recepção eficiente de filmes e programas legendados por parte de surdos estaria na velocidade da legenda. Na confecção das legendas para essa pesquisa de recepção foram observadas diretrizes para a segmentação das falas dos filmes em legendas preconizados por pesquisadores da área (DIAZ-CINTAS; REMAEL, 2007; REID, 1990; KARAMITROGLOU, 1998; IVARSSON; CARROLL, 1998), o que não tinha sido observado em pesquisas anteriores.

Na legendagem, a segmentação diz respeito à divisão da fala traduzida em porções de texto ao longo da tela, podendo ser relacionada à semântica, à lexicogramática e à

¹ Os ensurdidos não foram contemplados na pesquisa, pois acreditamos que uma legenda adequada para surdos contemplaria bastante suas necessidades, já que os surdos teriam menos acesso à língua portuguesa do que os ensurdidos, os quais, diferentemente dos surdos, tem o português como primeira língua.

² Quando a 1ª pessoa do plural (nós) for utilizada como sujeito do discurso ao invés da 1ª (eu) e 3ª pessoas (a pesquisa) ao longo da dissertação é porque se trata de uma ideia, assertiva ou ação que representa o ponto de vista do Grupo LEAD como um todo.

fonologia-fonética (segmentação linguística), ao ritmo da fala (segmentação retórica) e ao corte ou mudança de cena (segmentação visual) (DIAZ-CINTAS; REMAEL, 2007; REID, 1990; KARAMITROGLOU, 1998; IVARSSON; CARROLL, 1998). A segmentação linguística pode se tratar da divisão das legendas em blocos ao longo da legendagem, nomeada pelo termo geral ‘segmentação’, ou ainda entre linhas de uma mesma legenda, chamada de *line break* ou quebra de linha.

É grande o número de teóricos em Tradução Audiovisual (TAV) que destacam a relevância de uma boa segmentação para a recepção e compreensão da legenda (KARAMITROGLOU, 1998; IVARSSON; CARROLL, 1998; DIAZ-CINTAS; REMAEL, 2007; PEREGO, 2008; KALANTZI, 2008). Entre estes, Diaz-Cintas e Remael ressaltam que “uma segmentação cuidadosa da informação pode ajudar a reforçar a coerência e a coesão na legendagem³” (ibid., p. 173). No âmbito da segmentação quebra de linha, objeto de estudo da presente pesquisa, é necessário que as legendas sejam segmentadas de uma maneira que sejam autossuficientes sintática e semanticamente, dado que “quando segmentamos uma sentença, forçamos o cérebro a pausar o seu processamento linguístico por um momento até que os olhos captem a próxima informação⁴” (KARAMITROGLOU, 1998, p. 10), dessa forma, é imprescindível que essa divisão/segmentação do texto em tela siga determinadas diretrizes linguísticas. Karamitroglou (1998, p. 9-10) postula ainda que o processo de segmentação deve acontecer no mais alto nível sintático possível, e, idealmente, cada linha de legenda deveria conter uma sentença completa. Contudo, por ser inevitável que essa segmentação ocorra, essa divisão deveria ocorrer no ponto em que a carga semântica já transmitiu de forma satisfatória sua informação, o que coincide com o respeito às unidades sintagmáticas, ou seja, que não se quebre sintagmas no processo de divisão do texto legendado.

No que tange à descrição-classificação dos problemas de segmentação na LSE do tipo *closed caption*, dentre as poucas pesquisas que existem, nenhuma conseguiu, pelo pequeno escopo das mesmas, analisar uma quantidade grande de dados para um estabelecimento do perfil da segmentação (ou de suas inadequações) na LSE praticada pela TV do Brasil até o presente momento, apesar de alguns resultados em diferentes gêneros de LSE já estarem demonstrando certas regularidades em relação à sua natureza e seus padrões relacionados aos problemas de segmentação (ASSIS, 2013; ARAÚJO; ASSIS, 2014; GABRIEL, 2015; ARRAES, 2015), como uma grande quantidade de problemas de

³ Minha tradução de: "A careful segmentation of the information can help reinforce coherence and cohesion in subtitling".

⁴ Minha tradução de: "When we segment a sentence, we force the brain to pause its linguistic processing for a while, until the eyes trace the next piece of linguistic information."

segmentação nos sintagmas verbal e nominal, principalmente, e em legendas rápidas e de três linhas. Nesse contexto, há o projeto CORSEL (*Corpus*, Segmentação e Legendagem), que teve como objetivo a identificação e descrição-classificação dos problemas de segmentação na LSE da TV brasileira e é vinculado ao grupo de pesquisa LEAD da Universidade Estadual do Ceará, sendo o objeto desta pesquisa de dissertação de mestrado a LSE de cinco capítulos da telenovela *Amor Eterno Amor*, parte integrante e recorte do projeto em questão.

Os resultados de pesquisas anteriores (FRANCO; ARAÚJO, 2003; ARAÚJO 2004; ARAÚJO, 2007; ARAÚJO, 2008; ARAÚJO, 2009) corroboram com a necessidade de um número maior de pesquisas que proponham uma reflexão teórica e prática acerca da LSE que é veiculada na TV do país, sendo que essa modalidade de TAV acessível ainda se encontra distante das demandas e necessidades do público surdo/ensurdecido. Seguindo resultados de Nascimento e Araújo (2011) e Araújo (2012), o parâmetro da segmentação pode ser um dos responsáveis pela boa ou má recepção da LSE, em detrimento de outros parâmetros historicamente mencionados como problemáticos para os surdos – velocidade alta e densidade lexical, por exemplo. Dessa forma, a relevância de um estudo acerca da segmentação na LSE TV do Brasil e sua influência na recepção é inegável para que alcancemos parâmetros que tornem confortável a LSE destinada aos surdos brasileiros, sendo esta a justificativa social e motivação da presente pesquisa.

No contexto brasileiro, baseados nos resultados de Nascimento e Araújo (2011) e Araújo (2012) em relação à importância da segmentação para um processamento efetivo das legendas por surdos mencionados anteriormente, há as pesquisas de Chaves (2012) e Diniz (2012) – este resultado de um projeto de cooperação acadêmica entre a UECE e a Universidade Federal de Minas Gerais (UFMG) – acerca da categorização dos problemas de segmentação linguística em filmes brasileiros. Também se incluem dentre esses trabalhos, o meu Trabalho de Conclusão de Curso (ASSIS, 2013), um artigo oriundo deste Trabalho (ARAÚJO; ASSIS, 2014) e Gabriel (2015), estes dois sobre LSE de telenovelas e inseridos no projeto CORSEL supracitado. Em Araújo e Assis (2014), pesquisa que trouxe os resultados alcançados por Assis (2013) com alguns pequenos ajustes em relação à metodologia e à análise empregadas, apesar de ser analisada uma porção pequena do universo da LSE da telenovela *Amor Eterno Amor* (1 capítulo apenas), os resultados revelaram vários problemas de segmentação na legendagem analisada: 25,5% das legendas tinham algum tipo de problema, ou seja, mais de 1 em cada 4 legendas. Contudo, o tamanho do *corpus* analisado não nos permitiu lançar considerações mais conclusivas acerca da segmentação na LSE da telenovela em questão.

Tanto Assis (2013) e Araújo e Assis (2014) quanto os outros estudos supracitados foram baseados primariamente em Perego (2008), que, por sua vez, analisou a segmentação linguística em um *corpus* heterogêneo de filmes para DVD e cinema e, a partir dessa análise, propôs categorias para a investigação desse tipo de segmentação, como: quebra do sintagma nominal, quebra do sintagma preposicional, quebra do sintagma verbal e quebra das orações coordenadas e subordinadas.

No entanto, ao aliar o uso das ferramentas advindas da Linguística de *Corpus* e o estudo da segmentação, o trabalho de Chaves (2012) propôs uma vertente de análise que tem servido de modelo para as pesquisas do LEAD, um diferencial em relação ao estudo de Perego. A metodologia baseada em *corpus* através de programas de análise linguística, em nosso caso, o *Wordsmith Tools 5.0*, assim como proposta inicialmente por Baker (1993) para os Estudos da Tradução, tem se mostrado de grande valia para a análise descritiva da LSE e da Audiodescrição, dois dos tipos de TAV acessível que, de certa forma, ainda carecem de arcabouço teórico em relação à sua descrição.

Ao trabalhar com um número maior de dados do que Assis (2013) e Araújo e Assis (2014), irei fortalecer ou refutar com mais propriedade os resultados alcançados por estas pesquisas no que diz respeito aos problemas de segmentação no gênero telenovela, mais especificamente em *Amor Eterno Amor*, ou mesmo contrapor ou confirmar os resultados alcançados por Diniz (2012), Chaves (2012), Arraes (2015) e Gabriel (2015) em se tratando da análise de problemas de segmentação linguística em diferentes gêneros de ficção – filme e telenovela – veiculados na TV e em DVD. As regularidades nos resultados da LSE de telenovela – mesmo em *corpora* de pequena extensão – que apontam um maior número de problemas no sintagma verbal e nominal, assim como em legendas de três linhas e com velocidade considerada alta (acima de 180 palavras por minuto), por exemplo, serão confrontadas a partir da análise de um *corpus* mais extenso. Dessa forma, a expansão do *corpus* analisado em Assis (2013) e Araújo e Assis (2014) justifica a pesquisa academicamente do ponto de vista metodológico.

A minha intenção ao iniciar essa pesquisa de mestrado era descrever-classificar os problemas de segmentação linguística na LSE de uma telenovela (no caso, *Amor Eterno Amor*), vide a experiência proveitosa que tive em Assis (2013) e Araújo e Assis (2014), e, a partir dos resultados obtidos por essa descrição-classificação, investigar a recepção a legendas bem e mal segmentadas através de uma metodologia experimental cujo principal vértice seria a utilização de um rastreador ocular para mensurar a influência desses problemas de segmentação linguística na recepção de surdos e ensurdecidos. Esses dois vieses redundariam

em uma junção ainda inédita no campo de estudo da LSE, que iria, possivelmente, gerar resultados e discussões cada vez mais sólidos para a refutação ou comprovação de afirmações da literatura da área – LSE e Legendagem em um sentido mais amplo – ainda pouco estudados experimentalmente, como é o caso da relevância da segmentação na recepção do produto audiovisual legendado.

Entretanto, devido ao pouco tempo para se realizar uma pesquisa de mestrado, não foi possível que a parte experimental fosse realizada. Apesar disso, contudo, acredito que a análise baseada em *corpus* de problemas de segmentação linguística configura-se como uma problemática interessante a ser abordada por esta pesquisa. A parte experimental do trabalho ter sido relegada, possivelmente, para o futuro, propiciou que eu vislumbrasse algo ainda não realizado pelas pesquisas em LSE que sejam do meu conhecimento: a proposta de uma relegendagem de uma LSE a partir dos preceitos técnicos, linguísticos e paralinguísticos da área abalizados pelas pesquisas em Legendagem para Surdos e Ensurdidos e, mais especificamente, pelas pesquisas no âmbito do Grupo LEAD-UECE realizadas no Brasil e postos em prática pelos trabalhos de legendagem desenvolvidos pelo grupo. Desse modo, pretendo, através dessa proposta de relegendagem, demonstrar como todos esses aspectos – técnicos, linguísticos e paralinguísticos – se associam no momento da prática da legendagem, evidenciando quais parâmetros devem ser enfatizados pelo tradutor em detrimento de outros, quando necessário, como é o caso da proeminência e relevância que acreditamos que a segmentação linguística tenha em relação a outros parâmetros, por exemplo, e isso é algo que perpassa a proposta que será relatada. Em suma, através desse exercício de relegendagem, explicita-se a necessidade de se ver a legendagem holisticamente, já que todas essas características devem confluir para a entrega de uma LSE adequada para uma boa recepção de surdos e ensurdidos. Dessa maneira, a relegendagem contribui para justificar a pesquisa academicamente desta feita do ponto de vista teórico, dado que, por ser inédito, pretende preencher essa lacuna teórica no que diz respeito ao campo de estudo da LSE.

Em relação à justificativa institucional, o trabalho pode ser considerado como um recorte do projeto de pesquisa denominado CORSEL, tanto por se utilizar das mesmas prerrogativas teórico-metodológicas do projeto em questão como por utilizar como objeto de estudo um *corpus* vinculado ao mesmo. Almejo, a partir de minha pesquisa, trazer resultados sólidos para o objetivo do projeto em questão: uma descrição-classificação de como se dão os problemas de segmentação linguística na LSE veiculada pela TV brasileira. Pretendo ainda, através da proposta de relegendagem, evidenciar o *modus operandi* do Grupo LEAD-UECE a partir de parâmetros que acreditamos que, se postos em prática, possam contribuir para o

aprimoramento da LSE brasileira. Dessa forma, contribuo, de certo modo, no fortalecimento da pesquisa em LSE empreendida pelo Grupo no âmbito do Laboratório de Tradução Audiovisual (LATAV) desde 2002, pesquisa esta pioneira e referência no Brasil em se tratando de TAV acessível no país.

A partir do contexto delineado acima, tenho como intenção, a partir do presente trabalho, cumprir os objetivos e responder às perguntas indicadas a seguir:

OBJETIVO GERAL:

- a) Analisar a segmentação linguística da LSE da telenovela *Amor Eterno Amor*.

OBJETIVOS ESPECÍFICOS:

- a) Descrever-classificar os problemas de segmentação linguística da LSE de 5 capítulos da telenovela *Amor Eterno Amor*;
- b) Propor a ressegmentação de trechos da LSE de *Amor Eterno Amor* com problemas de segmentação linguística, segundo os parâmetros paralinguísticos, técnicos e linguísticos apregoados pelos pesquisadores de TAV.

QUESTÕES DE PESQUISA

- a) Quais problemas de segmentação linguística em LSE emergirão a partir da análise de 5 capítulos da telenovela *Amor Eterno Amor* e como eles classificam?
- b) Como seria a relegendagem de um trecho de *Amor Eterno Amor* com problemas de segmentação, segundo os parâmetros paralinguísticos, técnicos e linguísticos apregoados pelos pesquisadores de TAV?

Este trabalho está dividido nos seguintes capítulos para além desta introdução: no Capítulo 2, de Fundamentação Teórica, discorro sobre a natureza da LSE como modalidade de TAV acessível, além de apresentar uma revisão bibliográfica de trabalhos relacionados ao estudo da Legendagem/LSE e o parâmetro da segmentação e à interface LSE/segmentação/Linguística de *Corpus*; no Capítulo 3, reservado à metodologia, trato dos processos de extração do *corpus* e sua anotação, dentre outras etapas, incluindo o passo-a-

passo da proposta de relegendagem empreendida; no Capítulo 4, apresento e discuto os resultados advindos da análise do *corpus* e da proposta de relegendagem e, por fim, no Capítulo 5, teço algumas considerações finais acerca dos resultados e implicações da presente pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, teço considerações a respeito das áreas que se conjugam para o embasamento teórico-metodológico deste trabalho: as concepções teóricas acerca da LSE de uma forma geral; a Linguística de *Corpus*, que sustenta a metodologia descritiva a ser utilizada na quantificação dos dados, e sua interface com a Legendagem, assim como acerca da conceituação e relevância do parâmetro da segmentação linguística na Legendagem/LSE.

2.1. LEGENDAGEM E LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE)

A Legendagem para Surdos e Ensurdecidos é uma modalidade de Tradução Audiovisual Acessível destinada ao público que se encontra nessa situação, que permite que esta audiência tenha acesso, pela LSE, à informação veiculada por meio da língua oral. Além da questão linguística obviamente presente na tradução desse par oral/escrito, outros recursos presentes na língua oral e que não podem ser apreendidos apenas com a tradução das falas dos personagens em vídeo são características marcantes e que distinguem esse tipo de legendagem daquela acessada comumente pelo público ouvinte, a saber: a identificação dos falantes (explicitação na legenda do nome do personagem que detém o turno de fala em tela), a tradução dos efeitos sonoros (música e ruídos) e das características prosódicas da fala. Além dessas informações adicionais paralinguísticas presentes na LSE, questões de ordem técnica e concepções distintas de tradução são aspectos que diferenciam a LSE da legendagem para ouvintes (ARAÚJO, 2003, 2007, 2008; NASCIMENTO; ARAÚJO, 2011).

A LSE pode ser tanto um tipo de tradução intralinguística, ou seja, dentro de uma mesma língua, a tradução acontece do oral para o escrito, quanto interlinguística, de uma língua para outra, mas ainda do oral para o escrito, nos casos em que é feita a partir das versões dubladas de filmes e programas de TV em língua estrangeira⁵ (ARAÚJO; ASSIS, 2014). Considerando o fato de a tradução de efeitos sonoros estar presente na LSE, poderia se dizer, ainda, que a mesma se configura como um tipo de tradução intersemiótica, já que os efeitos sonoros são organizados como meio semiótico distinto do linguístico e, dessa forma,

⁵ Por um erro de revisão, em Araújo e Assis (2014, p. 159) dissemos que: “A LSE produzida para os canais de TV brasileiros são também intralinguísticas, visto que são produzidas a partir das versões dubladas de filmes e programas de TV em língua estrangeira”, quando, na verdade, para adicionar informação à oração anterior a esta o que faria mais sentido a partir do uso do “são também” seria, ao invés de “intralinguísticas”, “interlinguísticas”. Desse modo, retifico o erro de revisão e endosso a possibilidade de considerar a LSE como prática de tradução interlinguística também.

há, ao se traduzir sons em palavras, uma operação tradutória que perpassa dois meios semióticos.

Dessa forma, sendo as características linguísticas e técnicas da LSE as mesmas presentes na legendagem para ouvintes, irei fazer uma breve descrição dos parâmetros que caracterizam a legendagem como um todo para depois caracterizar de forma mais detalhada o objeto de estudo da pesquisa ora relatada, a LSE.

No que diz respeito às questões linguísticas da legendagem, fatores como redução do texto por condensação/reformulação, omissões/exclusões, coesão, coerência e segmentação linguística são importantes e têm participação decisiva na qualidade do produto final legendado e na sua boa recepção pelos telespectadores.

Em relação à redução do texto legendado, Diaz-Cintas e Remael (2007, p. 145) afirmam que “a versão escrita da fala em legendas é quase sempre uma versão reduzida do texto de partida oral⁶”, sendo isto devido às diversas restrições espaço-temporais que são pertinentes à LSE. Tentar transcrever exatamente, contudo, todas as falas é uma característica levada à cabo pela LSE praticada pela TV do Brasil e abalizada pelas políticas reguladoras da prática, erro de concepção de tradução que acaba por ocasionar outros erros linguísticos e técnicos nesse tipo de legendagem.

Voltando à redução do texto legendado, esta pode ocorrer de forma parcial ou total. Enquanto que a redução parcial é alcançada através de condensação e de uma versão mais concisa do texto oral de partida, a redução total surge a partir da omissão e apagamento de itens lexicais (DIAZ-CINTAS; REMAEL, 2007, p. 146). Normalmente, uma junção dessas duas estratégias linguísticas é o que leva a essa reescrita considerada inerente à prática da legendagem (ibid., p. 146). A opção em pôr em prática esses processos advém das limitações temporais e espaciais da legendagem. Ou seja, após verificar o tempo disponível para se inserir o texto legendado e observar que algum tipo de redução de texto é necessário, o legendista então procede a: “Eliminar o que não é relevante para a compreensão da mensagem e/ou reformular o que é relevante de uma forma tão concisa quanto possível ou requerido⁷” (ibid., p. 146).

A condensação ou reformulação – doravante apenas ‘condensação’, que funciona como redução parcial do texto legendado, depende das necessidades do que “pode” ser feito e o que “deve” ser feito, sendo essencial que essas reformulações sejam idiomáticas, ou seja,

⁶ Minha Tradução de: “The written version of speech in subtitles is nearly always a reduced form of the oral ST”.

⁷ Minha tradução de: “Eliminate what is not relevant for the comprehension of the message, and/or reformulate what is relevant in as concise a form as is possible or required”.

que soem naturais na língua de chegada do texto traduzido (DIAZ-CINTAS; REMAEL, 2007, p. 150). Essa condensação pode acontecer tanto no nível da palavra quanto no nível da oração. No nível da palavra, por sua vez, Diaz-Cintas e Remael (2007, p. 151-153) citam as seguintes estratégias para a economia de espaços na legenda: 1) simplificar perífrases verbais; 2) generalizar enumerações; 3) utilizar um sinônimo próximo menor ou expressão equivalente; 3) utilizar tempos verbais simples ao invés de compostos; 4) mudar a classes das palavras (transformar um verbo em um substantivo, por exemplo, pode oferecer formas menores em termos de caracteres); 5) e formas menores e contrações⁸ (no caso do inglês, por exemplo, em formas como *I'd*). No nível da oração, por sua vez, os autores citam (ibid., p. 154-161) estratégias como: 1) transformar negações ou perguntas em asserções ou orações afirmativas, perguntas indiretas em perguntas diretas, etc.; 2) simplificar modalizadores; 3) mudar o sujeito de uma oração ou frase; 4) manipular a ordem de tema⁹ (informação conhecida) e rema (nova informação); 5) transformar orações longas e/ou compostas em orações simples; 6) transformar a voz ativa em voz passiva ou vice versa; 7) utilizar pronomes (demonstrativos, pessoais, possessivos) e outros dêiticos para substituir substantivos ou sintagmas nominais e 8) fundir duas ou mais frases/orações em uma.

A omissão – também chamada apenas de ‘redução’, por sua vez, é algo inevitável em legendagem, trabalhando de forma conjunta à condensação, apesar de, algumas vezes, condensações serem mais efetivas do que omissões completas (ibid., 162). Normalmente, a redundância ou relevância irá pautar a escolha do legendista em omitir algum item lexical ou não, entretanto, a decisão por omitir o menos possível deve ser posta em prática (DIAZ-CINTAS; REMAEL, 2007, p. 162-163). As omissões também podem ocorrer no nível da palavra e da oração e são reguladas pela língua alvo da tradução. No nível da palavra, por um lado, modificadores, adjetivos e advérbios – por “apenas” modificarem a informação transmitida pelo verbo ou substantivo, palavras de função fática, elementos interpessoais que estabelecem relações de poder entre os interlocutores (cumprimentos, interjeições, vocativos, expressões de cortesia) e algumas repetições podem ser retiradas do texto traduzido, sendo a

⁸ No caso específico da LSE no Brasil, nossa experiência (Grupo LEAD-UECE) em pesquisas de recepção têm sugerido a preferência pelo português mais formal (considerado padrão) por parte dos surdos. Dessa forma, a utilização por formas abreviadas do português na LSE, o que, geralmente, caracteriza um português mais informal, não se aplica ao caso específico do público alvo da LSE no Brasil.

⁹ Essa conceituação de Tema e Rema difere da concepção de outras correntes teóricas. A Linguística Sistêmico-Funcional (HALLIDAY, 1994), por exemplo, vê o Tema como ponto de partida da mensagem de uma oração, ou seja, o primeiro constituinte da oração, enquanto que o Rema é o desenvolvimento do Tema. Para a Linguística Sistêmico-Funcional, há na oração tanto a estrutura temática Tema /Rema (oração como mensagem) quanto a estrutura da informação Dado/Novo (oração como informação), ao que penso que Diaz-Cintas e Remael se referiram com ‘informação conhecida’ e ‘nova informação’.

função do legendista refletir sobre o quão importante esses elementos são para que a legendagem consiga ser bem sucedida (DIAZ-CINTAS; REMAEL, 2007, p. 163-165). A omissão no nível da oração, por sua vez, apesar de não ser sempre aconselhável, em determinadas vezes também é inevitável. Isso ocorre, geralmente, em falas que possuem uma carga informacional bem pequena. Casos em que muitas pessoas falam ao mesmo tempo, diálogos em que o tópico de uma pergunta é incorporado pela resposta do próximo interlocutor na sequência, ou ainda orações ou frases com menor conteúdo proposicional e que introduzem um ponto de vista ou argumento, são exemplos de situação em que a omissão pode ocorrer no nível da oração (DIAZ-CINTAS; REMAEL, 2007, p. 166-170).

Apesar de característico ao texto legendado, condensação e redução podem levar a problemas de coerência na legenda. Dessa forma, apesar de haver uma contribuição dos canais auditivo e visual, é necessário que se evite uma baixa coesão textual, e, assim, problemas na coerência textual das legendas (ibid., p.171-172). Segundo Diaz-Cintas e Remael (ibid., p. 171), a coesão intersemiótica que acontece na legendagem diz respeito à maneira segundo a qual ela conecta a trilha sonora e a imagem em tela, utilizando a informação fornecida por esses meios para que se crie um todo linguístico-visual coerente.

Essas noções de coesão e coerência do texto legendado junto à imagem do filme são concepções que estão conectadas à visão de LSE enquanto modalidade de Tradução Audiovisual acessível do presente trabalho, assim como a condensação e redução/omissão supracitadas, inerentes à legendagem. Os aspectos linguísticos relativos à segmentação serão abordados separadamente mais adiante, devido à importância do parâmetro em questão para a problemática anteriormente delineada.

Diaz-Cintas e Remael (2007), no que diz respeito aos parâmetros técnicos da legendagem, os dividem em duas dimensões: espacial e temporal. Entre os parâmetros espaciais, estão: número de linhas da inserção de legenda, posicionamento da legenda em tela, tipo de fonte e número de caracteres por linha. Nos parâmetros considerados temporais, por sua vez, estão os aspectos de marcação e duração das legendas, sincronização, intervalo entre legendas consecutivas, velocidade da legenda e tempo de exposição das legendas. Abaixo, enfatizo os aspectos técnicos mais importantes para o presente trabalho: o número de linhas e o número de caracteres por linha na legenda, posicionamento da legenda em tela, marcação, sincronização, intervalo entre legendas, tempo de exposição das legendas e a velocidade da legenda.

Sobre o número de linhas por legenda, os autores afirmam (ibid., p. 82) que o consenso na literatura da área é que o número máximo de duas linhas por legenda seja

utilizado na legendagem para ouvintes, sendo que a LSE se diferencia por permitir o uso de três ou até mesmo quatro linhas, como no caso do Teletexto, sistema espanhol de LSE. Essa característica também é seguida pelo sistema de legendagem fechada para surdos utilizado no Brasil – o *closed caption*, que faz uso de um alto número de legendas com 3 linhas, como visto em Assis (2013), Araújo e Assis (2014) e Gabriel (2015). Como o Grupo LEAD-UECE advoga a favor de uma LSE que seja pautada pelo uso dos mesmos parâmetros técnicos e linguísticos da legendagem para ouvintes – diferenciando-se apenas pela adição de informações adicionais, e encontrando respaldo nos achados de D'ydewalle et al. (1987) em favor de duas linhas como sendo o número máximo por inserção de legenda, essa característica torna a LSE praticada no Brasil inadequada em relação a esse aspecto técnico específico, dentre outros.

Sobre a distribuição física do texto na legenda, é necessário que haja um equilíbrio no que diz respeito à parte linguística da legenda, já que as questões sintáticas e semânticas promovem uma melhor leitura, assim como questões estritamente visuais também importam, segundo Diaz-Cintas e Remael (2007, p. 86). Geralmente, se uma legenda curta cabe confortavelmente em uma linha, não há a necessidade de se usar duas linhas. Algumas empresas, no cinema e geralmente quando as legendas são centralizadas em tela, por questões estéticas, preferem duas linhas curtas de tamanho igual a uma muito grande de apenas uma linha. Além do mais, legendas muito longas forçam o olho a percorrer uma distância considerável da esquerda para a direita, ou direita para a esquerda, dependendo da língua, de forma particular em telas de cinema (ibid., p. 86).

Estudos sobre velocidade de leitura indicam que quão maior o número de palavras em uma legenda, menor o tempo gasto proporcionalmente para se ler cada uma das palavras (DIAZ-CINTAS; REMAEL, p. 93). Segundo Brondeel (1994, p. 28, *apud* DIAZ-CINTAS; REMAL, 2007, p. 93), a latência média na percepção aparentemente faz com que as legendas de duas linhas demandem menos esforço por parte do espectador, dado que duas legendas de uma linha seguidas, por exemplo, requerem dois *onsets* (começos); conseqüentemente, legendas de duas linhas parecem proporcionar ao leitor um conforto maior de leitura. Autores como Lomheim (1999, p. 192, *apud* DIAZ-CINTAS, REMAEL, 2007, p. 93), contudo, preferem o uso de legendas de uma linha por acreditar que estas são mais fáceis de serem lidas. O presente trabalho assume a preferência por legendas de duas linhas sempre que possível, em consonância com os dados encontrados por D'ydewalle e De Bruycker (2003) por meio de pesquisas experimentais com a utilização de rastreamento ocular. A questão da utilização de duas linhas ao invés de uma é importante para a pesquisa em questão, pois o

parâmetro linguístico a ser analisado nessa pesquisa – segmentação linguística entre linhas de legenda – acontece na quebra entre linhas de legenda.

O número de caracteres máximo por linha de legenda varia de acordo com o meio em que a legenda é veiculada. De acordo com Diaz-Cintas e Remael (2007, p. 84), o número máximo utilizado na legendagem de TV normalmente é 37, no caso de línguas que utilizam o alfabeto romanizado, o que incluiria os espaços e os sinais tipográficos. Dependendo da demanda do cliente, do prazo e do software utilizado, há casos em que o máximo pode variar de 33 à 39/41. Há, ainda, casos de festivais de filmes que permitem até 43 caracteres por linha, enquanto que para DVD e cinema o número de máximo de 40 caracteres parece ser o padrão. Apesar de não haver um número mínimo de caracteres para uma linha de legenda, os autores salientam que legendas com menos de 4 ou 5 caracteres são raras (p. 85). Nesse trabalho, assim como pelo Grupo LEAD-UECE como um todo, o número máximo de 37 caracteres por linha é adotado como padrão.

Sobre o posicionamento das legendas em tela, elas normalmente são alinhadas à esquerda ou centralizadas na parte de baixo da tela (um caso à parte, por exemplo, são as legendas de fãs de desenhos japoneses, que utilizam toda a extensão da tela para apresentar a legenda). A opção por centralizar a legenda na parte inferior da tela sempre foi utilizada no cinema e quase sempre no DVD, enquanto que na TV há uma variação entre centro e esquerda (DIAZ-CINTAS; REMAEL, p. 87-88). No Brasil, há, normalmente, a presença de ambos na legenda para surdos e ensurdecidos da TV. Enquanto que em filmes há uma tendência, pelo menos no caso da Rede Globo, de a legenda acompanhar o personagem que detém o turno de fala em tela (se o personagem está à esquerda, a legenda se move para a esquerda, se está à direita, da mesma forma), no caso das telenovelas, jornais e documentários, *sub-corpora* analisados no escopo do Projeto CORSEL, a legenda está normalmente alinhada à esquerda. Segundo Diaz-Cintas e Remael (2007, p. 88), uma razão para se adotar a centralização das legendas na tela tem a ver com o fato de a ação acontecer geralmente no meio da tela; dessa forma, o olho teria que percorrer um percurso menor da imagem para o texto da legenda se as legendas estivessem posicionadas no centro. Por esta razão, e também por ser, comumente, o posicionamento adotado na legendagem para ouvintes, a centralização da legenda será o posicionamento padrão adotado no presente trabalho, o que vem de acordo com a prática realizada no âmbito do Grupo LEAD-UECE, como tem sido frisado.

Outros aspectos técnicos, a marcação e a sincronização das legendas também são fatores importantes para conferir qualidade à legendagem. A marcação, por um lado, é

definida pelo momento em que uma legenda aparece e desaparece da tela (DIAZ-CINTAS; REMAEL, 2007, p. 88). A marcação deve levar em consideração o ritmo do filme e dos atores, as pausas e outras características prosódicas da fala original, e, se possível, a legenda deve aparecer no momento em que o personagem começa a falar e desaparecer no momento em que ele termina sua fala (DIAZ-CINTAS; REMAEL, 2007, p. 88-89). A sincronização, por sua vez, relaciona-se com a marcação das legendas, e diz respeito ao sincronismo necessário entre o som original do filme (as falas e efeitos sonoros que devem ser legendados, estes no caso da LSE), crucial para a uma boa legendagem, dado que ela é responsável por reforçar a coerência interna do produto audiovisual traduzido e auxilia o telespectador a identificar quem está falando (ibid., p. 90). Apesar da busca pelo sincronismo perfeito, devido a questões como a densidade semântica do diálogo original, em que é difícil condensar ou deletar alguma informação sem comprometer a mensagem como um todo, uma quantidade de assincronismo é permitida, sendo, assim, possível que a legenda apareça alguns momentos antes da fala e saia um pouco depois que a fala é terminada (ibid., p. 90-91). Essa ausência de sincronismo é utilizada com mais frequência na LSE em que a marcação deve se preocupar mais com a imagem do que com a trilha sonora em si, sendo que, em se tratando de legendagem para ouvintes, certa contenção é aconselhada (ibid., p. 90).

Em relação ao intervalo entre legendas, aspecto técnico relacionado ao sincronismo e marcação das legendas, não é aconselhado que uma inserção de legenda seja seguida imediatamente por outra sem que haja um tempo de intervalo entre as duas, de modo que os olhos podem encontrar dificuldade em perceber que uma nova legenda foi inserida (ibid., p. 92). De modo que isso não seja um problema, programas de legendagem possuem uma função que cria esse tempo mínimo de espaço após marcação final de uma legenda para que a outra já possa ser inserida respeitando esse intervalo (ibid., p. 92), o que será visto mais adiante no software de legendagem utilizado neste trabalho, o *Subtitle Workshop* 6.0B. Esse tempo mínimo pode ser estabelecido, mas um mínimo de 2 a 3 *frames* de intervalo é necessário, cerca de 80 a 120 milissegundos em uma taxa de 25 *frames* por segundo. Para esse trabalho, assim como para o grupo LEAD-UECE como um todo, um padrão de 100 milissegundos de intervalo mínimo entre legendas é utilizado.

Alguns estudos indicam que quando uma legenda fica em tela mais do que o tempo necessário para ser lida, o espectador tem a tendência de relê-la (ibid., p. 89). D'ydwalle et al. (1987), a partir de experimentos com rastreador ocular, estabeleceram um tempo mínimo e máximo em que uma legenda deve ficar em tela sem prejudicar a leitura pelos espectadores (1 e 6 segundos, respectivamente). Desse modo, sendo o limite do que

pode ser escrito em uma inserção de legenda de duas linhas com cerca de 37 caracteres (74 caracteres, ao todo) ou mais – diferentes mídias, espaços de exibição e softwares utilizados permitem o uso de até 43 caracteres por linha de legenda, como enfatiza Diaz-Cintas e Remal (2007, p. 84), não é recomendado que se utilize uma porção de legenda com essa quantidade máxima de caracteres por mais de 6 segundos, o que poderia levar à releitura (DIAZ-CINTAS; REMAEL, 2007, p. 89). Em situações assim, há a necessidade de se quebrar a legenda em unidades menores de pensamento, de preferência onde a lógica da sentença permite.

Assim como para evitar que haja releituras da legenda o tempo de exposição máximo de 6 segundos é utilizado, o tempo mínimo de 1 segundo é comumente utilizado para que legenda não ‘pisque’ apenas em tela (ibid., p. 89-90), de modo que os espectadores tenham tempo suficiente de ler o conteúdo em tela. Quando a fala do personagem dura menos do que um segundo, Diaz-Cintas e Remal (ibid., p. 90) apontam duas estratégias possíveis: se houver outra pessoa falando imediatamente após ou antes, deve ser analisado a possibilidade de apresentar as duas falas em forma de diálogo em uma legenda; no caso em que a fala é seguida e precedida por pausas, outra opção seria fazer uso de um assincronismo no começo e no fim da fala do personagem, ou seja, a legenda não iria ser apresentada em tela de acordo com o começo e/ou fim da fala do personagem.

O Grupo LEAD-UECE, em se tratando de LSE, utiliza como tempo máximo de exposição em tela 4 segundos, devido ao fato de ser um tempo padrão utilizado pelas empresas de legendagem no Brasil, e como tempo mínimo 1 segundo, o mesmo seguido comumente pela prática da legendagem. Esses tempos foram os utilizados para a proposta de relegendagem realizada como parte do capítulo de resultados.

Baseados na regra dos 6 segundos formulada por D’ydwalle et al. (1987), Diaz-Cintas e Remael (2007) propuseram, a partir da utilização do programa *Wincaps* (software de legendagem), tabelas que relacionam caracteres por segundo (cps) e *frames* – quadros em sequência que são responsáveis por estabelecer a sensação de movimento na gravação videográfica – de forma que os legendistas fossem instrumentalizados a fazer uso das três velocidades normalmente utilizadas na indústria da legendagem, de 145, 160 e 180 palavras por minuto (ppm).

Abaixo, as três tabelas em questão.

Tabela 1 – Velocidade de leitura da legenda a 145ppm em cps

145 palavras por minuto	Segundos: Frames	Caracteres	Segundos: Frames	Caracteres	
	01:00	16	02:00	29	
	01:04	17	02:04	32	
	01:08	18	02:08	34	
	01:12	20	02:12	36	
	01:16	23	02:16	38	
	01:20	25	02:20	40	
Segundos: Frames	Caracteres	Segundos: frames	Caracteres	Segundos: Frames	Caracteres
03:00	44	04:00	58	05:00	71
03:04	46	04:04	60	05:04	71
03:08	48	04:08	62	05:08	73
03:12	50	04:12	64	05:12	73
03:16	52	04:16	65	05:16	74
03:20	54	04:20	67	05:20	74

Fonte: Gabriel (2015, p. 26) adaptada de Diaz-Cintas e Remael (2007, p. 97)

Tabela 2 – Velocidade de leitura da legenda a 160ppm em cps

160 palavras por minuto	Segundos: Frames	Caracteres	Segundos: Frames	Caracteres	
	01:00	17	02:00	31	
	01:04	18	02:04	34	
	01:08	20	02:08	37	
	01:12	23	02:12	40	
	01:16	26	02:16	42	
	01:20	28	02:20	44	
Segundos: Frames	Caracteres	Segundos: Frames	Caracteres	Segundos: Frames	Caracteres
03:00	48	04:00	63	05:00	75
03:04	50	04:04	65	05:04	75
03:08	53	04:08	67	05:08	76
03:12	56	04:12	69	05:12	76
03:16	58	04:16	71	05:16	77
03:20	60	04:20	73	05:20	77
				06:00	78

Fonte: Gabriel (2015, p. 26) adaptada de Diaz-Cintas e Remael (2007, p. 98)

Tabela 3 – Velocidade de leitura da legenda a 180ppm em cps

180 palavras por minuto	Segundos: Frames	Caracteres	Segundos: Frames	Caracteres	
	01:00	17	02:00	35	
	01:04	20	02:04	37	
	01:08	23	02:08	39	
	01:12	26	02:12	43	
	01:16	28	02:16	45	
01:20	30	02:20	49		
Segundos: Frames	Caracteres	Segundos: Frames	Caracteres	Segundos: Frames	Caracteres
03:00	53	04:00	70	05:00	78
03:04	55	04:04	73	05:04	78
03:08	57	04:08	76	05:08	78
03:12	62	04:12	76	05:12	78
03:16	65	04:16	77	05:16	78
03:20	68	04:20	77	05:20	78
				06:00	78

Fonte: Gabriel (2015, p. 27) adaptada de Diaz-Cintas e Remael (2007, p. 99)

A velocidade da legenda depende da velocidade de fala dos personagens em vídeo, assim como da capacidade, digamos assim, dos leitores de lerem essa legenda de maneira confortável ou não (DIAZ-CINTAS; REMAEL, 2007, p. 95), o que acaba por estabelecer uma relação entre a velocidade de leitura do público alvo da legenda e a velocidade da legenda. Segundo Ivarsson e Carroll, “a velocidade leitora difere grandemente entre o público ouvinte e isso se aplica a uma extensão ainda maior para os surdos e ensurdecidos¹⁰” (1998, p. 131), fato agravado pelo fato de algumas legendas veiculadas na TV brasileira não apresentarem sequer sincronismo entre imagem e legenda, sendo a atividade de leitura forçosa, cognitivamente falando, para o surdo. Vale ressaltar este fato, pois isso amplifica a responsabilidade de fazermos legendas que minimizem ao máximo esse esforço cognitivo do participante surdo ao receber o produto legendado. Desse modo, uma legenda adequada em termos linguísticos – principalmente no que concerne a segmentação linguística, abordada mais a fundo aqui, paralinguísticos e técnicos, tem mais chances de ser bem recebida, como visto em Nascimento e Araújo (2011) e Araújo (2012), a serem detalhados a seguir.

Ainda sobre a velocidade da legenda e sua relação com a recepção da LSE por surdos, a velocidade de 145ppm era a considerada ideal para a LSE nos produtos legendados por nosso grupo de pesquisa, estando de acordo com os resultados apresentados até então

¹⁰ Minha tradução de: “Viewer reading speeds differ greatly among hearing audiences and this applies to an even greater extent to the deaf and hard of hearing”.

pelas pesquisas de recepção realizadas com o público surdo cearense (ARAÚJO, 2008). Contudo, as pesquisas de recepção de Nascimento e Araújo (2011) e Araújo (2012) sugeriram que, com uma legenda bem segmentada, uma pessoa surda se torna capaz de ler em uma velocidade de até 180ppm, colocando em cheque a primazia do parâmetro da velocidade na compreensão do filme legendado pelos surdos.

Sobre as características que definem a LSE e a diferenciam da legendagem para ouvintes, há a identificação dos falantes e tradução dos efeitos sonoros, além da tradução das características prosódicas da fala. Sobre a identificação dos falantes, Ivarsson e Carroll (1998, p. 131) apontam que o surdo deve ter conhecimento de quem está falando o que em cena, ainda mais quando há vários interlocutores em tela ou quando a câmera não está focalizada em quem detém o turno de fala. Na Europa, a identificação dos falantes é feita por meio do uso de diferentes cores para cada personagem. O modelo adotado em Portugal, por exemplo, faz uso das cores branca (falas) e amarela (identificação dos falantes e efeitos sonoros), além da movimentação da legenda sobre o personagem que detém o turno de fala. No Brasil, por sua vez, há uma preferência pelo uso de colchetes com os nomes dos personagens dentro.

Os efeitos sonoros, da mesma maneira, são relativos à trilha sonora e aos sons naturais do produto audiovisual. A relevância da tradução de efeitos sonoros na LSE é indiscutível. Como consumidores de produtos audiovisuais que inter-relacionam códigos semióticos distintos, no caso, o imagético, o verbal e o sonoro, os surdos devem ser empoderados pela tradução a construir significados e pontos de vista acerca dessas produções a partir da correlação entre esses meios semióticos, sendo isso absolutamente necessário para o sucesso da recepção do produto. Sobre a importância do código semiótico sonoro, Nascimento (2013, p. 16) diz que os “[...] efeitos sonoros devem ser acrescentados à LSE, pois são relevantes para a compreensão da trama, já que muitas vezes contêm elementos que só podem ser compreendidos através de determinado som [...]”. Um exemplo, segundo a autora, seria o som da porta que bate, mas que não aparece na imagem. Sem a presença da tradução desse som na LSE, os surdos não compreenderiam porque os personagens reagem àquele ou à outro efeito sonoro.

Pesquisas de recepção com surdos feitas pelo Grupo LEAD-UECE (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004, 2007, 2008; NASCIMENTO; ARAÚJO, 2011) têm auxiliado na obtenção de parâmetros de LSE que estejam de acordo com as necessidades específicas dos surdos brasileiros. Entre as constatações dessas pesquisas de recepção estão os seguintes parâmetros: i) a cor deve ser, preferencialmente, amarela em um fundo transparente; ii) a identificação de falantes e os efeitos sonoros devem aparecer entre colchetes; iii) a voz

filtrada (voz fora da tela, vinda da televisão, vinda do rádio etc.) deve apresentar-se em itálico; e iv) a linguagem utilizada deve ser o português formal. Esses parâmetros, contudo, não são comumente colocados em prática na LSE que é veiculada pelos canais de TV brasileiros. Esses parâmetros sugeridos pelas pesquisas de recepção supracitadas foram levados em conta na atividade de relegendagem presente no segundo momento do capítulo de resultados.

2.1.1 O *closed caption*

O *closed caption* é um tipo de legendagem fechada ou oculta (que pode ser acessado pelo controle e não é “colada” no vídeo) destinada a surdos e ensurdecidos. É o sistema utilizado no Brasil para a veiculação de LSE pelos canais de TV. Este segue os mesmos princípios do sistema americano, em que há a transformação da legenda em códigos eletrônicos que são inseridos na linha 21 do intervalo vertical em branco da TV, sendo esta uma barra horizontal posta entre as imagens (ARAÚJO, 2008, p. 62). O *closed caption* brasileiro comumente não edita a fala dos interlocutores em tela, sendo uma transcrição dessa fala e, desse modo, estando em dissonância com o que se considera como parâmetro da legendagem para ouvintes e para a LSE segundo a literatura da área, assim como com os parâmetros apontados pelas pesquisas em LSE no Brasil (ARAÚJO, 2008; NASCIMENTO; ARAÚJO, 2011; ARAÚJO, 2012). Essa questão é central em relação à concepção de tradução na produção da LSE na TV do Brasil, tendo respaldo na definição da atividade pela portaria 310 do Ministério das Comunicações acerca dos recursos de acessibilidade na TV brasileira:

3.2. Legenda Oculta: corresponde à transcrição, em língua portuguesa, dos diálogos, efeitos sonoros, sons do ambiente e demais informações que não poderiam ser percebidos ou compreendidos por pessoas com deficiência auditiva.

A meu ver, essa concepção de tradução enquanto transcrição da fala é uma dos responsáveis pela baixa qualidade apresentada por esse tipo de LSE no Brasil.

O *closed caption* no Brasil pode aparecer em tela em dois tipos: como *roll-up* e como *pop-on*. No tipo de legendagem *roll-up*, as palavras são digitadas da esquerda para a direita e rolam de baixo para cima na parte inferior da televisão. Esta é produzida,

normalmente, por um estenotipista¹¹ ou por refalamento (*respeaking*), que funciona através de um software de reconhecimento de voz. A legendagem *roll-up* deveria aparecer somente em programas ao vivo, pois, devido à instantaneidade da fala em tempo real, não haveria a possibilidade de produzir uma legenda editada de acordo com os preceitos técnicos e linguísticos da legendagem. Contudo, isso não acontece sempre no *closed caption* disponibilizado pelos canais de TV, o que vai de encontro com o que defendemos no âmbito do Grupo LEAD-UECE acerca da prática da LSE. Um exemplo desse tipo de utilização de *closed caption* do tipo *roll-up* pode ser visto abaixo, na Figura 1, em imagem extraída do *Programa do Jô*, da Rede Globo.

Figura 1 – Exemplo de legenda *roll-up* no Programa do Jô



Fonte: Chaves (2009, p. 19)

A legendagem *pop-on*, por sua vez, que pode ser feita pelo estenótipo ou por um software específico de legendagem, aparece na tela em bloco e sai em bloco, assim como a legendagem para ouvintes, sendo reservada para programas pré-gravados, em que há tempo hábil para sincronização, identificação dos falantes, legendagem dos sons, etc. (FRANCO; ARAÚJO, 2003). Abaixo, na Figura 2, um exemplo de legenda *pop-on*.

¹¹ Profissional que digita em tempo real em posse de um estenótipo, teclado especial de 24 teclas em que a pressão de várias destas ao mesmo tempo pode formar uma palavra.

Figura 2 – Exemplo de legenda *pop-on* em O Auto da Compadecida



Fonte: Chaves (2009, p. 20)

Este foi o tipo de legendagem utilizada neste trabalho, dado que, supostamente, por ser editada, daria a chance ao legendista de elaborá-la de acordo com os preceitos da literatura da área, dentre eles, o parâmetro da segmentação, foco do trabalho ora relatado.

A LSE de DVD comercial no Brasil apresenta-se sempre no formato *pop-on*. Diferentemente do *closed caption*, que apresenta características técnicas não recomendadas pela literatura em TAV, como um número maior de linhas e velocidade e densidade lexical altas, além de uma concepção equivocada de tradução, A LSE em DVD comercial costuma seguir os mesmos parâmetros técnicos e linguísticos adotados pela legendagem para ouvintes.

2.2 A SEGMENTAÇÃO NA LEGENDAGEM

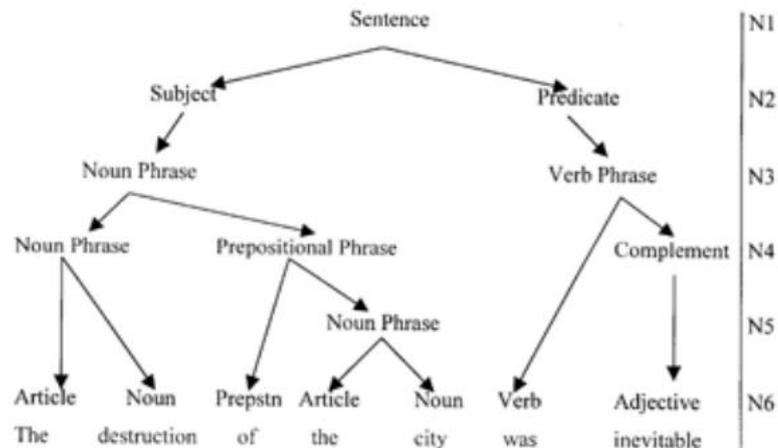
Uma dos parâmetros linguísticos presentes na legendagem e principal objeto de estudo do presente trabalho, a segmentação refere-se à divisão da fala traduzida em porções de texto escrito na parte inferior da tela. Ela pode acontecer de três formas: pautada pela linguística (sintaxe, semântica, lexicogramática e fonologia-fonética), pela retórica (fluxo da fala) e pelo visual (corte ou mudança de cena) (REID, 1990; GOTTLIEB, 1994; KARAMITROGLOU, 1998; IVARSSON; CARROLL, 1998; DIAZ-CINTAS; REMAEL, 2007; CHAVES, 2012; ARAÚJO; ASSIS, 2014). A segmentação linguística¹² pode ainda

¹² Reid (1990) utiliza o termo 'gramatical' ao invés de 'linguística'. Como apontamos em Araújo e Assis (2014, p. 160), além da problematização exposta por Chaves (2012, p. 45), a qual sugeriu que a escolha por 'linguística' deveu-se ao fato de a palavra 'gramatical' ser "arraigada de preconceito e por isso acaba conduzindo à discussão por um viés normativo da língua", no âmbito dos trabalhos sobre o tema vinculados ao Grupo LEAD-UECE, preferimos a utilização do termo 'linguística' porque este é, já que a segmentação não envolve somente a gramática, ou seja, o léxico e a sintaxe, mas também a semântica e fonologia-fonética.

ocorrer de duas maneiras: por meio da divisão das legendas ao longo da legendagem, ou seja, a quebra que ocorre entre a última palavra/sintagma de uma inserção de legenda e a primeira palavra/sintagma da inserção de legenda seguinte, chamada de segmentação; ou ainda entre linhas de uma mesma legenda, chamada de *line break*, ou segmentação quebra de linha (DIAZ CINTAS; REMAEL 2007, p. 173). Este é o tipo de segmentação linguística analisado na pesquisa ora relatada, a qual irei referir-me apenas como ‘segmentação linguística’.

Para que haja uma boa segmentação linguística das legendas, sendo essa segmentação responsável por facilitar o processamento linguístico do produto audiovisual legendado por parte dos espectadores, a divisão dessas legendas deve cumprir diretrizes linguísticas. Ainda sobre a segmentação linguística, Karamitroglou (1998) afirma que esse processo de segmentação deve acontecer no mais alto nível sintático possível. Com isso em mente, o autor exemplifica esse “mais alto nível sintático” através de uma perspectiva gramatical de cunho gerativista, chegando a utilizar uma descrição sintática baseada em árvore, comum ao gerativismo. Abaixo, a Figura 3 apresenta a árvore sintática utilizada para exemplificar a segmentação linguística no mais alto nível sintático.

Figura 3 – Árvore sintática para explicar a Segmentação Linguística



Fonte: Karamitroglou (1998)

Karamitroglou (1998) divide a oração *the destruction of the city was inevitable* (a destruição da cidade foi inevitável) a partir de seis nós: no N1 está a oração constituída de sujeito e predicado (N2), formados, por sua vez, por sintagmas nominais e verbais respectivamente (N3). O sujeito *the destruction of the city* é formado por um sintagma nominal (*the destruction*) e por um sintagma preposicional (*of the city*). O predicado possui um complemento, o qual está localizado no N4 com os sintagmas nominal e preposicional que

formam o sujeito. O sintagma preposicional também possui, dentre seus constituintes, um sintagma nominal (*the city*), localizado no N5. Por fim, temos as classes de palavras dos itens lexicais que formam a estrutura gramatical, formando o N6. Para Karamitroglou (1998), sendo essa oração uma legenda e tendo que ser segmentada em duas linhas, essa segmentação deveria privilegiar as unidades *The destruction of the city* (sujeito) na linha de cima e *was inevitable* (predicado) na linha de baixo idealmente. Dessa maneira, a segmentação aconteceria no N2, o mais alto nível sintático possível, e, quando isso ocorre, “[...] maior é o agrupamento de carga semântica e mais completa é a informação a ser apresentada ao cérebro” (1998, p. 9), facilitando, assim, a compreensão da legenda como um todo (ARAÚJO; ASSIS, 2014, p. 161).

Contudo, alguns legendistas preferem segmentar pelo número de caracteres, em detrimento das questões linguísticas apontadas por Karamitroglou (1998) para a distribuição das legendas em tela. Essa segmentação baseada em uma espécie “geometria” da legenda produz três formatos: i) um semelhante ao de um retângulo, em que cada linha de legenda apresenta quase o mesmo número de caracteres nas duas linhas e ii) dois formatos que lembram um triângulo, em que há mais caracteres na linha de cima ou na de baixo (ARAÚJO; ASSIS, 2014, p.161-162).

O Quadro 1 mostra os três formatos:

Quadro 1 – Formato das legendas na tela

Formato	Legendas
Em forma de retângulo	O guardinha me parou por causa de uma bobagem da placa que caiu!
Em forma de triângulo com a linha de cima maior	Um tutuzinho de feijão, um lombinho.
Em forma de triângulo com a linha de cima menor	[Deolinda] já imaginava, por isso fiz o tutuzinho logo hoje.

Fonte: Araújo e Assis (2014, p. 162)

Karamitroglou (1998, p. 10) postula que, ao invés do formato triangular, o ideal seria que a legenda de duas linhas, a qual normalmente tem até 37 caracteres em cada linha na legenda veiculada pela TV, tivesse um formato retangular, já que este é o formato ao qual os telespectadores estão mais acostumados. Entretanto, os triangulares também podem ser acessados quando as questões linguísticas que envolvem uma boa segmentação linguística não permitem esse formato.

2.3. LEGENDAGEM E LINGUÍSTICA DE *CORPUS*

A Linguística de *Corpus* é o ramo da Linguística que trata da análise eletrônica de linguagem natural através de computadores, coletando e explorando *corpora*, ou conjunto de dados linguísticos textuais criteriosamente armazenados, para o estudo de uma língua ou de uma variedade linguística (BERBER SARDINHA, 2004, p. 3). O uso da Linguística de *Corpus* pelos Estudos da Tradução remonta ao trabalho de Baker (1993), pioneiro na junção da metodologia que a Linguística de *Corpus* propicia à área de estudos em questão. Em relação à Tradução Audiovisual, por sua vez, a análise baseada em corpus tem ganhado espaço e respaldo entre os pesquisadores da área. Entretanto, ainda é pequeno o arcabouço teórico-metodológico formado pela união dessas duas vertentes. No que diz respeito à interface LSE/Linguística de *Corpus*, a partir dos objetivos do presente trabalho, irei detalhar, a seguir, as pesquisas que embasam o estudo aqui relatado em algum ponto. A saber, são: Perego (2008), Diniz (2012), Chaves (2012), Assis (2013), Araújo e Assis (2014), Arraes (2015) e Gabriel (2015).

Perego (2008) investigou problemas de segmentação linguística entre linhas (*line breaks*) na legendagem para ouvintes de um *corpus* heterogêneo de legendas de filmes para DVD e para cinema. A autora definiu categorias para a análise desses problemas, a saber: quebra do sintagma nominal, quebra do sintagma preposicional, quebra do sintagma verbal, quebra das orações coordenada e subordinada. Apesar de fazer uso de uma análise manual, diferentemente da análise semiautomática baseada em *corpus* proposta pelo presente trabalho, Perego (2008) serviu como referencial para a elaboração das etiquetas discursivas para a análise da segmentação em LSE por Chaves (2012, p. 29), sendo este o trabalho que serviu como base metodológica para o presente trabalho e para o projeto CORSEL como um todo.

Diniz (2012) analisou o fenômeno da segmentação na legendagem para surdos e para ouvintes do filme brasileiro *Irmãos de Fé*. Com esse intuito, a autora utilizou Lemle (1989), ou seja, uma concepção gerativista de sintagmas para embasar sua descrição linguística dos casos de segmentação gramatical. A leitura e interpretação da obra de Lemle resultou por nomear como ‘desvios’ os casos em que o texto legendado segmentado apresentava alguma ruptura indevida dos sintagmas localizados entre os intervalos de linha. Os resultados de Diniz apontaram que não havia diferença na segmentação entre linhas na legendagem para ouvintes e na LSE do filme. O desvio mais recorrente em seu *corpus* foi no nível do sintagma verbal. Outra consideração interessante advinda do estudo em questão foi o fato de a maior parte das legendas que sofreram desvios linguísticos quando segmentadas

terem sido motivadas por aspectos estéticos e geométricos, e não linguísticos, durante o processo de legendagem.

De forma concomitante a Diniz (2012), com o objetivo de analisar os problemas de segmentação no filme brasileiro *Nosso Lar*, Chaves (2012) propôs etiquetas para a análise de aspectos técnicos e, principalmente, de aspectos relativos à segmentação linguística utilizando os pressupostos teóricos de Karamitroglou (1998) expostos na subseção anterior acerca da necessidade da segmentação acontecer no mais alto nível sintático possível. As etiquetas presentes em Chaves (2012) foram inspiradas na pesquisa de Perego (2008). No que diz respeito à perspectiva sintática utilizada pelo trabalho em questão, a autora fez uso da concepção descritiva de Perini (2010) acerca do português brasileiro. Sobre as etiquetas criadas pela autora, 3 delas indicavam os tipos de problemas de segmentação (linguística, retórica e visual), e 8 indicavam os parâmetros técnicos da LSE (número de linhas e caracteres, velocidade, tempo inicial e final da legenda). A partir dos conceitos, nomenclaturas e observações de Perini (2010), a autora formulou 19 subetiquetas para categorizar problemas de segmentação linguística. Os resultados elencados por Chaves (2012) mostraram uma maior quantidade de problemas de segmentação linguística no nível dos sintagmas verbal, especificamente na quebra verbo+verbo, e nominal (41% e 26%, respectivamente), e em legendas com velocidade considerada média e alta (160 e 180 palavras por minuto).

O trabalho de Chaves (2012) foi de grande importância para a pesquisa desenvolvida por mim em Assis (2013) por suas metodologia e etiquetas discursivas voltadas para a análise da segmentação, sendo assim importante também para a fundamentação da presente pesquisa e para o desenvolvimento do projeto CORSEL. Abaixo, o Quadro 2 referente às etiquetas de características e problemas de segmentação categorizadas por Chaves (2012).

Quadro 2 – Etiquetas de problemas de segmentação de Chaves

ETIQUETAS DE ANÁLISE DE PARÂMETROS TÉCNICOS DA LEGENDAGEM	
Número da legenda	<sub1>legenda 1</sub1>
Linhas por legenda	<1L>, <2L>
Tempos inicial e final de cada legenda	<t>início --> final</t>
Número de caracteres por linha (aplicada em legendas de 2 linhas)	<cpl>
Velocidade da legenda baixa (até 13cps).	<velocidade da legenda_baixa>
Velocidade ideal (14 a 15cps).	<velocidade da legenda_ideal>
Velocidade alta (a partir de 16cps)	<velocidade da legenda_alta>
ETIQUETA INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO GRAMATICAL	
<PROSEGG>	
ETIQUETA INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO RETÓRICA	
<PROSEGR_antecipouinformação>	
<PROSEGR_atrasouinformação>	
ETIQUETA INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO VISUAL	
<PROSEGV_vazou>	
ETIQUETAS DE ANÁLISE DE SINTAGMA NOMINAL (SN)	
<SN_pre-nucleares+subst>	
<SN_nominal+modif/modif+nominal>	
<SN_superlativo+adj>	
<SN_relativo+oração_incompleta>	
<SN_nome próprio>	
<SN_título+nome próprio>	
<SN_colocações/idiom/conv>	
ETIQUETAS DE ANÁLISE DE SINTAGMA PREPOSICIONADO (SP)	
<SP_prep+subst>	
ETIQUETAS DE ANÁLISE DE SINTAGMA VERBAL (SV)	
<SV_verbo+verbo>	
<SV_verbo+adv>	
<SV_colocações>	
<SV_negação+verbo>	
<SV_(verbo)+oblíquo+verbo>	
ETIQUETAS DE ANÁLISE DE SINTAGMA ADVERBIAL (SAdv)	
<SAdv>	
ETIQUETAS DE ANÁLISE DE SINTAGMA ADJETIVO (SAdj)	
<SAdj_subst+adj>	
ETIQUETAS DE ANÁLISE DE ORAÇÃO COORDENADA (COORD)	
<COORD_coordenador+oração>	
<COORD_negativa>	
ETIQUETAS DE ANÁLISE DE ORAÇÃO SUBORDINADA (SUBORD)	
<SUBORD_conj+oração>	
<SUBORD_se>	

Fonte: Chaves (2012, p. 62)

Em Assis (2013) e Araújo e Assis (2014), tivemos como objetivo descrever e analisar como acontecia a segmentação linguística na legendagem *closed caption pop-on* da

TV brasileira, especificamente em 1 capítulo da telenovela brasileira *Amor Eterno Amor*, *corpus* pertencente ao projeto CORSEL e extraído automaticamente do vídeo com o programa *CCExtractor*. A metodologia do trabalho envolveu uma dimensão descritiva pautada por análises quanti-qualitativas baseadas em *corpus* dos problemas relativos à segmentação linguística do objeto de estudo. Estes foram identificados a partir de etiquetas discursivas próprias para esse tipo de análise, criadas com base no trabalho de Chaves (2012) e na gramática de cunho funcionalista de Castilho (2012) sobre o Português Brasileiro. Após o processo de anotação, o *corpus* foi analisado com o auxílio do programa *Wordsmith Tools 5.0*. Os resultados da pesquisa apontaram uma grande quantidade de problemas de segmentação no *corpus* analisado, cerca de 25,5% do total de legendas. Os problemas apareceram com mais frequência nos níveis dos sintagmas verbal, principalmente, assim como em legendas de 3 linhas e com velocidade considerada alta.

Conforme apontado em Araújo e Assis (2014, p. 164-166), embasados pelas etiquetas de Chaves (2012) e pela pressuposição de que “[...] mesmo a ação mais simples imaginável, a de contar palavras ou identificar a pontuação pressupõe uma teoria linguística” (GREFENSTETTE; TAPANAINEN, 1994; NUNBERG, 1990, *apud* SANTOS, p. 58), passamos, entre os pesquisadores do Grupo LEAD, no âmbito do Projeto CORSEL, a pautar a análise da segmentação linguística por um viés funcionalista, devido ao fato de que essa perspectiva linguística conjuga epistemologicamente com a metodologia baseada em *corpus*. Nesse sentido, a Nova Gramática do Português Brasileiro (CASTILHO, 2012) – por ser menos sintética¹³ que a descrição de Perini (2010) e devido a sua abundância de detalhes no que diz respeito à descrição dos sintagmas e das sentenças¹⁴ complexas¹⁵ da língua portuguesa

¹³ Sendo menos sintética, permitiu-nos compreender de forma mais substancial os problemas de segmentação linguística (ARAÚJO; ASSIS, 2014, p. 165)

¹⁴ Conforme destacado por Araújo e Assis (2014, p. 165), a “designação da sentença não é pacífica na literatura” (CASTILHO, 2012, p. 58), sendo possível encontrar “[...] termos tais como oração, frase, período (conjunto de orações) etc.” (CASTILHO, 2012, p. 58). Uma definição de ‘sentença’ no Glossário da Nova Gramática do Português Brasileiro diz que: “1. Sentença ou oração é a unidade da sintaxe estruturada por um verbo que seleciona o seu sujeito e seus complementos. Os adjuntos também integram uma sentença, mas não são selecionados pelo verbo [...]” (p. 691). Dessa maneira, o autor, aparentemente, considera os dois termos como intercambiáveis, apesar de preferir utilizar ‘sentença’. Esta pressuposição pode ser confirmada através da definição do vocábulo ‘oração’ na mesma obra, em que há uma remissão ao vocábulo ‘sentença’ expressa por ‘Veja Sentença’ (p. 686). Em minha análise, assim como no Projeto CORSEL, preferi fazer uso do termo ‘oração’, pois, sendo ‘sentença’ e ‘oração’ termos intercambiáveis, a utilização deste último se torna menos obscura. O termo ‘sentença’, inclusive, nem mesmo se encontra na ‘Nomenclatura Gramatical Brasileira’, ao contrário do que acontece com ‘oração’.

¹⁵ Da mesma forma anteriormente destacado por Araújo e Assis (2014, p. 165), Castilho (2012), no capítulo intitulado ‘A Sentença Complexa e sua Tipologia’, ressalta que o termo ‘sentença complexa’ é preferível à ‘período’, devido ao fato de este não apresentar uma unidade sintática diferente da sentença simples, “ou seja, tudo o que ocorre numa sentença simples ocorre numa sentença complexa”(p. 336). Para efeitos de análise, não havendo uma diferença além de terminológica no que diz respeito a esses dois termos, irei utilizar os termos ‘oração coordenada’ e ‘oração subordinada’ (períodos compostos) pelo mesmo motivo supracitado.

em sua variante brasileira – foi adotada como referencial teórico para os objetivos da pesquisa ora relatada, assim como em relação ao Projeto CORSEL como um todo. Segundo Castilho (2012, p. 249), “a estrutura sintática da sentença fundamenta-se nos arranjos lexicais de que ela é formada, os sintagmas, bem como nas funções que decorrem do relacionamento desses sintagmas”, sendo os sintagmas um somatório de constituintes que apresentam um lugar previsível na estrutura da sentença. Conforme Castilho (2012), o sintagma apresenta uma estrutura composta por: i) NÚCLEO, ii) uma margem esquerda, preenchida ou não pelos ESPECIFICADORES do sintagma e iii) uma margem direita, preenchida ou não pelos COMPLEMENTADORES. No Quadro 3, presente em Araújo e Assis (2014, p. 165-166), há, de forma resumida, os tipos de sintagma conforme a descrição de Castilho, assim como a estrutura geral de cada um deles. Do mesmo modo, há a descrição acerca dos dois tipos de sentença mencionados.

Quadro 3 – Estrutura geral dos sintagmas e sentenças complexas

Sintagma Verbal (SV)	O SV é a construção que tem como núcleo o verbo. Como a sentença é um verbo que articula seus argumentos, a única diferença entre ele e uma sentença é que no SV não figura o sujeito.
Sintagma Nominal (SN)	O SN é uma construção cuja estrutura tem por núcleo um substantivo ou um pronome, tendo por Especificador o artigo e os pronomes e por Complementadores os SAdjs e os SPs.
Sintagma Adjetival (SAdj)	O SAdj tem por Núcleo o adjetivo, que é uma classe basicamente predicadora, funcionando como adjunto adnominal, enquanto constituinte do sintagma nominal, ou como predicativo, enquanto constituinte do sintagma verbal (CASTILHO, 2012, p. 516). O SAdj tem como Especificador advérbios predicativos qualificadores e como Complementador SPs e sentenças substantivas objetivas.
Sintagma Preposicional (SP)	O SP tem por núcleo a preposição, por Especificador o advérbio, enquanto que o Complementador em um SP também pode ser: (i) um substantivo, como em ‘para casa’; (ii) um verbo, como em ‘para comer’; (iii) um pronome, como em ‘para mim’, ‘para quem’; (iv) um quantificador definido, como em ‘para dois’; (v) um quantificador indefinido, como em ‘para muitos’, dentre outras possibilidades de realização.
Sintagma Adverbial (SAdv)	O SAdv tem como núcleo o advérbio e como Especificador e Complementador outros advérbios. Sintaticamente, Castilho (2012) diz que os advérbios mantêm relação entre si, com os adjetivos e os verbos.
Sentenças Complexas	Enunciados que possuem mais de um verbo, ou seja, contém mais de uma sentença, podendo estabelecer relação de coordenação, subordinação e correlação entre si.

Fonte: Castilho (2012, *apud* ARAÚJO; ASSIS, 2014, p. 165-166)

Após a realização dessa releitura a partir de Castilho (2012), algumas etiquetas formuladas por Chaves (2012) permaneceram intactas, diferentemente de outras, que foram

reelaboradas em termos de nomenclatura. Houve, ainda, casos em que diferentes etiquetas foram aglutinadas em apenas uma de modo a facilitar a operacionalização da etiquetagem semiautomática; outras surgiram ainda pelos casos inéditos¹⁶ encontrados no *corpus* do presente estudo.

Outras duas pesquisas realizadas no âmbito do projeto CORSEL, e, dessa maneira, relevantes para o presente estudo, foram Arraes (2015) e Gabriel (2015).

Arraes (2015) teve como objetivo analisar a segmentação linguística na LSE do filme *Virada Radical* (2006), transmitido pela Rede Globo em junho de 2014 através do programa *Sessão da Tarde*. Utilizando uma metodologia quanti-qualitativa com o auxílio do programa *Wordsmith Tools 5.0* para a análise baseada em *corpus* de etiquetas discursivas próprias para a análise da segmentação linguística em legendas, o trabalho em questão difere da pesquisa ora relatada pelo meio pelo qual o autor obteve o seu *corpus*: ao invés de utilizar algum programa de extração automática das legendas, Arraes filmou a tela de uma TV para capturar o filme e suas legendas *closed caption* para, então, transcrevê-las e sincronizá-las através do programa de legendagem *Subtitle Workshop 2.51*. Os resultados mostraram um número baixo de problemas de segmentação, cerca de apenas 5,41% das legendas apresentaram algum tipo de problema, predominantemente nos sintagmas verbal (70,6%) e nominal (14,2%) e nas orações subordinadas (10,8%). A maioria desses problemas ocorreu em legendas de velocidade considerada alta (56,5%).

Gabriel (2015), por sua vez, é um trabalho que apresenta muitas características semelhantes com a proposta de análise quanti-qualitativa de segmentação linguística em LSE de uma telenovela realizada pela pesquisa ora relatada. A autora realizou a análise em questão da LSE da telenovela *Cheias de Charme*, *corpus* integrante do Projeto CORSEL, utilizando os mesmos pressupostos teórico-metodológicos adotados pelo projeto e também adotados aqui na análise descritiva de *Amor Eterno Amor*. Os resultados alcançados pela autora mostraram, assim como em Diniz (2012), Chaves (2012), Assis (2013), Araújo e Assis (2014) e Arraes (2015), um maior número de ocorrências de problemas de segmentação linguística no nível do sintagma verbal (35%). Em segundo lugar, bem próximo ao sintagma verbal, ficou o sintagma nominal (33%), o que reforça a relevância desses dois sintagmas no que diz respeito aos problemas de segmentação linguística ocorridas nos gêneros de ficção telenovela e filme. Assim como em Assis (2013) e Araújo e Assis (2014), a maior parte desses

¹⁶ Chaves (2012) apenas criou etiquetas referentes aos problemas de segmentação encontrados em seus corpus. Aqueles casos que não apresentaram ocorrências no corpus, dessa maneira, não foram categorizados no quadro sinótico de etiquetas de quebra de segmentação linguística formulado pela autora.

problemas de segmentação linguística ocorreu em legendas de 3 linhas (53%) e com velocidade de 180ppm, considerada alta (54,5%).

As pesquisas mencionadas até aqui influenciaram o presente estudo e são importantes no que diz respeito às questões teórico-metodológicas aqui abarcadas, dentre elas minhas pesquisas anteriores (ASSIS, 2013; ARAÚJO, ASSIS, 2014), as quais se configuram como os principais pilares para a dimensão descritiva desse trabalho, já que tenciono o aumento do *corpus* anteriormente analisado para a obtenção de dados mais conclusivos no que diz respeito à segmentação linguística da legendagem *closed caption pop-on* da telenovela *Amor Eterno Amor*. Sendo assim, pretendo replicar boa parte da metodologia de meus trabalhos anteriores, e, conseqüentemente, do Projeto CORSEL como um todo, em relação aos métodos de extração, armazenamento, tratamento e análise, do tipo quanti-qualitativa, do *corpus* da pesquisa ora relatada.

No próximo capítulo, discorro sobre a metodologia utilizada para alcançar os objetivos aqui elencados.

3 METODOLOGIA

Nesta seção, irei detalhar o percurso metodológico a ser traçado na pesquisa para os dois propósitos principais da pesquisa: i) a análise descritiva do parâmetro da segmentação linguística em 5 capítulos da LSE da telenovela *Amor Eterno Amor* e ii) a proposta de relegendagem de um trecho de 10 minutos e 9 segundos de *Amor Eterno Amor* a partir dos parâmetros técnicos, linguísticos e paralinguísticos apregoados pela literatura em Tradução Audiovisual acessível/LSE e compartilhados pelo Grupo LEAD-UECE.

3.1 TIPO DE PESQUISA

A pesquisa proposta tem o suporte teórico-metodológico dos Estudos da Tradução, mais especificamente da Tradução Audiovisual Acessível e da Linguística de *Corpus*. A metodologia, de cunho quanti-qualitativo, envolve uma dimensão descritiva pautada por análise baseada em *corpus* e uma proposta de relegendagem à luz dos parâmetros vigentes da pesquisa em legendagem. Por se tratar de um estudo de caso, os resultados aqui alcançados só tem validade para *Amor Eterno Amor*, podendo ser generalizados em menor ou maior grau na medida em que são contrapostos com os resultados de Assis (2013), Araújo e Assis (2014), Gabriel (2015) ou outros trabalhos envolvendo os problemas de segmentação linguística na LSE de telenovelas ou ficção.

3.2 CONTEXTO DA PESQUISA

O presente trabalho trata-se de um recorte do projeto que responde pela sigla CORSEL, que tem como objetivo o estudo da segmentação linguística na programação da TV brasileira. No Projeto CORSEL, estive vinculado como bolsista de Iniciação Científica durante os anos de 2012/2014, período no qual defendi meu Trabalho de Conclusão de Curso (ASSIS, 2013), também dentro do âmbito do projeto e sob a orientação da professora Dra. Vera Lúcia Santiago Araújo. Este trabalho, aliás, foi o primeiro a ser desenvolvido com as prerrogativas e objeto de estudo do projeto em questão, e foi publicado em formato de artigo em Araújo e Assis (2014).

A pesquisa como um todo está atrelada ao Grupo LEAD, que se constitui como uma linha de pesquisa do grupo Tradução e Semiótica, registrado no CNPq, e ao seu respectivo ambiente de trabalho, o LATAV (Laboratório de Tradução Audiovisual). Dado o

contexto em que está situado este trabalho, a grande parte das prerrogativas teóricas e metodológicas a serem utilizadas aqui provém do percurso trilhado pela pesquisa em Legendagem para Surdos e Ensurdecidos no âmbito do Grupo LEAD, através das pesquisas descritivas e exploratórias realizadas com surdos cearenses e de todo o Brasil desde 2002.

3.3 O CORPUS

O *corpus* utilizado na parte descritiva da pesquisa foi a LSE do tipo *pop-on* de 5 capítulos da telenovela *Amor Eterno Amor*, *corpus* integrante do projeto CORSEL, formando um total de 1849 legendas e 14.788 *tokens*¹⁷, excluindo-se os números do *corpus*.

Dentro do escopo do projeto CORSEL, de modo a tentar encontrar padrões de ocorrência de problemas de segmentação linguística na LSE (*closed caption*) do tipo *pop-on* veiculada na TV brasileira, foram selecionados três gêneros televisivos distintos: telenovelas, documentários e seriados tipo comédias de situação¹⁸. Todos esses programas foram gravados através de um sistema de gravação de uma operadora de TV à cabo ao longo de duas semanas. Contudo, em alguns capítulos de algumas das telenovelas do *corpus* não foi possível a visualização ou extração das legendas por meio do software utilizado para esse propósito (*CCExtractor*), sendo essa a razão de se descartar alguns capítulos e fechar um número de cinco capítulos por telenovela (4 telenovelas, no Projeto como um todo), cujas legendas pudessem ser visualizadas, para a composição do *corpus*, inclusive em *Amor Eterno Amor*, objeto do presente trabalho. As duas semanas de gravação de *Amor Eterno Amor* foram transmitidas de 18 a 23 de Junho e de 25 a 30 de Junho de 2012. Os capítulos que foram selecionados para compor o *corpus* foram veiculados nos dias 25, 26, 27, 28 e 29 de Junho e, para fins de organização, foram nomeados durante a análise como capítulos 1, 2, 3, 4 e 5, respectivamente.

Além das circunstâncias técnicas apresentadas acima, a escolha em utilizar 5 capítulos da LSE de *Amor Eterno Amor* para os propósitos desta pesquisa – diferentemente de

¹⁷ *Tokens*, segundo Berber Sardinha (2004, p. 165), são: “Palavras, ocorrências ou itens (*tokens*): contagem de palavras corridas (*running words*); cada palavra conta como uma ocorrência, mesmo que seja repetida”.

¹⁸ Como está destacado em Araújo (2015), a escolha desses três gêneros foi pautada pelo fato de que esses são os itens da programação que normalmente contém a legenda do tipo *pop-on*. Apesar do desejo em montar um corpus com exemplares de diferentes emissoras de TV, a falta de uma regularidade na transmissão de uma programação legendada por parte de outras emissoras fez com que o corpus do Projeto fosse composto apenas por programas da Rede Globo.

apenas um capítulo¹⁹ da LSE da mesma telenovela em questão em Assis (2013) e Araújo e Assis (2014) – visa expandir as considerações acerca dos problemas de segmentação linguística encontrados na LSE de *Amor Eterno Amor*. Considero possível, com a quantidade de 5 capítulos da LSE em questão, traçar considerações mais sólidas acerca da caracterização da segmentação linguística na LSE do tipo *closed caption* da telenovela em questão, reforçando ou refutando os resultados encontrados em Assis (2013) e Araújo e Assis (2014). Esse posicionamento em favor da representatividade do *corpus* da presente pesquisa ocorre pelo fato de que não há critérios objetivos para a determinação dessa representatividade, a não ser aqueles estipuladas pelo próprio analista. Ou seja, como pontua Leech (1991, p. 27, *apud* BERBER SARDINHA, 2004, p. 25), “os usuários de um corpus atribuem a ele a função de ser representativo de certa variedade”. Dessa maneira, considero os 5 capítulos da LSE de *Amor Eterno Amor* analisados na pesquisa ora relatada como representativos do universo da LSE da telenovela em questão como um todo.

3.3.1 Extração do *corpus* e armazenamento

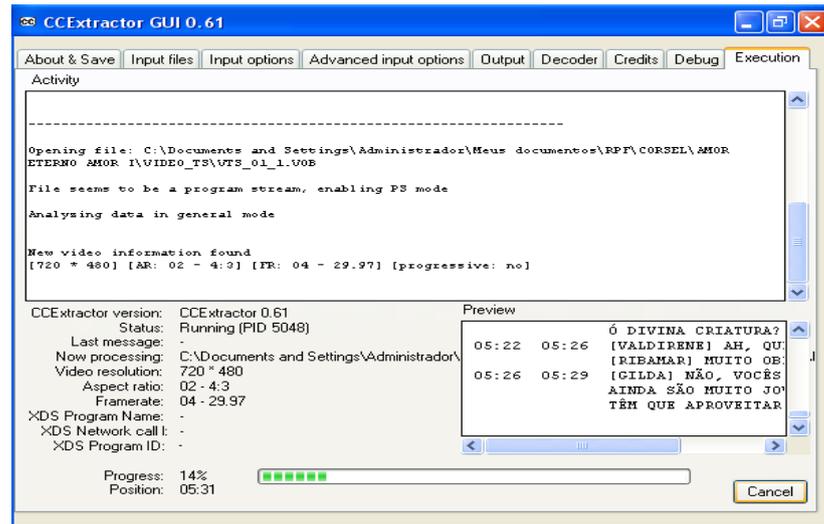
Para a extração automática das legendas que serão utilizadas na análise, utilizei o software *CCExtractor 0.61*²⁰, cuja interface está na Figura 4. Este programa extrai, quase instantaneamente, legendas *closed caption* que porventura possam existir em um arquivo de DVD²¹, formato do *corpus* integrante do projeto CORSEL.

¹⁹ O capítulo analisado em Assis (2013) e Araújo e Assis (2014) também está presente no corpus analisado, por ele fazer parte dos 5 capítulos de *Amor Eterno Amor* (Capítulo 3) selecionados pelo Projeto CORSEL. Dessa maneira, não encontrei razões suficientes para que ele fosse descartado dessa análise maior.

²⁰ Programa *freeware* que pode ser baixado no domínio <http://ccextractor.sourceforge.net/>

²¹ Os capítulos dos quais foram extraído a LSE da pesquisa ora relatada foi obtido através da gravação do sinal televisivo por um gravador de DVD.

Figura 4 – Interface CCEXtractor

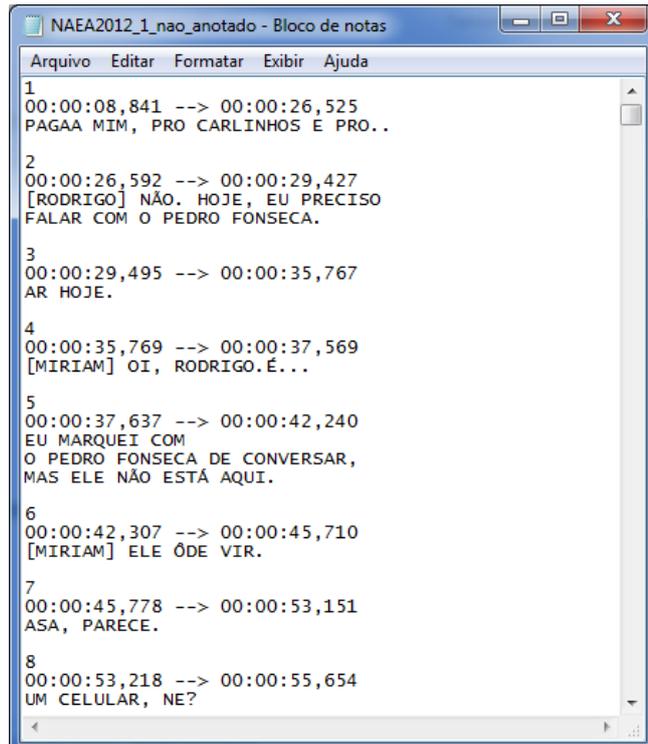


Fonte: Elaborada pelo autor

O programa gera um arquivo de legendagem srt. com o número de inserção das legendas, assim como a marcação e sincronização das mesmas. Por vezes, contudo, o arquivo gerado apresenta pequenas falhas que precisam ser corrigidas manualmente, como no caso das legendas 1, 3, 6 e 7²² na Figura 5.

²² A nossa experiência nos sugere que as legendas 1, 3, 6 e 7 apresentaram problemas, provavelmente, devido a instabilidade do sinal em que é veiculado a legendagem closed caption. Contudo, há a possibilidade de o erro ter ocorrido na própria confecção das legendas por parte da emissora, sendo difícil detectar com absoluta certeza em que instância desse processo o erro aconteceu.

Figura 5 – Arquivo de legenda em formato srt. aberto no Bloco de Notas²³



Fonte: Elaborado pelo autor

Tendo como exemplo a primeira inserção de legenda, pode-se observar a linha composta do número '1' indicando a posição daquela legenda na sequência de inserções. Na indicação do tempo de entrada daquela legenda, logo abaixo, podemos ver '00:00:08,841' e, ao seu lado, o tempo de saída '00:00:26,525', definindo assim o tempo em que aquela legenda permanecerá na tela.

Sendo o *corpus* composto pela LSE de cinco capítulos de *Amor Eterno Amor*, parte integrante do Projeto CORSEL, os parâmetros utilizados para a construção do nome de cada arquivo foram os mesmos do projeto em questão, que funcionam como espécies de cabeçalho²⁴ de cada arquivo. Para a confecção desses títulos, com fins de organização dos dados, fez-se uso das seguintes premissas (ARAÚJO, 2015): tipo de filme ou programa, seguido de sublinhado – Título (sempre as iniciais), ano de produção, número do capítulo e seu tipo de anotação - não-annotado ou anotado. Sendo assim, para a novela *Amor Eterno Amor* o título ficou dessa maneira: N = novela (também poderia ser de filme de longa

²³ Editor de texto disponível no sistema operacional *Windows*, da *Microsoft*.

²⁴ Segundo Berber Sardinha (2004, p. 73), "Cabeçalhos são uma parte do arquivo de cada texto do *corpus* que contém informações sobre o texto, tais como origem, a data de coleta, o grupo de pesquisa responsável, o tamanho do texto, sistemas de transcrição [...]", dentre outras características. Esse título, que foi dado para cada arquivo do CORSEL, funciona como uma espécie de cabeçalho, com a diferença que essas informações estão no nome do arquivo, não inseridos no texto do arquivo.

metragem - FLM - ou média metragem – FMM); AEA= Amor Eterno Amor (sempre as iniciais do programa/filme); 2012 = ano do programa; 1, 2, 3, 4 e 5 = números dos capítulos gravados; não-annotado = arquivo “limpo”, ou seja, sem a inserção das etiquetas de segmentação e de parâmetros técnicos.

3.3.2 Etiquetagem

As etiquetas²⁵ utilizadas nesta pesquisa para a descrição quanti-qualitativa da segmentação linguística na LSE brasileira foram as etiquetas e subetiquetas de segmentação linguística utilizadas no âmbito do Projeto CORSEL, estas baseadas, primeiramente, nas etiquetas e subetiquetas criadas por Chaves (2012), Assis (2013) e Araújo e Assis (2014) e calcadas nos pressupostos teóricas funcionalistas de sintaxe de Castilho (2012). As etiquetas e subetiquetas de Assis (2013) e Araújo e Assis (2014), contudo, foram remodeladas pelo nosso grupo de pesquisa a fim de eliminar algumas poucas incongruências terminológicas, assim como para abarcar, a partir de termos mais gerais, algumas ocorrências que se tornavam problemáticas quando confrontadas com as etiquetas previamente utilizadas. Elas foram criadas com base no padrão *SGML* (*Standard Generalized Markup Language*), que fornece códigos escritos no formato <etiquetas de abertura>informação</etiqueta de fechamento/>, caracterizando-as e delimitando o conteúdo a ser analisado. Assim como em Chaves (2012) para a confecção das etiquetas, o sinal de (+) de modo a indicar a o lugar em que a estrutura sintagmática havia sido quebrada indevidamente, o que caracteriza o problema de segmentação linguística em legendagem.

Abaixo uma explicação breve sobre o que significa cada uma das subetiquetas de quebra indevida dos sintagmas:

- **SUBETIQUETAS DE ANÁLISE DE QUEBRA DE SIN TAGMA NOMINAL (SN)**
 - <SN_especif+SN> (quebra entre o núcleo do SN – substantivo – e o seu especificador; quebra entre os constituintes do especificador)
 - <SN_núcleo> (quebra no núcleo de um SN composto, como um nome próprio)
 - <SN_subst+SAdj> (quebra entre um substantivo e um SAdj que se relacionam)
 - <SN_subst+SP> (quebra entre o núcleo de um SN e um SP quando este possui função adjetiva)

²⁵ A etiquetagem aqui utilizada pode ser denominada como “discursiva”, Segundo a classificação de Tagnin (2010), que é um processo no qual o conteúdo do corpus recebe etiquetas demarcando uma determinada parte do texto. Entretanto, no escopo do projeto CORSEL e da presente pesquisa, elas são situadas como ‘etiquetas de segmentação’ apenas, dada a especificidade da etiquetagem aqui empreendida.

- **SUBETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA VERBAL (SV)**
 - <SV_composto> (quebras entre dois verbos do SV, um auxiliar + um verbo na forma nominal – infinitivo, gerúndio, particípio)
 - <SV_verbo+SAdv> (quebra entre verbo e um SAdv)
 - <SV_(verbo)+oblíquo+SV> (quebra entre pronome oblíquo (precedido ou não de verbo) e o verbo aquele está relacionado)

- **SUBETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA PREPOSICIONAL (SP)**
 - <SP_prep+SN> (quebra entre o núcleo de um SP – preposição – e um substantivo)
 - <SP_prep+SV> (quebra entre o núcleo de um SP e um verbo)

- **SUBETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA ADJETIVAL (SAdj)**
 - <SAdj_especif+SAdj> (quebra entre o especificador - um advérbio, por exemplo - e o núcleo do SAdj – adjetivo)

- **SUBETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA ADVERBIAL (SAdv)**
 - <SAdv_adv+adv> (quebra da estrutura interna de um SAdv complexo)

- **SUBETIQUETAS DE ANÁLISE DE QUEBRA DE ORAÇÃO COORDENADA (COORD)**
 - <COORD_conj+oração> (quando há quebra entre a conjunção coordenativa (e, mas, e etc.) e oração coordenada)

- **SUBETIQUETAS DE ANÁLISE DE QUEBRA DE ORAÇÃO SUBORDINADA (SUBORD)**
 - <SUBORD_conj/pron_rel+oração> (quando há quebra entre a conjunção subordinativa/pronome relativo (quando, porque e etc.) e oração subordinada)

Além das etiquetas para designar os problemas de segmentação, outras, de caráter técnico, referentes ao número das legendas, número de linhas da legenda, número de caracteres por linha e velocidade da legenda foram utilizadas de modo a traçar, a partir da análise quanti-qualitativa desses dados, o contexto no qual as legendas com problemas de segmentação linguística aparecem e, dessa forma, verificar se o padrão de legendas com velocidade alta e de 3 linhas no qual ocorrem mais problemas de segmentação LSE da telenovela analisada se repete na análise de uma quantidade maior de *corpus* do que em Araújo e Assis (2014). Abaixo, no Quadro 4, as etiquetas técnicas e de segmentação utilizadas na parte descritiva do trabalho.

Quadro 4 – Etiquetas de características técnicas e de problemas segmentação linguística em legendagem

ETIQUETAS DE ANÁLISE DE PARÂMETROS TÉCNICOS DA LEGENDAGEM	
Número da legenda	<sub1 ²⁶ >legenda1</sub1>
Linhas por legenda	<1L>, <2L> e <3L>
Tempos inicial e final de cada legenda	<t>início --> final</t>
Número de caracteres por linha (aplicada em legendas de 2 e 3 linhas)	<cp1 ²⁷ >
Velocidade da legenda baixa (145ppm ²⁸) – 14cps à 15,9 cps ²⁹	<velocidade da legenda_baixa>
Velocidade de legenda média (160ppm) – 16cps à 16,9 cps	<velocidade da legenda_média>
Velocidade de legenda alta (180ppm) – 17cps à 18,9 cps ³⁰	<velocidade da legenda_alta>
ETIQUETA INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO LINGÜÍSTICA	
<PROSEGL>	
SUBETIQUETAS PARA O SINTAGMA VERBAL (SV)	
<SV_composto>	
<SV_verbo+SAdv>	
<SV_(verbo)+oblíquo+SV>	
<SV_verbo+SP>	
SUBETIQUETA PARA O SINTAGMA NOMINAL (SN)	
<SN_especif+SN>	
<SN_subst+SAdj>	
<SN_subst+SP>	
<SN_núcleo>	
SUBETIQUETA PARA O SINTAGMA ADJETIVAL (SAdj)	
<SAdj_especif+SAdj>	
SUBETIQUETA PARA O SINTAGMA ADVERBIAL (SAdv)	
<SAdv_adv+adv>	
SUBETIQUETA PARA O SINTAGMA PREPOSICIONAL (SP)	
<SP_prep+SN>	
<SP_prep+SV>	
SUBETIQUETA PARA A ORAÇÃO COORDENADA (COORD)	
<COORD_conj+oração>	
SUBETIQUETA PARA A ORAÇÃO SUBORDINADA (SUBORD)	
<SUBORD_conj/pron_rel+oração>	

Fonte: Arquivos do CORSEL

Após o processo de anotação do *corpus*, feito a partir da edição do arquivo txt. aberto em qualquer editor de texto, as etiquetas foram analisadas pelo programa de análise

²⁶ Abreviatura de *subtitle*, ‘legenda’ em inglês.

²⁷ Acrônimo de ‘caracteres por linha’.

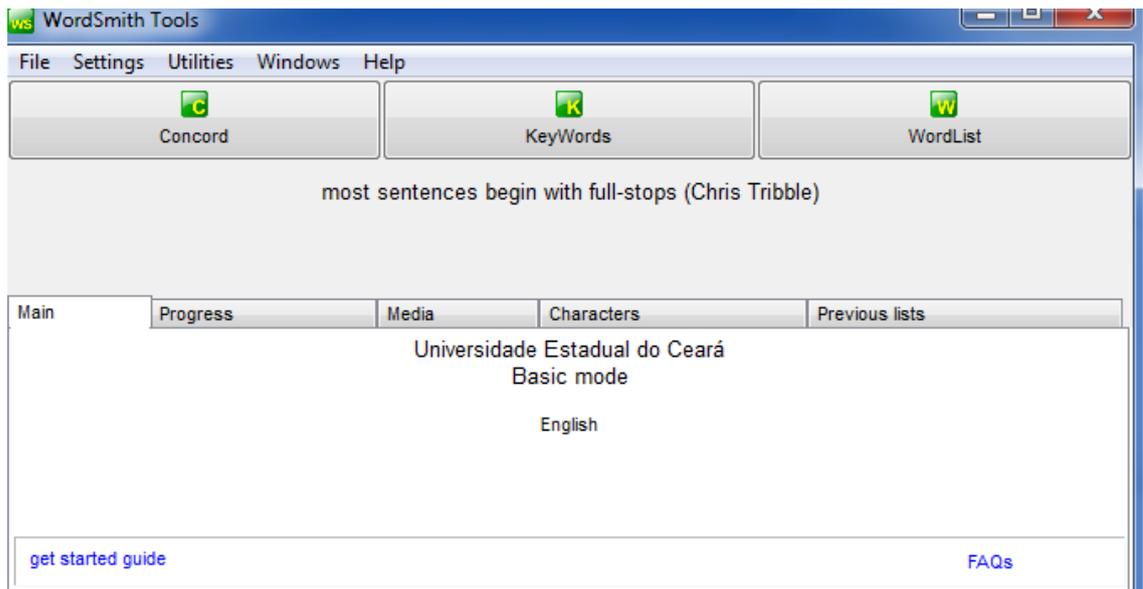
²⁸ Acrônimo de ‘palavras por minuto’.

²⁹ Legendas com velocidade abaixo de 14cps (145ppm) também foram categorizadas como sendo de ‘velocidade baixa’.

³⁰ Legendas com velocidade acima de 18,9 cps (180ppm) também foram categorizadas como sendo de ‘velocidade alta’.

linguística *Wordsmith Tools* 5.0. O mesmo possui três ferramentas principais: *Wordlist*, *Concord* e *Keywords*, mas no trabalho apenas duas fazem-se necessárias: a *Wordlist*, para a obtenção do número de *tokens* (palavras corridas) que compõem o corpus a partir dos arquivos selecionados, e a *Concord*, para buscar o número de ocorrências dos nódulos que descrevem problemas de segmentação linguística e mostrar o contexto das ocorrências encontradas. Na Figura 6, abaixo, imagem da interface inicial do *Wordsmith Tools* 5.0.

Figura 6 – Interface inicial do *Wordsmith Tools* 5.0



Fonte: elaborada pelo autor

Na Figura 7, há um exemplo de um trecho da LSE de *Amor Eterno Amor* etiquetado.

Figura 7 – Arquivo anotado aberto no Bloco de Notas

```

Arquivo Editar Formatar Exibir Ajuda
|<sub1><1L>1
<t>00:00:12,879 --> 00:00:21,653</t>
<cp126>VOCÉ SABE DO QUE ECIÊNCIA.
<veloc_leg_baixa 26c/8,7s></sub1>

<sub2><2L>2
<t>00:00:21,721 --> 00:00:26,892</t>
<cp126>QUE NO LUGAR DO INTELLECTO,
<cp124>USA A FORÇA FIEM MULHER!
<veloc_leg_baixa 50c/5,1s></sub2>

<sub3><1L>3
<t>00:00:26,959 --> 00:00:30,428</t>
<cp118>LÁ DO SEU PASSADO.
<veloc_leg_baixa 18c/3,4s></sub3>

<sub4><2L>4
<t>00:00:30,496 --> 00:00:34,633</t>
<cp118>É FEZ? AH, SAIU NO <PROSEGL> <SP_prep+SN>
<cp132>JORNAL, VOCÉ DANDO PANCADA NELA!
<veloc_leg_baixa 50c/4,1s></sub4>

<sub5><2L>5
<t>00:00:34,700 --> 00:00:39,504</t>
<cp116>[DIMAS] MELISSA,
<cp132>EU NÃO ESTOU BRINCANDO, MELISSA.
<veloc_leg_baixa 48c/4,8s></sub5>

<sub6><3L>6
<t>00:00:39,572 --> 00:00:45,977</t>
<cp115>O QUÉ QUE VOCE,
<cp125>MELISSA SOBRAL, FOI FAZER
<cp127>NUMA DELEGACIA DE SUBÚRBIO?
<veloc_leg_baixa 67c/6,4s></sub6>

```

Fonte: Elaborada pelo autor

Na Figura 7, pode-se observar como as informações relativas aos parâmetros técnicos e linguísticos da LSE são etiquetadas. Após esse processo feito manualmente pelo analista, o arquivo está pronto para ser analisado quantitativamente pelo *Wordsmith Tools*. Para fins de organização em relação ao armazenamento do *corpus* compilado, o nome do arquivo exemplificado na Figura 7 ficou ‘NAEA2012_5_annotado’, conforme explicado na subseção anterior.

3.4 A RELEGENDAGEM

Esta subseção tem como intuito apresentar as etapas seguidas para a relegendagem da LSE de um trecho da telenovela *Amor Eterno Amor*.

Amor Eterno Amor foi uma telenovela brasileira produzida e exibida pela Rede Globo entre 5 de março e 7 de setembro de 2012, em 161 capítulos, escrita por Elizabeth Jhin e com direção geral de Pedro Vasconcelos. Em linhas gerais, a trama gira em torno, principalmente, da história de Carlos, que aos 10 anos fugiu do interior de Minas Gerais por causa dos maus-tratos infringidos por seu padrasto. Ao fugir, foi encontrado na estrada por

um caminhoneiro chamado Xavier, que acabou criando o menino como se fosse o seu filho. Com o passar dos anos, Carlos se transformou em um homem bonito e respeitado em sua cidade, possuindo o talento de domar búfalos, pelo qual ganhou o apelido de Barão. Sua mãe, Verbena Borges, viúva milionária residente no Rio de Janeiro, nunca soube o motivo de seu filho ter sumido, sem nunca sequer ter suspeitado ter sido por causa de seu ex-marido, Virgílio. A trama desenrolou em volta da busca de Verbena por seu filho desaparecido, e dos esforços de sua irmã, Melissa, e seu cunhado, Dimas, interessados em sua herança, em impedir este encontro.

O trecho da novela e, conseqüentemente, sua LSE (objeto da relegendagem aqui proposta) foi retirado do capítulo 5 do *corpus*. Este trecho, composto de 10 minutos e 9 segundos de vídeo, está presente no final do capítulo em questão e traz 5 principais cenas em sua narrativa, a saber: 1 – Laís e Julinho no acampamento após sua fuga; 2 – Reunião de várias pessoas preocupadas com o desaparecimento de Laís e Julinho; 3 – Conversa de pai e filha entre Gabriel e Miriam; 4 – conversa entre Fernando e sua mãe; 5 – Dimas na delegacia denunciando sua esposa Melissa.

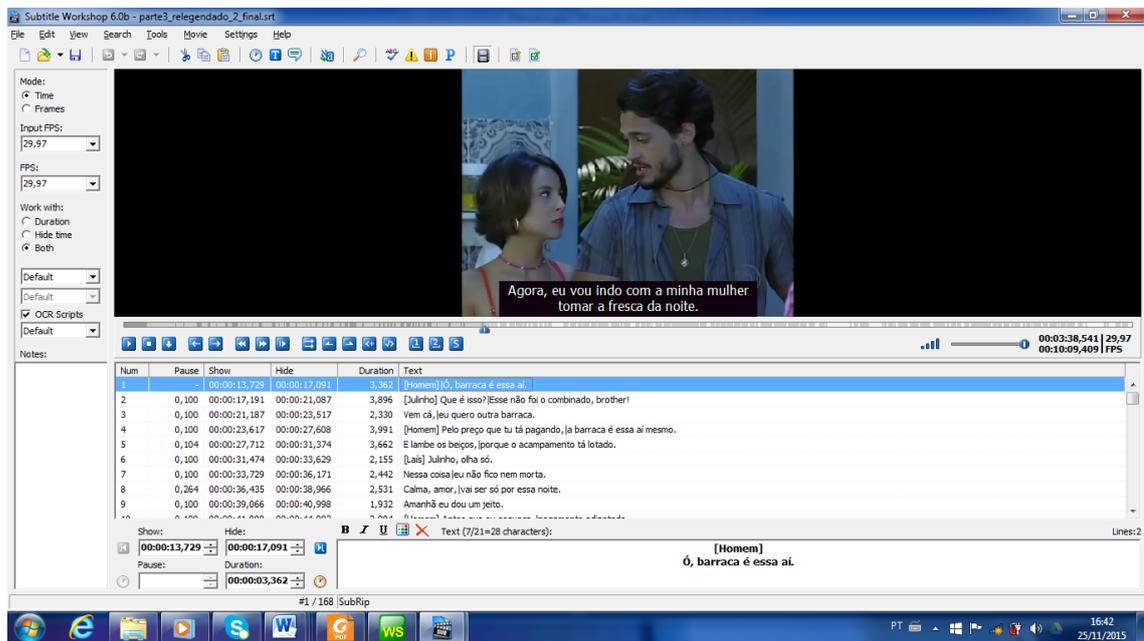
A LSE de partida do trecho em questão era formada por 154 legendas, com 36 problemas de segmentação em sua totalidade, e foi escolhido dentre tantos trechos da telenovela que formava o *corpus* da pesquisa por conter um número de legendas suficiente para que a análise da atividade de relegendagem fosse proveitosa, a meu ver, e, ao mesmo tempo, por conter um alto número de problemas de segmentação linguística. Este fato propiciou que fossem postos em prática os resultados discutidos durante a análise quantitativa dos problemas de segmentação do *corpus* em termos de uma tarefa prática de legendagem que considerasse esses problemas vinculados ao material audiovisual legendado como um todo, ou seja, em relação aos seus parâmetros técnicos, linguísticos e paralinguísticos,

3.4.1 Subtitle Workshop 6.0B

A relegendagem foi feita com o auxílio do programa freeware *Subtitle Workshop 6.0B S.W.* O *Subtitle Workshop* é uma ferramenta completa, eficiente e prática utilizada para a edição de legendas de vídeo. Através do programa, é possível realizar as tarefas de marcação, tradução e pré-visualização do produto audiovisual legendado. O *Subtitle Workshop* foi responsável em grande parte pela democratização do acesso à tarefa de legendar por *fansubbers* (legendistas piratas ou fãs), que tinham interesse em legendar suas séries ou

animes (desenhos animados japoneses) antes de elas serem disponibilizadas oficialmente, assim como foi adotado por muitos legendistas profissionais que veem nele a praticidade e completude necessárias para o exercício da legendagem. O *Subtitle Workshop* foi desenvolvido pela URUsoft - <http://www.urusoft.net>. A Figura 8 traz a tela principal de trabalho do programa.

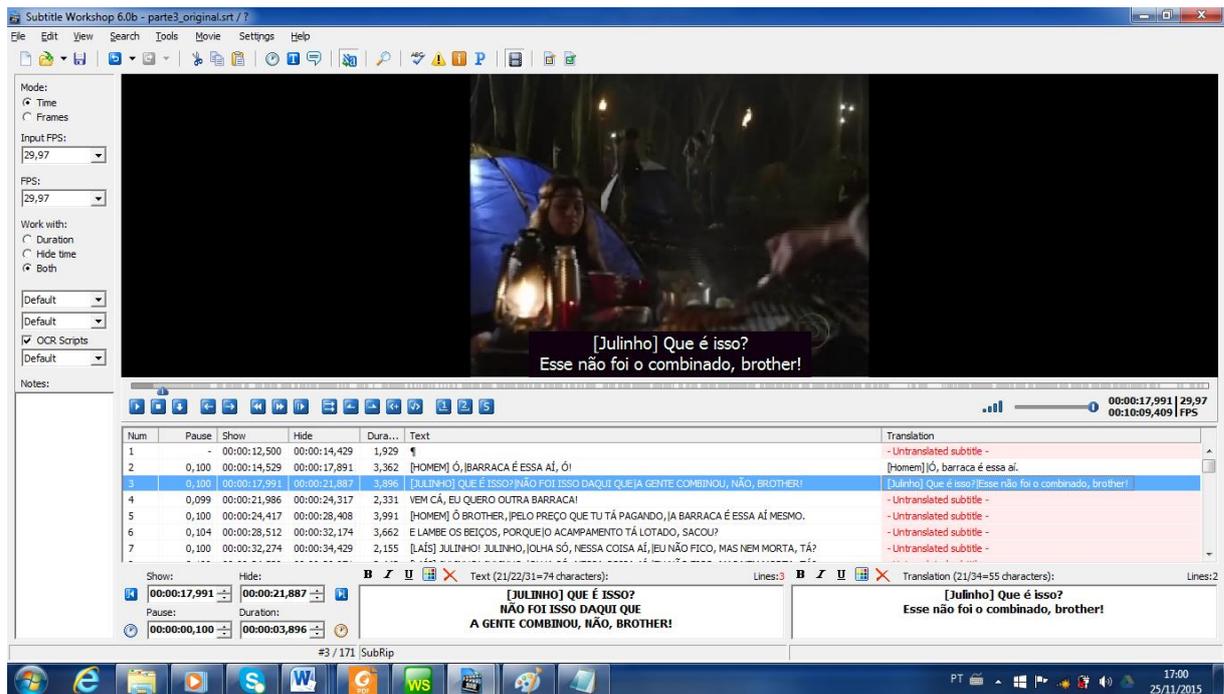
Figura 8 – Interface do *Subtitle Workshop* 6.0B



Fonte: elaborada pelo autor

Algumas ferramentas do *Subtitle Workshop* foram preponderantes para a fluidez da atividade de relegendagem proposta. Um dos dois principais recursos utilizados foi o *TRANSLATOR MODE* (Modo Tradutor), que ao me permitir visualizar a LSE de partida do lado esquerdo da tela e a tradução proposta na parte direita da tela, propiciou uma maior organização e sistematização dos comentários e análises qualitativas empreendidos em relação às escolhas feitas durante a atividade de relegendagem. A Figura 9 traz a tela do *Subtitle Workshop* na ferramenta *TRANSLATOR MODE*, que virou uma espécie de estação de trabalho durante a proposta de relegendagem.

Figura 9 – Interface do *Subtitle Workshop* utilizando a ferramenta *TRANSLATOR* MODE

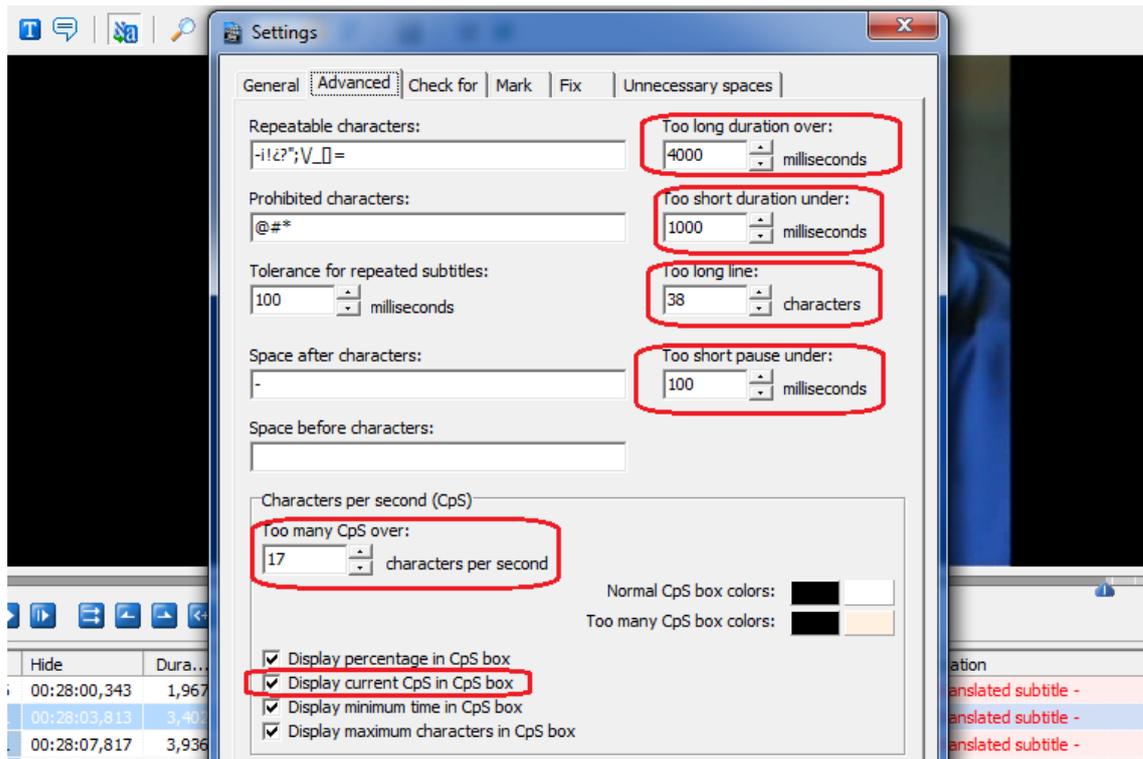


Fonte: Elaborada pelo autor

Na Figura 9, pode-se perceber como esse processo de relegendagem ocorreu. Todas as traduções propostas visavam adequar a LSE de partida em relação aos parâmetros linguísticos, técnicos e paralinguísticos apreçados pela literatura em LSE e advogadas pelo Grupo LEAD-UECE.

Os ajustes feitos na relegendagem em relação aos parâmetros técnicos e linguísticos foram auxiliados por outra função do *Subtitle Workshop* 6.0B que permitiu verificar, entre outras coisas, se havia, na LSE de partida, legendas com mais de duas linhas, duração insuficiente/excessiva, linhas longas demais, pausas curtas demais e excesso de caracteres por segundo. É possível determinar esses parâmetros a partir da guia *TOOLS* (ferramentas) > *INFORMATION AND ERROS* (informações e erros) > *SETTINGS* (configurações) > *ADVANCED* (avançado). Na Figura 10, é possível ver a tela em que se configura os parâmetros em questão.

Figura 10 – Tela do *Subtitle Workshop* na guia *SETTINGS > ADVANCED*

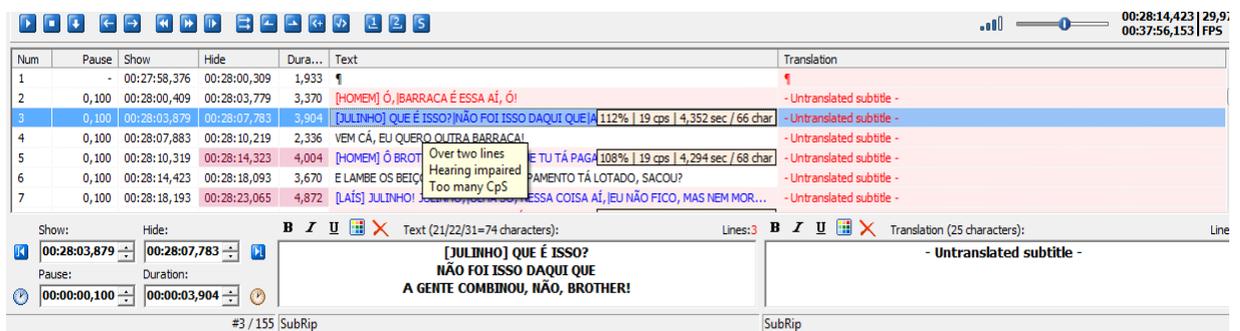


Fonte: Elaborada pelo autor

Estão destacadas em vermelho na Figura 10 algumas das configurações mencionadas. Na opção *Too long duration over* (duração muito longa acima de) eu estabeleci que o máximo de tempo da legenda seria de 4.000 milissegundos (4 segundos); que a legenda teria uma *too short duration under* (duração muito curta abaixo de) de 1000 milissegundos (1 segundo); que seria uma *too long line* (linha muito longa) a partir de 38 caracteres, ou seja, seria permitido um máximo de 37 caracteres por linha; que seria uma *too short pause under* (pausa muito curta abaixo de) 100 milissegundos, ou seja, intervalo mínimo de 100 milissegundo entre duas inserções de legenda; que seriam *too many CpS over* (muitos caracteres por segundo acima de) 17 caracteres por segundo, que corresponde à velocidade de legenda de 180 palavras por minuto utilizada como limite para a relegendagem proposta. A opção *Display current CpS in CpS box* (mostrar a caixa atual de caracteres por segundo) sinaliza as legendas que estão com velocidade acima daquela estabelecida, no caso, 17 caracteres por segundo, e qual a velocidade corrente daquela legenda. Todas essas opções foram definidas de acordo com os pressupostos técnicos assumidos pela análise, baseados na literatura em LSE e nos parâmetros do Grupo LEAD-UECE.

Todas essas opções funcionam de modo a verificar a existência de características na legendagem que não estejam de acordo com as assinaladas, ou seja, elas não são corrigidas automaticamente, são apenas destacadas através de outra ferramenta presente na guia *TOOLS* > *INFORMATION AND ERROS* > *INFORMATION AND ERRORS*. Ao clicar nesta opção, se houver casos de inserções que ultrapassem os critérios pré-definidos nas configurações, surge uma pequena tela com todas as informações relativas aos erros presentes no arquivo de legendagem, discriminando a quantidade, o tipo de problema e onde eles ocorrem. Ao mesmo tempo que isso ocorre, as legendas “problemáticas” são ressaltadas através de cores que significam o tipo de problema e através de pequenas caixinhas de informações relativas aos caracteres por segundo (excedentes) daquela legenda. Um exemplo de como essas informações relativas aos erros presentes no arquivo de legendagem são visualizadas está na Figura 11.

Figura 11 – Tela do *Subtitle Workshop* evidenciando os erros da legendagem



Fonte: Elaborada pelo autor

No exemplo presente na Figura 11, pode-se ver essa “visualização” dos erros presentes na legendagem. Em fundo de uma cor-de-rosa bem clara, são destacadas as legendas que contém algum tipo de problema. Em fonte azul, são ressaltadas as legendas que contém mais de 2 linhas (esse parâmetro, contudo, é padrão do programa, não tendo sido determinado por mim, o que enfatiza a importância desse parâmetro técnico e que, mesmo assim, não é seguido pelas TVs na confecção de suas legendas). Em fonte vermelha, é destacado a presença de colchetes, ou seja, de legendas próprias para surdos. Apesar de isso não ser um erro, o programa destaca quando há esse tipo de legenda. Em um fundo vermelho mais escuro, é destacado o tempo que excede o máximo estabelecido nas configurações (no caso, 4 segundos). Quando se posiciona a seta do mouse sobre as legendas “problemáticas” (continuemos chamando-as assim, apesar de legendas com colchetes, próprias para surdos,

não serem, de fato, problemáticas), aparecem as informações relativas às características daquela legenda, no caso ressaltado no exemplo, que a legenda está *over two lines* (acima de duas linhas), específica para *hearing impaired* (surdos) e que contém *too many CpS* (muitos caracteres por segundo).

Outra questão ressaltada na Figura 11 é a da quantificação dos caracteres por segunda das inserções de legenda, ou seja, a velocidade de cada legenda quando essa velocidade excede àquela estabelecida como limite nas configurações. Essas informações são visualizadas através de *boxes* (caixas) em formato retangular de fundo amarelo que ficam sobre a legenda “problemática”. Essas informações foram selecionadas ainda na seção de configurações representada na Figura 10 através das opções elencadas a seguir e que serão exemplificadas com as informações relativas à legenda 3 presente na Figura 11 : *Display percentage in CpS box* (Mostrar porcentagem na caixa dos Cps), que mostra quantos por cento a mais de 100% (representando o limite de CpS) estão os caracteres por segundo daquela legenda; no caso da legenda 3 usada como exemplo, 112%; *Display current CpS in CpS box* (Mostrar o CpS corrente na caixa de CpS), que mostra a velocidade da legenda corrente em termos de caracteres por segundo, no caso, 19 caracteres por segundo (a mais do teto de 17 CpS estabelecido); *Display minimum time in CpS box* (Mostrar tempo mínimo na caixa de CpS), que mostra o tempo mínimo que aquela legenda em questão deveria ter para acomodar o seu número de caracteres, no caso, 4,352 segundos para os 74 caracteres da legenda de partida, ao invés de seus 3,904 segundos (aqui, no caso, não houve, na indicação da caixa, o respeito ao limite de 4 segundos pré-estabelecido); e *Display maximum characters in CpS box* (Mostrar máximo de caracteres na caixa de CpS), que mostra o número máximo de caracteres que aquela legenda deveria ter com o tempo disponível, no caso, 66 caracteres ao invés de 74.

As informações disponibilizadas por essas pequenas caixas de informação relativas à velocidade da legenda foram de grande ajuda no momento da relegendagem do trecho em questão para adequá-lo à velocidade de 180ppm (17 caracteres por segundo) proposta.

No próximo capítulo, trato dos resultados alcançados pelas duas etapas de análise da pesquisa: a descrição-classificação dos problemas de segmentação linguística da LSE de 5 capítulos da telenovela *Amor Eterno Amor* e a proposta de relegendagem de um trecho de *Amor Eterno Amor* envolvendo os parâmetros paralinguísticos, linguísticos e técnicos inerentes à legendagem.

4 RESULTADOS DA ANÁLISE DOS DADOS

O presente capítulo divide a apresentação dos resultados da análise empreendida na pesquisa ora relatada em dois momentos distintos. Em um primeiro momento, apresento os resultados da análise dos problemas de segmentação linguística na LSE de 5 capítulos da telenovela *Amor Eterno Amor*, aliando as perspectivas quantitativa – gerada pelo uso das ferramentas *Wordlist* e *Concord* provenientes do *Wordsmith Tools* – e qualitativa – ao interpretar o contexto no qual ocorre os problemas de segmentação linguística no *corpus* e observar possíveis regularidades para a ocorrência desse parâmetro – da pesquisa. Em um segundo momento, apresento os resultados da análise da LSE de um trecho do *corpus* e proponho uma relegendagem sobre esse excerto. Ao fazer esse exercício, pretendo utilizar como base para essa proposta o que é abalizado pela literatura em LSE como um todo, e, mais especificamente, os resultados das pesquisas descritivas e exploratórias do Grupo LEAD-UECE em LSE com surdos do Ceará e de todo o Brasil.

4.1 RESULTADOS: PANORAMA GERAL DO CORPUS E PROBLEMAS DE SEGMENTAÇÃO LINGUÍSTICA EM *AMOR ETERNO AMOR*

4.1.1 Panorama geral do corpus

Nesta subseção, são apresentados os resultados dos dados gerados pela *Wordlist*, uma das três ferramentas oferecidas pelo suíte *Wordsmith Tools 5.0*, a qual oferece dados relativos ao panorama geral do *corpus*, como número de palavras, palavras mais recorrentes, entre outros. Esses dados são relevantes, pois podem auxiliar a compreender a natureza lexical do *corpus* analisado. Abaixo, na Figura 12, um *printscreen* da ferramenta *Wordlist* na guia *statistics*, a qual revela alguns dados sobre a natureza linguística do *corpus* analisado.

Figura 12 – Tela da Wordlist na guia statistics

	N	Overall	1	2	3	4	5
text file		Overall	NAEA2012_1_nao_annotado.srt	NAEA2012_2_nao_annotado.srt	NAEA2012_3_nao_annotado.srt	NAEA2012_4_nao_annotado.srt	NAEA2012_5_nao_annotado.srt
file size	156,645		32,104	29,571	26,047	18,714	50,209
tokens (running words) in text	31,433		6,412	5,967	5,177	3,837	10,040
tokens used for word list	14,788		2,956	2,844	2,530	1,695	4,763
sum of entries							
types (distinct words)	2,451		872	794	741	610	1,167
type/token ratio (TTR)	16.57		29.50	27.92	29.29	35.99	24.50
standardised TTR	23.24		23.42	22.46	23.56	23.00	23.45
standardised TTR std. dev.	72.66		64.54	63.89	62.69	58.71	68.47
standardised TTR basis	1,000		1,000	1,000	1,000	1,000	1,000
mean word length (in characters)	3.19		3.21	3.18	3.25	3.09	3.20

Fonte: Elaborada pelo autor

Ao se observar os dados da coluna *overall*, a qual traz os dados da LSE dos 5 capítulos analisados, pode-se ver que a linha *tokens (running words) in text* mostra que há a presença de 31.433 palavras corridas – *tokens* – nos arquivos em questão, incluindo os números relativos à marcação dos tempos inicial e final de cada inserção. Mais importante que essa linha, contudo, é a que está localizada logo abaixo, a *tokens used for word list*, que trata dos itens lexicais – excluindo-se os números – presente no *corpus*, 14.788 palavras. Esse dado é relevante, pois é a partir dele que será feito o cálculo à respeito do Índice de Frequência Simples, a ser apresentando mais adiante. Há ainda os dados da linha *types (distinct words)*, que apresenta as palavras distintas – *types* – que ocorrem no *corpus*, ou seja, que aparecem pelo menos uma vez, não importando a repetição de ocorrências da mesma, perfazendo um total de 2.451 *types*.

4.1.2 Resultados da Etiquetagem

Nesta subseção, apresento os resultados numéricos relativos aos problemas de segmentação linguística encontrados no *corpus*, exibindo dados provenientes da busca por concordâncias da etiqueta PROSEGL e suas subetiquetas no *corpus*. Ao mesmo tempo em que apresento os resultados desses dados quantitativos, também exemplifico a ocorrência de uma subetiqueta de problemas de segmentação linguística no *corpus*, analisando-a enquanto legenda como um todo – ou seja, considerando os seus parâmetros técnicos e linguísticos para além do parâmetro da segmentação – ao propor possíveis ‘ressegmentações’ dessas quebras indevidas. Ou seja, além de apontar e discriminar o porquê da inadequação do ponto de vista sintagmático das quebras entre linhas das legendas exemplificadas, tento demonstrar possíveis

soluções para a correção dessas quebras utilizando uma noção macro de legendagem, e não só “desfazendo” a ruptura problemática do sintagma.

A análise quantitativa do *corpus* apresentou um total de 333 casos de PROSEGL. O nível sintagmático com o maior número de ocorrências de problemas de segmentação linguística no *corpus* foi o verbal, com 146 ocorrências (43,8%), seguido, em ordem decrescente, pelo sintagma nominal com 109 ocorrências (32,8%), em consonância com os resultados obtidos por Chaves (2012), Araújo e Assis (2014), Arraes (2015) e Gabriel (2015). Seguindo a ordem decrescente de ocorrências dos problemas de segmentação no *corpus*, o sintagma preposicional apresentou 32 ocorrências (9,6%); a categoria das orações subordinadas apresentou 30 das ocorrências (9%); a categoria das orações coordenadas e o sintagma adjetival apresentaram 6 ocorrências cada (1,8%); e, por último, o sintagma adverbial apresentou 4 das ocorrências (1,2%). A Tabela 4 traz os resultados gerais e os resultados por tipos de problemas de segmentação linguística encontrados no *corpus* em termos de valor absoluto (VA) e porcentagem, trazendo para isso as subetiquetas apresentadas no Quadro 4 (cf. Capítulo 3).

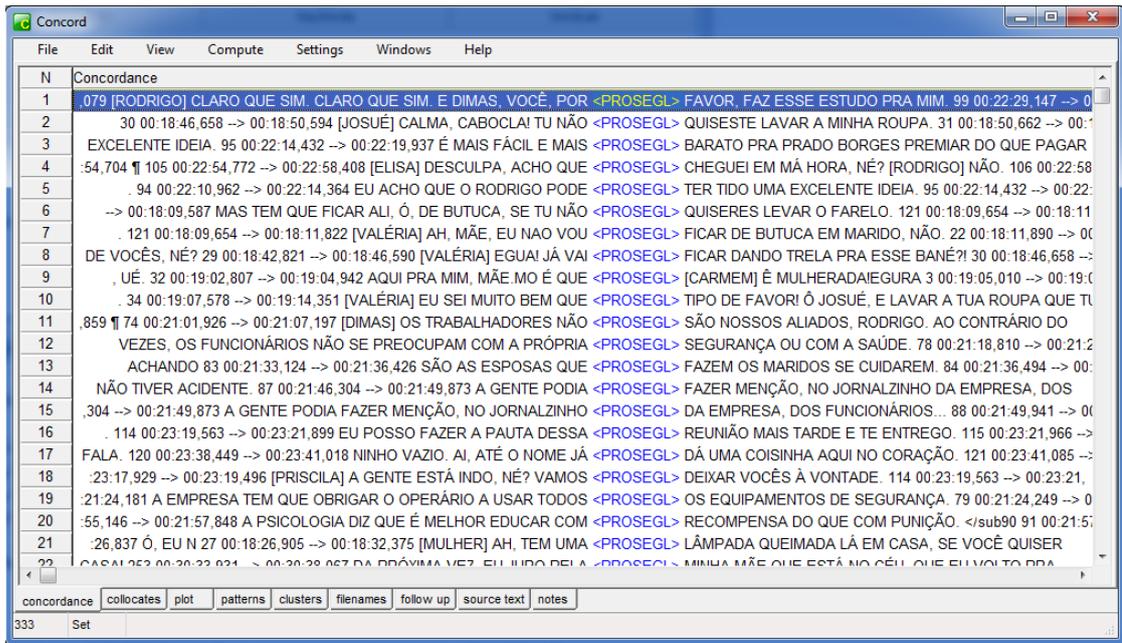
Tabela 4 – VA e porcentagem dos tipos de PROSEGL em *Amor Eterno Amor*

AEA	SV	SN	SP	SUBORD	SAdj	COORD	SAdv
Capítulo 1	27	21	8	7	0	0	2
Capítulo 2	25	20	9	4	1	0	1
Capítulo 3	39	21	6	5	0	2	0
Capítulo 4	7	4	1	3	2	0	0
Capítulo 5	48	43	8	11	3	4	1
TOTAL 1	146	109	32	30	6	6	4
TOTAL 2 (%)	43,8	32,8	9,6	9	1,8	1,8	1,2

Fonte: Elaborada pelo autor

A Figura 13 traz a tela da ferramenta *concord* após a busca pela concordância <PROSEGL> no *corpus*.

Figura 13 – Tela da *concord* após a busca pela etiqueta <PROSEGL>



Fonte: Elabora pelo autor

Nas subseções a seguir, irei apresentar os problemas de segmentação linguística identificados através das ocorrências de quebra de linha nas legendas em que ou as estruturas de sintagmas e/ou de orações foram separadas de forma indevida. Os resultados da análise descritivo-classificatória dos problemas identificados também serão apresentados, o que ocorrerá a partir da alocação dos problemas por tipo de sintagma e de oração em diferentes subseções. Sendo impossível exemplificar os 333 problemas de segmentação linguística apontados pela análise do *corpus* a partir do *Wordsmith Tools 5.0*, dada a restrição de espaço, optei por trazer exemplos que mostrem apenas uma só categoria de problema por tipo de sintagma e oração. Para cada exemplo, mostro uma proposta de ‘ressegmentação’ que considera não só o aspecto da segmentação, mas vê a legenda como um todo através de suas características linguísticas e técnicas. As partes do texto que deveriam aparecer na mesma linha de legenda foram sublinhadas. Há ainda uma subseção dedicada ao Índice de Frequência Simples (IFS) da categoria PROSEGL no *corpus* e uma subseção dedicada à análise da inter-relação dos parâmetros técnicos de velocidade da legenda e número de linhas da legenda com os problemas de segmentação linguística presentes no *corpus*.

4.1.2.1 Sintagma Verbal (SV)

Os problemas de segmentação linguística no SV acontecem normalmente dentro de um sintagma verbal composto, normalmente composto de seu núcleo preenchido por verbo pleno numa forma nominal – infinitivo, particípio ou gerúndio – e especificado, ou seja, antecedido por um verbo auxiliar, como nos casos da quebra em um SV composto e na quebra **SV+oblíquo+SV**, etc. Outro caso é o da quebra **SV+SAdv**, já que, mesmo não levando em conta constituintes relacionados em uma estrutura verbal complexa, esses dois sintagmas estão fortemente inter-relacionados semanticamente (o segundo modificando o primeiro) que, em situações em que os dois se relacionam e não estão lado a lado em uma linha de legenda, mas em linhas diferentes, creio que a recepção daquela inserção de legenda possa ser comprometida, levando-se em consideração a literatura da área.

Dentre as 146 ocorrências de PROSEGL no nível do sintagma verbal, foram encontrados 73 (22%) do tipo **SV_composto**, 55 (16,5%) do tipo **SV_verbo+SAdv** e 18 (5,4%) do tipo **SV_(verbo)+ oblíquo+SV**.

Para exemplo dos casos de quebra da estrutura **SV_composto** do *corpus*, observemos o exemplo a seguir, presente no Quadro 5. Esse problema ocorreu na legenda 89 do Capítulo 5, que possui um sintagma verbal composto – ‘queria trazer’ – mal segmentado entre as duas linhas da legenda em questão.

Quadro 5 – Problema de segmentação linguística - quebra no sintagma verbal (SV_composto)

TCR ³¹	No. de linhas	Duração	Velocidade	Legenda
00:06:28,754 --> 00:06:31,555	2	2,8s	180ppm (alta)	EU NÃO QUERIA TRAZER PROBLEMA PRA VOCÊS

Fonte: Elaborado pelo autor

Considerando o parâmetro da segmentação linguística, a legenda em questão ressegmentada ficaria da seguinte forma:

Eu não queria trazer
problema para vocês.

³¹ Sigla para *Time Code Reader*, que indica o tempo de entrada e saída da legenda.

Como os 48 caracteres da legenda em 2,8s de exposição em tela estão de acordo com a velocidade de 180ppm (DIAZ-CINTAS; REMAEL, 2007, p. 99), não seriam necessárias outras estratégias, como condensação e a redução (DIAZ CINTAS; REMAEL, 2007) para tornar a legenda adequada, sendo a simples colocação do verbo ‘trazer’ para a primeira o suficiente para tornar a legenda em questão bem segmentada sintagmaticamente. A única modificação linguística seria a substituição do ‘pra’ por ‘para’, mais formal, assim como preferem os surdos, aumentando a legenda em apenas 1 caractere.

Em relação aos problemas de segmentação abarcados pela subetiqueta **SV_verbo+SAdv**, o Quadro 6, logo abaixo, traz um exemplo dessa subetiqueta presente na legenda 164 do capítulo 2 do *corpus*.

**Quadro 6 – Problema de segmentação linguística - quebra no sintagma verbal
(SV_verbo+SAdv)**

TCR	No. de linhas	Duração	Velocidade	Legenda
00:25:44,342 --> 00:25:46,276	2	1,9s	180ppm (alta)	[RODRIGO] O MAURO <u>SE EMPENHOU</u> <u>MUITO</u> NESSA CAUSA.

Fonte: Elaborado pelo autor

No caso dessa legenda, uma boa segmentação linguística significaria deixar na mesma linha ‘se empenhou’ e ‘muito’, SV e advérbio, respectivamente, que se relacionam na oração. Contudo, essa legenda tem uma densidade lexical alta. A mesma possui 47 caracteres em apenas 1,9s de tempo de exposição, enquanto que o permitido para a velocidade de 180ppm seria de cerca de 32-33 caracteres para esse mesmo tempo. Mesmo utilizando a condensação e a redução, dificilmente seria possível chegar ao número de caracteres em questão sem ter o sentido da oração um pouco comprometido, ainda mais devido à proximidade de tempos das falas que antecedem e sucedem essa legenda. Sendo assim, uma possibilidade seria condensar o sentido de ‘se empenhou muito’ através do verbo ‘doar-se’, que tem uma conotação forte, suprindo, assim, um pouco a questão do verbo e advérbio retirados. Outro ponto seria a condensação do SN ‘nessa causa’ por apenas ‘nisso’, por acreditar que, a partir do diálogo presente na cena, seria possível inferir sobre o que ‘nisso’ se refere, não havendo, assim, prejuízo em termos de compreensão ao utilizar essa referência. Abaixo, uma sugestão para a possível legenda, com 32 caracteres.

[Rodrigo]
O Mauro se doou nisso.

Vejam agora o caso das ocorrências da subetiqueta **SV_(verbo)+oblíquo+SV**, presente na legenda 536 do capítulo 5 do *corpus* e exposta no Quadro 7, abaixo.

**Quadro 7 – Problema de segmentação linguística - quebra no sintagma verbal
(SV_(verbo)+oblíquo+SV)**

TCR	No. de linhas	Duração	Velocidade	Legenda
00:33:26,504 --> 00:33:30,773	2	4,2s	145ppm(baixa)	EM TODO CASO, EU <u>VOU</u> <u>TE MANDAR</u> UM OUTRO ARTIGO...

Fonte: Elaborado pelo autor

Podemos observar nessa ocorrência a quebra indevida da estrutura verbal ‘vou te mandar’. Nessa legenda, os parâmetros técnicos estão de acordo com o que apregoa a literatura em legendagem, havendo a sugestão de modificação, para uma adequação total da inserção em questão, apenas no que concerne a ressegmentação linguística da legenda, que ficaria como está abaixo:

Em todo caso, eu
vou te mandar um outro artigo...

4.1.2.2 Sintagma Nominal (SN)

No que diz respeito às categorias de quebra indevida no âmbito do sintagma nominal, foram consideradas as quebras entre o i) Especificador e o núcleo do SN, ii) na estrutura do núcleo do SN, iii) entre o substantivo e o adjetivo que o modifica, iv) além dos casos que há a separação indevida do substantivo com o SP, enquanto Complementador do sintagma nominal, quando o mesmo cumpre função adjetiva.

Dentre os 109 problemas de segmentação linguística encontrados no nível do Sintagma Nominal, 56 (16,8%) foram do tipo **SN_especif+SN**, 28 (8,4%) do tipo **SN_subst+SP**, 19 (5,7%) do tipo **SN_subst+SAdj** e 6 (1,8) do tipo **SN_núcleo**. Para exemplificar os problemas de quebra no sintagma nominal, detalho, novamente, um exemplo de cada tipo de problema relativo às subetiquetas supracitadas.

Quanto à subetiqueta **SN_especif+SN**, o exemplo destacado para análise pertence a legenda de número 95, do capítulo 4 do *corpus*, e está exposto no Quadro 8 abaixo.

Quadro 8 – Problema de segmentação linguística - quebra no sintagma nominal (SN_especif+SN)

TCR	No. de linhas	Duração	Velocidade	Legenda
00:24:31,602 --> 00:24:40,243	2	8,6s	145ppm (Baixa)	CHEGA DE ESCONDER <u>O NOSSO</u> <u>NAMORO</u> , LAÍS. A VIDA É AGORA!

Fonte: Elaborado pelo autor

No caso acima, o problema de segmentação linguística acontece a partir da separação do pronome possessivo ‘nosso’ do núcleo do SN – ‘namoro’. A ressegmentação da legenda acima é simples. O parâmetro da segmentação linguística pode ser facilmente contornado com a colocação de ‘o nosso’ para a linha de baixo da legenda, sem que haja prejuízo de outras características da legenda. Contudo, em uma possível readequação completa da legenda em questão aos parâmetros vigentes de legendagem, a mesma não poderia ficar tanto tempo em tela, absurdos 8,6s, já que o máximo preconizado literatura em TAV é de 6s em tela. Este tempo absurdo não foi ocasionado pelo tempo da fala que a legenda representa, mas por erros na confecção da legenda ou por questões tecnológicas relacionados à transmissão do sinal do *closed caption*. A sugestão de ressegmentação encontra-se logo abaixo.

Chega de esconder
o nosso namoro, Laís. A vida é agora!

Na legenda 370 do capítulo 1, no Quadro 9, encontra-se a ocorrência escolhida para exemplificar a quebra da categoria **SN_subst+SP**.

Quadro 9 – Problema de segmentação linguística - quebra no sintagma nominal (SN_subst+SP)

TCR	No. de linhas	Duração	Velocidade	Legenda
00:35:30,661 --> 00:35:35,131	2	4,4s	145ppm (Baixa)	ELE É <u>DA CIDADE</u> <u>DE RIO FUNDO</u> , MINAS.

Fonte: Elaborada pelo autor

No caso da legenda em questão, ‘da cidade de Rio Fundo’ deveria permanecer junto pela forte relação mantida entre o substantivo ‘cidade’ (além da preposição ‘da’ que o acompanha) e ‘de Rio Fundo’, SP que cumpre função adjetiva nesse caso, qualificando e dando mais detalhes acerca do substantivo. A inadequação dessa legenda também reside em relação a seu ao tempo excessivo em tela (4,4s) ao passo que, na menor velocidade possível (145ppm), seriam necessários apenas 2,6s para a leitura os 38 caracteres da legenda. Esse tempo de exposição muito maior que o necessário pode acabar por causar releituras da legenda e, assim, atrapalhar a recepção do produto audiovisual. Uma maneira de ressegmentar essa legenda em apenas uma encontra-se abaixo.

Ele é
da cidade de Rio Fundo, Minas.

A legenda de número 392 do Capítulo 5 configura-se como um exemplo para a categoria **SN_subst+SAdj**. O exemplo encontra-se no Quadro 10 abaixo.

**Quadro 10 – Problema de segmentação linguística - quebra no sintagma nominal
(SN_subst+SAdj)**

TCR	No. de linhas	Duração	Velocidade	Legenda
00:24:28,366 --> 00:24:31,468	2	3,1s	145ppm (Baixa)	VAMOS DEIXAR <u>OS SENTIMENTOS</u> <u>NEGATIVOS</u> DE LADO...

Fonte: Elaborado pelo autor

Nesse caso, vemos o núcleo do SN e o seu especificador – ‘Os sentimentos’ – separados do complementador do SN, o adjetivo ‘negativos’. Estando os outros parâmetros que não a segmentação linguística em conformidade com o que já foi discutido aqui, a ressegmentação da legenda está presente abaixo.

Vamos deixar
os sentimentos negativos de lado...

O exemplo da subetiqueta **SN_núcleo** provém do capítulo 5, da legenda de número 270, e está presente no Quadro 11.

**Quadro 11 – Problema de segmentação linguística - quebra no sintagma nominal
(SN_núcleo)**

TCR	No. de linhas	Duração	Velocidade	Legenda
00:17:55,506 --> 00:18:00,711	2	5,2s	145ppm (Baixa)	A CULPA FOI NOSSA, DE <u>NÓS</u> <u>DOIS</u> , DA GENTE QUE SE AMA.

Fonte: Elaborado pelo autor

O problema de segmentação linguística ocorre pela separação do núcleo do SN ‘nós dois’ em linhas de legenda distintas. Outro problema está relacionado ao tempo excessivo (5,2s) em que ela fica em tela sem necessidade, já que a legenda consta de apenas 51 caracteres, o que, na velocidade de 145ppm, precisaria de apenas 3,5s. Em relação à ressegmentação linguística, uma proposta seria a da seguinte legenda:

A culpa foi nossa,
de nós dois, da gente que se ama.

4.1.2.3 Sintagma Preposicional

Os problemas de segmentação linguística no SP quebra indevida entre o núcleo do sintagma preposicional e o seu Complementador, que, no caso do *corpus* do estudo ora relatado, só se realizou por substantivo e por verbo.

Dos 32 PROSEGL de quebra de sintagma preposicional, 31 (9,3%) são do tipo **SP_prep+SN** e apenas 1 (0,3%) do tipo **SP_prep+SV**.

Um exemplo de **SP_prep+SN** é o da legenda 131 (Capítulo 3), a qual contém um advérbio de lugar – ‘lá’ – e um sintagma preposicional com função de SAdv igualmente de lugar – ‘na empresa’ – que, juntos, formam uma só estrutura adverbial. Contudo, a quebra ocorre exatamente entre a preposição ‘na’ e o substantivo ‘empresa’. Como os dois segmentos estão em linhas distintas, houve a separação entre a preposição e o substantivo a ela relacionado. Esse está presente logo abaixo, no Quadro 12.

Quadro 12 – Problema de segmentação linguística - quebra no sintagma preposicional (SP_prep+SN)

TCR	No. de linhas	Duração	Velocidade	Legenda
00:26:00,758 --> 00:26:04,794	3	4s	180ppm (Alta)	ENTROU COM UM PROCESSO <u>LÁ NA</u> <u>EMPRESA</u> EM QUE ELE TRABALHAVA, GANHOU UMA GRANA FERRADA.

Fonte: Elaborado pelo autor

Essa legenda possuiu uma alta densidade lexical (compõe-se de 83 caracteres quando o máximo deveria ser 74 caracteres ou 37 caracteres por linha) e ultrapassa o limite de caracteres para 4 segundos, de acordo com a velocidade de 180ppm, que é de 70 (DIAZ-CINTAS; REMAEL, 2007, p. 99). Para adequar-se aos padrões, essa legenda tem que ser condensada e/ou ser ressegmentada em duas legendas. Algumas estratégias de redução do texto da legenda seria retirar o advérbio ‘lá’, redundante no contexto nas legendas, e condensar por substituição sinonímica "uma grana ferrada" por ‘boa grana’. Ao fazer essas readequações, é possível atingir o número de caracteres ideal para 4s de exposição na velocidade de 180ppm (70 caracteres).

Entrou com um processo na empresa
em que trabalhava e ganhou boa grana.

ou

Entrou com um processo
na empresa em que trabalhava

e ganhou boa grana.

Sobre as duas legendas propostas, a primeira teria 4s, 70 caracteres e duas linhas; a segunda, dividida em duas inserções de legenda, teria 2,8s e 50 caracteres na primeira legenda, enquanto que a segunda legenda teria 1,2s e 19 caracteres.

A única ocorrência encontrada no *corpus* da categoria **SP_prep+SV** na legenda 201 do capítulo 1 presente no Quadro 13.

Quadro 13 – Problema de segmentação linguística - quebra no sintagma preposicional (SP_prep+SV)

TCR	No. de linhas	Duração	Velocidade	Legenda
00:16:14,873 --> 00:16:19,710	2	4,8s	145ppm (Baixa)	UMA VONTADE <u>DE</u> <u>ESTAR</u> PERTO, DE AJUDAR.

Fonte: Elaborado pelo autor

Nessa ocorrência, pode-se observar a separação da preposição ‘de’ e do verbo no infinito ‘estar’, causando, assim, um problema de segmentação linguística na categoria **SP_prep+SV**. Há, ainda, um grande tempo de exposição da legenda sem necessidade – apenas 37 caracteres em 4,8s – sendo que na menor velocidade, de 145ppm, são necessários apenas 2,6s para essa quantidade de caracteres. Desse modo, além da ressegmentação linguística, outra mudança em relação a essa legenda seria adequá-la em termos de tempo de exposição, diminuindo esse tempo, se possível, ou ressegmentando a legenda em duas legendas para que o ritmo de fala do personagem fosse respeitado (segmentação retórica). Uma sugestão para a ressegmentação linguística da presente legenda encontra-se logo abaixo:

Uma vontade de estar perto,
de ajudar.

4.1.2.4 Sintagma Adjetival (SAdj)

A quebra indevida na estrutura do sintagma adjetival para efeitos da pesquisa ora relatada ocorreu entre o adjetivo e o sintagma adverbial enquanto seu especificador. A quebra no sintagma adjetival não apareceu muito no *corpus*. Foram encontrados apenas 6 ocorrências (1,8%) da categoria **SAdj_especif+SAdj** no *corpus* inteiro. Um exemplo dessa categoria encontra-se logo abaixo, no Quadro 14, referente à legenda 157 do capítulo 5 do *corpus*.

**Quadro 14 – Problema de segmentação linguística - quebra no sintagma adjetival
(SAdj_especif+SAdj)**

TCR	No. de linhas	Duração	Velocidade	Legenda
00:11:09,334 --> 00:11:15,005	3	5,6s	145ppm(Baixa)	NÓS ESTAMOS <u>BEM MAIS</u> <u>TRANQUILOS</u> VENDO O SENHOR, ASSIM, FORTE, CONTENTE.

Fonte: Elaborado pelo autor

Na ocorrência em questão vemos a separação do SAdv ‘bem mais’, agindo enquanto especificador do SAdj ‘tranquilos’. Para a ressegmentação linguística da legenda em questão, privilegiando uma segmentação linguística adequada, teríamos que fazer alguns ajustes de redução, eliminando repetições e redundâncias para que fosse obtido o exemplo abaixo, de 54 caracteres. Como a velocidade original da legenda e desse exemplo de ressegmentação são baixas, também haveria a possibilidade de transformar essa legenda em duas inserções, não havendo, assim, necessidade de redução do texto da legenda. Nessa segunda possibilidade, a primeira inserção teria 30 caracteres em 1,8s, enquanto que a segunda inserção teria 38 caracteres em 2,2s, totalizando, assim, 68 caracteres em 4s de inserção, enquadrando-se na velocidade de 180ppm, a mais rápida. Pode-se ver, então, que não há a necessidade de um tempo tão grande de exposição (5,6s) para o número de caracteres da legenda de partida e das ressegmentações aqui propostas, pois, segundo estudiosos da área, quando a legenda fica na tela por um período maior do que o necessário, há a possibilidade de que o espectador releia as legendas, dificultando, assim, o processo de recepção do produto audiovisual legendado (DIAZ-CINTAS; REMAEL, 2007, p. 89).

Estamos mais tranquilos
vendo o senhor forte, contente.
Ou

Nós estamos
bem mais tranquilos

vendo o senhor,
assim, forte, contente.

4.1.2.4 Sintagma Adverbial (SAdv)

Os problemas de segmentação linguística no nível do sintagma adverbial acontecem em estruturas adverbiais compostas, isto é, quando há a separação entre um dado advérbio e outro que lhe acrescenta informação. A categoria de quebra no SAdv, representada pela subetiqueta **SAdv_adv+adv**, também apresentou poucas ocorrências no *corpus*. Foram apenas 4 (1,2%). A ocorrência a seguir provém da legenda 184, presente no capítulo 5 do *corpus*, e representada no Quadro 15.

Quadro 15 – Problema de segmentação linguística - quebra no sintagma adverbial (SAdv_adv+adv)

TCR	No. de linhas	Duração	Velocidade	Legenda
00:12:49,801 --> 00:12:54,538	3	4,7s	145ppm(Baixa)	NOSSA, JÁ FOI COMPLICADO PRA GENTE QUE <u>ESTAVA ALI</u> <u>PERTO</u> , EU IMAGINO PRA VOCÊ.

Fonte: Elaborado pelo autor

O problema de segmentação linguística encontrado no fragmentado acima reside na separação inadequada dos advérbios ‘ali’ e ‘perto’. Devido à densidade lexical da legenda, com 75 caracteres, e a dificuldade de ressegmentar a legenda de forma que não houvesse um comprometimento sintático das orações em apenas uma inserção de legenda, penso que a melhor solução para a adequação da segmentação linguística seria dividir a legenda em duas inserções, ficando assim:

Nossa, já foi complicado
pra gente que estava ali perto,

imagino pra você.

4.1.2.5 Orações Subordinadas (SUBORD) e Coordenadas (COORD)

Alguns manuais sobre a prática da legendagem afirmam que, sempre que possível, é necessário deixar cada uma das orações que formam um período composto em uma linha de legenda, de modo que o sentido completo da oração esteja em cada linha (IVARSSON;

CAROLL, 1998). Conciliar essa instrução com uma boa segmentação linguística não foi fácil (ARAÚJO; ASSIS, 2014, p. 177), já que é complicado segui-la durante a legendagem devido aos diversos fatores linguísticos e técnicos que estão em interação na legendagem.

Aliando as considerações de Chaves (2012) e de Perego (2008) sobre a segmentação em orações subordinadas e coordenadas às perspectivas gramático-funcionais de Castilho (2012) acerca dessas orações complexas, os problemas de segmentação no nível das orações subordinadas e coordenadas acontecem quando as conjunções coordenativas e subordinativas – incluindo nestas os pronomes relativos – não aparecem na mesma linha de legenda da oração coordenada ou subordinada que iniciam (ARAÚJO; ASSIS, 2014, p. 177).

Dos 15 PROSEGL em que houve quebra no nível das orações, 6 (1,8%) ocorrem com orações coordenadas e 9 (2,7%) com orações subordinadas. Serão apresentadas aqui uma ocorrência de cada um dos dois tipos de problemas de segmentação em questão.

No que diz respeito às orações coordenadas, vejamos o exemplo da categoria **COORD_conj+oração** da legenda 52 do capítulo 3 do *corpus* no Quadro 16.

Quadro 16 – Problema de segmentação linguística - quebra em oração coordenada (COORD_conj+oração)

TCR	No. de linhas	Duração	Velocidade	Legenda
00:21:56,380 --> 00:22:00,851	3	4,4s	160ppm (Média)	EU ESTAVA TÃO ANSIOSA, QUE EU JÁ ARRUMEI <u>E</u> <u>DESARRUMEI</u> ISSO AQUI MIL VEZES!

Fonte: Elaborado pelo autor

A quebra nessa ocorrência ocorre entre a conjunção e a oração que ela é responsável por conectar, iniciada pelo verbo flexionado ‘desarrumei’. Antes de se ressegmentar essa legenda, é necessário que o tempo máximo de 4 segundos – limite considerado pelo presente trabalho – ao invés de 4,4 segundos seja utilizado, e, dessa forma, um máximo de 70 caracteres seja adotado – máximo permitido para 4 segundos em uma velocidade de 180ppm – ao invés dos 72 caracteres da legenda como foi ao ar pela Rede Globo. Para diminuir esses dois caracteres, a simples omissão de ‘eu’ na oração ‘que eu já arrumei’ bastaria, dado que esse pronome pessoal, que já está presente na primeira oração da legenda, pode perfeitamente ficar implícito sem que haja prejuízo de sentido na oração. Uma ressegmentação que transformasse essa legenda em uma legenda bem segmentada e com

apenas duas linhas – já que este é o padrão recomendando por teóricos em TAV, ficaria da seguinte forma, estando os outros parâmetros linguísticos e os parâmetros técnicos da legenda de acordo:

Eu estava tão ansiosa, que já arrumei
e desarrumei isso aqui mil vezes!

Quanto ao PROSEGL nas orações subordinadas, representadas pela categoria **SUBORD_conj/ pron_rel+oração**, pode-se ver o exemplo³² da legenda 21 (Capítulo 3) presente no Quadro 17.

**Quadro 17 – Problema de segmentação linguística - quebra em oração subordinada
(SUBORD_conj/ pron_rel+oração)**

TCR	No. de linhas	Duração	Velocidade	Legenda
00:20:17,248 --> 00:20:19,583	3	2,3s	180ppm (Alta)	NÃO TEM NADA NESSE MUNDO <u>QUE</u> <u>SUBSTITUA</u> CARINHO DE MÃE, NÉ?

Fonte: Elaborado pelo autor

A quebra aqui ocorre na separação do pronome relativo ‘que’ do restante da oração subordinada por ele introduzida: ‘que substitua carinho de mãe, né?’. Conforme já explicado, essa separação se configura, a partir de ótica da presente pesquisa, como um problema de segmentação linguística na legendagem. Como a legenda possui 57 caracteres, seriam necessários cerca de 3,3s em uma velocidade de 180ppm para que a legenda estivesse adequada. Dessa forma, como estratégia de redução e condensação, sugiro a substituição de alguns elementos do texto por outros que possuam significado similar além, principalmente, da transformação da oração principal e da oração subordinada presentes na legenda em questão em apenas uma oração, deixando o texto da legenda com 42 caracteres. Uma sugestão para a ressegmentação da legenda 21 seria:

Nada no mundo
substitui carinho de mãe, né?

³² Esse exemplo está presente em Araújo e Assis (2014, p. 177), mas a proposta de ressegmentação aqui presente considera o aspecto técnico da velocidade da legenda em sua concepção.

Os resultados relativos aos dados apresentados apontam para uma recorrência dos padrões de problemas de segmentação linguística na LSE de ficção, como em filmes (CHAVES, 2012; ARRAES, 2015) e telenovelas (ARAÚJO; ASSIS, 2014; GABRIEL, 2015). As quebras no nível dos sintagmas verbal e nominal se configuram como os mais problemáticos nesses estudos.

4.1.3 Índice de Frequência Simples e problemas de segmentação linguística no *corpus*

Para verificar a relação do número de problemas de segmentação linguística com o número de *tokens* do *corpus*, fiz uso de um recurso estatístico chamado Índice de Frequência Simples³³, que serve para se compensar os números de palavras diferentes da LSE dos 5 capítulos diferentes analisados. Esse recurso estatístico foi incorporado ao Projeto CORSEL em 2014, com o projeto em andamento, de modo que fosse possível comparar os sub-*corpora* localizados no seio do projeto, os quais tinham extensões (tamanhos) distintos, sendo assim difícil, às vezes, comparar os resultados provenientes da análise de sub-*corpora* entre si por meio de porcentagens (ARAÚJO, 2015; ARRAES, 2015).

O cálculo do IFS mostra a ocorrência da categoria a ser analisada – um dado traço qualquer, no caso da presente pesquisa, a etiqueta PROSEGL, que representa os problemas de segmentação linguística – a cada mil *tokens*. A Tabela 5, abaixo, traz, além das informações em relação ao valor absoluto (VA) (de cada capítulo separadamente e de todos juntos) e o IFS (de cada capítulo separadamente e da média de todos) das ocorrências de PROSEGL, o número de legendas e *tokens* da LSE de cada capítulo da novela e a soma de todos.

Tabela 5 – IFS e VA dos PROSEGL em *Amor Eterno Amor*

AEA	Nº de legendas	<i>Tokens</i>	PROSEGL	
			VA	IFS
Capítulo 1	384	2.966	65	21,9
Capítulo 2	347	2.848	60	21
Capítulo 3	294	2.534	73	28,8
Capítulo 4	238	1.695	17	10
Capítulo 5	586	4.813	118	24,5
TOTAL	1849	14.856	333	22,4

Fonte: Elaborada pelo autor

³³ Número de ocorrências de uma dada categoria dividido pelo número total de palavras do *corpus* usado como referência, com o resultado multiplicado por 1.000 (PRAXEDES FILHO; MAGALHÃES, 2013, p. 80).

Os Índices de Frequência Simples dos 5 capítulos de *Amor Eterno Amor*, no que diz respeito aos problemas de segmentação linguística encontrados, foram relativamente próximos, com quatro deles variando entre 21 e 28,8, sendo IFS dos cinco capítulos 22,4, ou seja, 21,2 ocorrências de PROSEGL a cada mil palavras do *corpus* analisado. O capítulo 4, contudo, o menor em termos de *tokens* e de número de legendas, apresentou também o menor número de problemas de segmentação linguística tanto em valores absolutos (17 ocorrências), quanto em relação ao IFS (10), comportamento que fugiu do padrão encontrado nos outros 4 capítulos, de uma quantidade considerável de problemas de segmentação em valores absolutos e IFS na casa dos 20, como mencionado acima.

É interessante ressaltar que o IFS encontrado em *Amor Eterno Amor* em relação aos problemas de segmentação linguística foi quase o mesmo apresentado no sub-*corpus* ‘Novelas’ do Projeto CORSEL (19,8), mas ligeiramente superior, como pode-se ver na Tabela 6.

Tabela 6 – IFS e VA dos PROSEGL em por gênero no projeto CORSEL

Gênero	Nº de legendas	<i>Tokens</i>	PROSEGL	
			VA	IFS
Documentários	1467	11.821	577	48,8
Novelas	7703	60.710	1207	19,8
Seriados	1596	11.625	169	14,5

Fonte: Arquivos do CORSEL

Em relação ao IFS dos gêneros analisados pelo CORSEL, pode-se observar a relevância dos problemas de segmentação linguística principalmente nos resultados provenientes do gênero ‘Documentários’, sendo a categoria Novelas o gênero de LSE com o segundo maior IFS em termos de ocorrência desse tipo de problema de segmentação.

4.1.4³⁴ Relação entre problemas de segmentação e parâmetros técnicos

Esta subseção traz a relação dos problemas de segmentação linguística encontrados com os parâmetros técnicos analisados: a velocidade da legenda, na qual estão diretamente relacionados o número de caracteres, e o número de linhas por legenda.

³⁴ Como destacado em Araújo e Assis (2014, p. 178), “Para efeitos de análise, apenas alguns dos parâmetros técnicos foram levadas em conta, como caracteres por linha, número de linhas e velocidade da legenda, por apresentarem relações mais intrínsecas com o parâmetro da segmentação linguística. Os outros parâmetros etiquetados foram importantes para a localização da legenda no *corpus*, como a etiqueta que indica o número da legenda e a que delimita os tempos inicial e final da legenda”.

A Tabela 7 mostra a relação dos problemas de segmentação linguística encontrados no *corpus* com o número de linhas das legendas problemáticas.

Tabela 7 – Relação entre PROSEGL e número de linhas em *Amor Eterno Amor*

AEA	Legendas de 2L	Legendas de 3L
Capítulo 1	32	33
Capítulo 2	25	35
Capítulo 3	27	45
Capítulo 4	7	10
Capítulo 5	59	59
TOTAL 1	150	183
TOTAL 2 (%)	45	55

Fonte: Elaborada pelo autor

Os resultados da análise quantitativa dos dados mostraram um número maior de problemas de segmentação linguística aparecendo em legendas de 3 linhas, com 183 ocorrências ou 55% do total, enquanto que houve 150 ocorrências – ou 45% do total – problemas em legendas de 2 linhas. Esses resultados reforçam aqueles encontrados por Assis (2013), Araújo e Assis (2014) na análise da LSE de um capítulo de *Amor Eterno Amor*, que, em primeira instância, também observaram uma maior quantidade de problemas de segmentação linguística em legendas de 3 linhas, assim como Gabriel (2015) na análise de problemas de segmentação linguística em outra telenovela.

Abaixo, a Tabela 8 mostra os problemas de segmentação linguística em relação ao nível de velocidade da legenda, outro parâmetro técnico analisado. Os dois parâmetros técnicos – velocidade da legenda e número de linhas – analisados de forma concomitante ao parâmetro da segmentação podem auxiliar a compreender o contexto mais propício no qual surgem os problemas de segmentação na LSE da *Amor Eterno Amor*.

Tabela 8 – Relação entre PROSEGL e velocidade da legenda em *Amor Eterno Amor*

AEA	145ppm Vel. Baixa	160ppm Vel. Média	180ppm Vel. Alta
Capítulo 1	36	1	28
Capítulo 2	21	4	35
Capítulo 3	17	3	52
Capítulo 4	9	1	7
Capítulo 5	60	5	53
TOTAL 1	143	14	176
TOTAL 2 (%)	42,9	4,2	52,9

Fonte: Elaborada pelo autor

Os dados dispostos na Tabela 8 mostram que, assim como em Chaves (2012), Assis (2013) e Araújo e Assis (2014), Gabriel (2015) e Arraes (2015), os resultados mostraram um maior número de ocorrências de problemas de segmentação linguística em legendas com velocidade alta (180ppm), com 176 ocorrências ou 52,9% do total. Em ordem decrescente, os problemas também apareceram em legendas de velocidade baixa (160ppm), com 143 ocorrências ou 42,9% do total, e, por último e de forma pouco representativa em número absolutos, em legendas com velocidade média (145ppm), com 14 ocorrências ou 4,2% do total.

Os resultados mostram uma predominância de problemas de segmentação linguística em legendas de 3 linhas e de velocidade alta, reforçando os resultados alcançados por Assis (2013), Araújo e Assis (2014) e Gabriel (2015) na análise da LSE de telenovelas. Um resultado curioso a ser destacado, da mesma maneira, é o fato de as legendas de duas linhas e com a velocidade baixa, considerados padrões que garantiriam uma boa recepção (D'YEDEWALLE ET AL, 1987; DIAZ-CINTAS E REMAEL, 2007) também apresentarem consideráveis problemas de segmentação linguística.

4.2 RELENDAGEM DE UM TRECHO DE *AMOR ETERNO AMOR*

Nesta seção do capítulo de resultados, proponho a relegendagem de um trecho da LSE de *Amor Eterno Amor* pertencente ao *corpus*. O vídeo que acompanha essa LSE trata-se de um recorte do *corpus*, de 10 minutos e 9 segundos de duração, o qual foi retirado das cenas finais do capítulo 5 do *corpus*. A LSE que acompanha o vídeo em questão, e que foi analisada e relegendada – quando necessário – na proposta desenvolvida, é composta 154 legendas, com 36 problemas de segmentação em sua totalidade.

A proposta de relegendagem seguiu os parâmetros técnicos, linguísticos e paralinguísticos provenientes da literatura em LSE-TAV e das pesquisas exploratórias e descritivas desenvolvidas pelo grupo LEAD-UECE desde 2002 com surdos do Ceará e de todo o Brasil (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004, 2007, 2008; NASCIMENTO; ARAÚJO, 2011; CHAVES, 2012; ASSIS, 2013; ARAÚJO; ASSIS, 2014; GABRIEL, 2015; ARRAES, 2015).

Para tornar esta subseção mais dinâmica e menos cansativa, decidi apresentar as legendas – e propostas de novas legendas – que pudessem exemplificar o *modus operandi* dos legendistas do Grupo LEAD – ao qual me incluo – no que concerne o lidar com as limitações e restrições tradutórias da LSE, relacionando, principalmente, sempre que possível, as

decisões sobre as questões técnicas e linguísticas da LSE que privilegiassem o parâmetro linguístico da segmentação, apontado como o principal responsável pela boa recepção dos telespectadores surdos, mesmo em legendas com velocidade de 180ppm. Esta velocidade, inclusive, foi a máxima utilizada como parâmetro para a confecção das legendas aqui propostas.

4.2.1 Parâmetros Paralinguísticos

Nesta subseção, apresento uma pequena análise sobre as questões tradutórias envolvidas na identificação de falantes e na tradução de efeitos sonoros (ruídos e sons) e música na LSE relegendada. Apresento, primeiramente, as escolhas tradutórias feitas na versão de partida para, depois, contrastando-a ou ratificando-a com o que é considerado adequado em LSE, como exposto na fundamentação teórica da presente pesquisa, propor a minha versão – que não é unânime, sendo, inclusive, passível de contestação em alguns momentos, dado o caráter idiossincrático da tradução – e justificar o caminho percorrido pela LSE proposta. A questão da segmentação linguística, mote principal deste trabalho, não está tão intrinsicamente conectada aos aspectos paralinguísticos a serem discutidos aqui, mas como estes são aspectos pertinentes à LSE, e aqui tento discutir todas as implicações que podem ocorrer durante a feitura dessa relegendagem, vale a pena fornecer a reflexão em torno de como devem/deveriam se orientar as traduções em torno delas.

A questão da identificação dos falantes no excerto analisado traz algumas considerações a respeito desse parâmetro paralinguístico característico da LSE. O trecho relegendado conta com algumas cenas de diferentes núcleos dentro da narrativa da telenovela. A LSE de partida seguiu os parâmetros indicados pela literatura em TAV e aos quais os surdos participantes das pesquisas exploratórias da UECE (FRANCO e ARAÚJO, 2003; ARAÚJO, 2004; 2007; 2008; NASCIMENTO; ARAÚJO, 2011): identificando os falantes através de colchetes e havendo a identificação sempre que houvesse troca do turno de fala entre pelos menos 2 pessoas na cena. Nas 5 cenas principais do trecho de 10 minutos e 9 segundos (1– Laís e Julinho no acampamento; 2 – Reunião de várias pessoas preocupadas com o desaparecimento de Laís e Julinho; 3 – Conversa de pai e filha entre Gabriel e Miriam; 4 – conversa entre Fernando e sua mãe; 5 – Dimas na delegacia), apenas na cena número 2 há mais de duas pessoas trocando turnos de fala, ainda assim em apenas alguns momentos. A maior parte dos diálogos é estabelecida com apenas 2 personagens bem definidos em tela,

sendo, através dos enquadramentos bem próximos realizados pela câmera, característica da telenovela, bem claro quem detém o turno de fala na maioria das cenas.

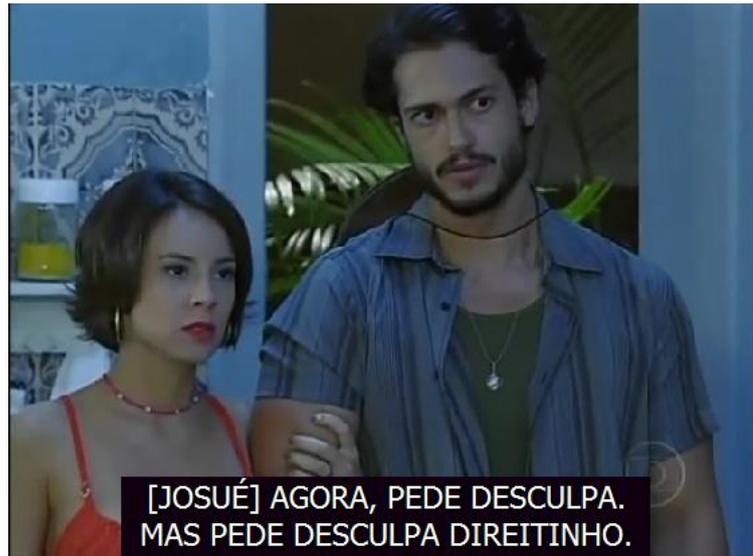
Contudo, a proposta para a identificação dos falantes na relegendagem diferiu, ligeiramente, daquela seguida pela LSE de partida e pela literatura em TAV. Além da identificação do falante quando havia mais de 2 pessoas em tela, o que poderia gerar confusão para o surdo em se tratando de referenciação, sempre que apenas duas pessoas estavam em tela e era claro quem detinha o turno de fala, aquele personagem não foi identificado através de colchetes. Em minha opinião, quando está bem claro quem detém o turno de fala do diálogo em cena, não há a necessidade de se nomeá-lo em toda ocasião em que ocorrer o seu turno de fala, já que uma legenda confeccionada a partir de parâmetros técnicos e linguísticos adequados vai permitir ao telespectador o cotejo adequado entre imagem e legenda, ocorrendo a referenciação correta de qual dos dois personagens em tela, a partir do movimento da boca e do foco em seu rosto, está falando naquele exato momento. Outra questão que pode parecer menor, mas que deve ser levada em conta, é a alta densidade lexical presente em uma LSE de um gênero como uma telenovela. Com a necessidade, no caso do trecho relegendado, de se adequar à velocidade de legenda máxima bem recebida pelos surdos brasileiros (NASCIMENTO; ARAÚJO, 2011), a de 180ppm, há a necessidade de se recorrer à estratégias que diminuam o texto escrito em tela, para além daquelas corroboradas pelos parâmetros de redução e condensação. A não identificação em casos redundantes dentro da condição supracitada (2 falantes em tela em que, a partir de um enquadramento próximo da câmera, seja claro quem detém o turno de fala) pouparia preciosos caracteres no processo de legendagem. É claro, contudo, que essa é uma proposta ainda não testada e que necessita ser objeto de pesquisas de recepção no futuro.

Um exemplo dessa “não identificação” encontra-se abaixo. Na cena em questão, Josué e Valéria trocam turnos de fala e é bem nítido quando cada um dos dois está falando. Na Figura 14³⁵ está a legenda de partida e na Figura 15³⁶ encontra-se a proposta de relegendagem.

³⁵ Os *printscreens* da LSE de partida acompanhada do seu respectivo vídeo foram tiradas no programa *Subtitle Workshop* 6.0B, o que justifica o fato de a legenda estar centralizada, diferentemente da transmissão original, na qual a mesma é alinhada à esquerda.

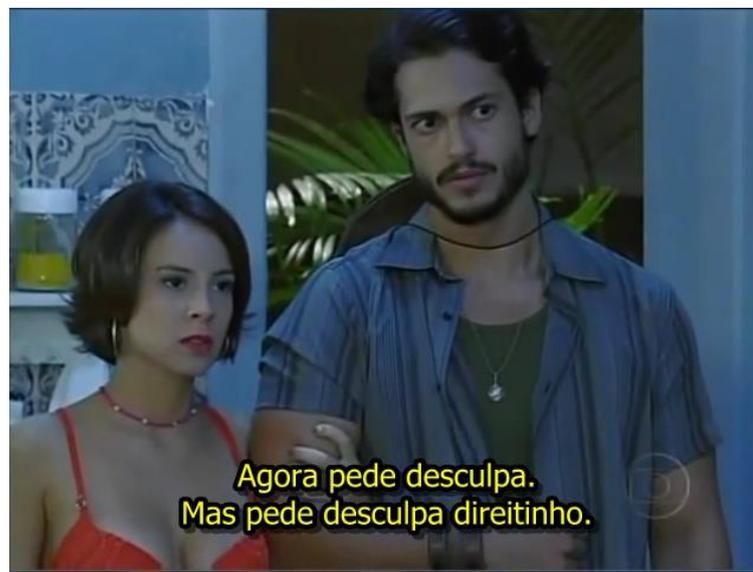
³⁶ Os *printscreens* das propostas de relegendagem da LSE acompanhada do vídeo original traz os parâmetros técnicos e linguísticos editados conforme os preceitos aqui estabelecidos. Essa legenda foi colada no vídeo com o auxílio do programa *freeware* AVI ReComp. O vídeo com a legendagem proposta encontra-se no formato de DVD enquanto Apêndice deste trabalho.

Figura 14 – Exemplo de identificação de falante desnecessária



Fonte: Elaborada pelo Autor

Figura 15 – Proposta de relegendagem sem a identificação do falante

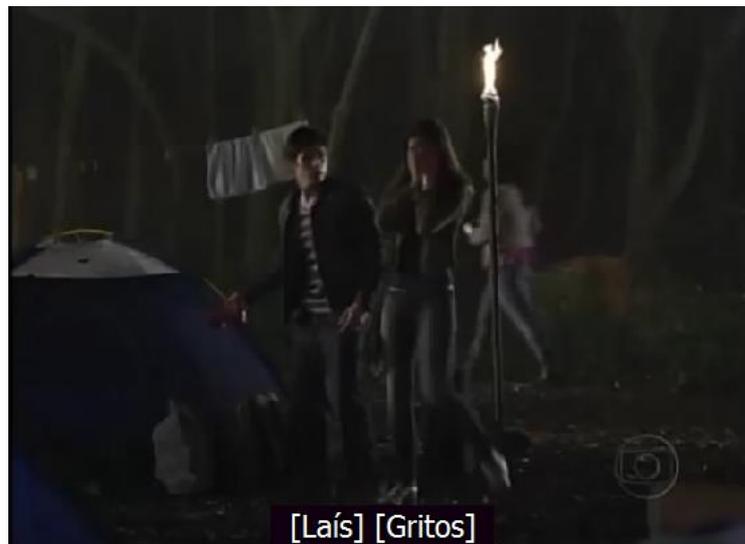


Fonte: Elaborada pelo autor

O trecho da LSE utilizado como *corpus* não apresentou uma grande quantidade de tradução de efeitos sonoros e música, até mesmo porque no trecho da telenovela em questão não há uma presença marcante desses componentes sonoros na composição narratológica do excerto. É válido ressaltar que as concepções tradutórias em relação a esses elementos paralinguísticos em questão foram retirados, em grande parte, de Nascimento (2013), sobre a tradução de efeitos sonoros e música em filmes brasileiros.

Em relação aos ruídos, poucos são os que detêm influência na narrativa da telenovela e, dessa maneira, realmente, deviam ser traduzidos em forma de legenda, de forma a auxiliar o telespectador surdo a capturar a importância de um determinado som na construção da ficção. Exemplos de ruídos relevantes no excerto em questão são as legendas [Gritos] e [Grito], em uma cena em que Laís gritava por ter visto um sapo em sua barraca e que, pela escuridão da fotografia da cena, não era totalmente perceptível que, de fato, a personagem gritava (mesmo que fosse minimamente possível ver a abertura da sua boca), não sendo assim um som “redundante” e, dessa forma, necessária a sua tradução, como se vê na Figura 16 abaixo.

Figura 16 – Legenda [Gritos]



Fonte: Elaborada pelo autor

Ainda em relação aos ruídos na LSE em questão, houve apenas duas traduções de ruídos acrescentados em minha proposta de relegendagem. Uma delas foi o caso da legenda [Gracinha] [Risos], em que a personagem Gracinha, após fazer um comentário jocoso à respeito das habilidades de “domador” de Josué, complementa com risos a verdadeira intenção do seu comentário em um momento em que a personagem se encontra *offscreen*, ou seja, fora de tela. Abaixo, no Quadro 18, o trecho relegendado em que ocorreu a proposta de legenda em questão.

Quadro 18 – Trecho relegendado da fala de Gracinha

TCR	Legenda
00:03:17,843 --> 00:03:20,777	Eu achando que só tu, maninho, que amansava onça.
00:03:20,877 --> 00:03:23,671	Mas o Josué... Estás podendo, hein, mano.
00:03:23,971 --> 00:03:25,347	[Gracinha] [Risos]

Fonte: Elaborado pelo autor

O outro caso é passível de discussão, em minha opinião. Trata-se da proposta de legenda [Barulho de sapos], inserida em um momento antes de Laís romper-se em gritos, exatamente no momento em que ela encontra alguns sapos dentro de sua barraca e, ao mesmo tempo em que a câmera enfoca os anfíbios, há o som dos seus coaxos. A primeira questão a ser debatida é a necessidade ou não da legendagem desse som, já que é possível ver o sapo em tela e presumir que a reação de Laís ao gritar seja motivada pela presença do bicho. Contudo, a meu ver, a tradução desse som reforça a presença do bicho na cabana e contribui para a reação de Laís. E, uma vez que não é presumível o barulho feito pelos sapos apenas pela imagem, a tradução desse som se justifica nesse ponto. Outra questão seria a forma como nomear o som feito pelos sapos. O Priberam Dicionário³⁷ traz a seguinte definição para coaxar: 1. Gritar (a rã ou o sapo). Sendo assim óbvia a existência de um nome para o som produzido por esses bichos. Contudo, a meu ver, a opção por [Os sapos coxam] ou [Coaxar de sapos] não seria de compreensão tão automatizada por parte dos surdos, já que é uma palavra pouco difundida até mesmo entre os falantes nativos de português. Dessa maneira, optei por [Barulho de sapos] de forma a explicitar ainda mais um som que não possui grande difusão no que concerne o seu nome “oficial”. Faz-se necessário ressaltar a maleabilidade que permeia as questões envolvidas no processo tradutório. Há a necessidade, por certo, de profissionais se pautarem na literatura produzida dentro da subárea LSE/TAV no momento do exercício do ofício, sem, entretanto, esquecer que sempre haverá espaço para as idiosincrasias do tradutor/legendista, mesmo em um processo tradutório tão restringido pelas circunstâncias espaço-temporais como é a legendagem. Abaixo, imagem da cena em que está presente a legenda em questão, na Figura 17.

³⁷ <https://www.priberam.pt/DLPO/coaxar>. Acesso em: 6 de Novembro de 2015.

Figura 17 – Legenda [Barulho de sapos]



Fonte: Elaborada pelo autor

Em relação às traduções de música na LSE de partida, houve 6 traduções de música, todas marcadas pelo uso da colcheia, normalmente utilizada como símbolo de música na LSE. Contudo, o uso da colcheia não atrela a música à diegese do filme, condição *sine qua non* para a tradução de música na LSE, conforme apontado pelas pesquisas de recepção realizadas com surdos (NASCIMENTO; ARAÚJO, 2011) e ressaltado por Nascimento (2013). Em minha opinião, em apenas 1 das 6 ocorrências de tradução de música no *corpus* em questão havia a necessidade de se legendar a música, pois a música acrescentava informação à narrativa da história e, sendo assim, deveria ser ressaltada em forma de legenda.

A legenda em questão trata-se da legenda [Música de suspense], por mim assim traduzida. Essa legenda aparece no momento em que, no final do capítulo, Dimas Sobral vai até a delegacia pra relatar algo sobre sua própria esposa, Melissa Sobral. Essa legenda tem influência sobre a narrativa da cena em questão, pois constrói o clima de suspense que encerra o capítulo. Abaixo o momento em que aparece a legenda em questão, na Figura 18.

Figura 18 – Legenda [Música de suspense]



Fonte: Elaborada pelo autor

4.2.2 Relegendagem dos parâmetros técnicos e linguísticos

Esta subseção pretende abordar as questões levadas em consideração para a proposta de relegendagem no que diz respeito aos aspectos técnicos e linguísticos da LSE, correlacionando-os, principalmente, com o parâmetro da segmentação linguística. Não é o meu propósito aqui expor todos os pormenores envolvidos no fazer da LSE, mas sim ressaltar o que vem sendo discutido desde o início deste trabalho sobre o que realmente pode garantir uma boa recepção ao telespectador surdo em termos de LSE.

Não há, de fato, um modo definitivo, ou uma cartilha sagrada, sobre quais parâmetros devem ser abordados primeiramente no passo-a-passo de uma legendagem, dado que nessa modalidade de tradução audiovisual parâmetros técnicos, linguísticos e paralinguísticos ocorrem simultaneamente. Primeiramente, acerca dos parâmetros técnicos, há aos parâmetros de velocidade da legenda, número de linhas, número de caracteres por linha, tempo máximo e mínimo de exibição da legenda em tela etc.

O sincronismo entre falas e legendas esteve deficiente na LSE de partida, com um adiantamento da legenda em relação à fala de aproximadamente um minuto. É importante ressaltar que esse atraso pode ter sido causado pela ferramenta com a qual a legenda foi gravada, no caso, o *CCExtractor*. Contudo, ao atrasar toda a legendagem através do programa *Subtitle Workshop* até o momento em que houvesse o sincronismo entre a primeira

fala e a primeira linha de legenda, pude observar que, ao fazer isso, todas as inserções de legendas subsequentes ficaram sincronizadas, deixando margem para também acreditar que houve falta de sincronismo em relação à transmissão original da TV responsável, já que o programa conseguiu captar o intervalo entre as legendas como ele de fato existiu.

Nesse quesito, foi interessante observar um padrão de pausas entre legendas (intervalos) no trecho de partida. Esses intervalos, majoritariamente, variavam entre 0,67 e 0,68, o que parece ser a pausa mínima utilizada pelo canal em questão. Abaixo, na Figura 19, a recorrência desses intervalos está destacada.

Figura 19 – Tela do *Subtitle Workshop* com os intervalos da LSE de partida destacados

Pause	Show	Hide	Dura...	Text
0,068	00:36:03,994	00:36:06,095	2,101	< />[FERNANDO] VEM CÁ, E O PAPI?< />
0,068	00:36:06,163	00:36:09,298	3,135	DEIXA EU FALAR COM ELE. ELE DEVE ESTAR NUM HUMOR MELHOR DO QUE O SEU. ELE TÁ AÍ?
0,068	00:36:09,366	00:36:13,069	3,703	[MELISSA] PAPI ESTÁ NO SOFÁ DA SALA, E EU NÃO VOU CHAMAR.
0,068	00:36:13,137	00:36:18,141	5,004	[FERNANDO] POXA, MÃE! QUE MALDADE! VOCÊ JOGOU O VELHO DE NOVO PRO SOFÁ DA SALA?
0,067	00:36:18,208	00:36:19,776	1,568	O QUE FOI QUE ELE FEZ DESSA VEZ?
0,067	00:36:19,843	00:36:23,946	4,103	[MELISSA] SEU PAI CONSEGUE SER INSUPORTÁVEL 99% DO TEMPO.
0,068	00:36:24,014	00:36:26,649	2,635	[FERNANDO] MAS NESSE 1% DEVE SER UMA MARAVILHA, NÉ?
0,068	00:36:26,717	00:36:30,086	3,369	PORQUE VOCÊS DOIS ESTÃO CASADOS ATÉ HOJE.
0,068	00:36:30,154	00:36:34,824	4,670	[MELISSA] AMANHÃ, EU TE TELEFONO, TÁ BOM? BOA NOITE E DURMA COM OS ANJOS.
0,068	00:36:34,892	00:36:36,292	1,400	E ME DEIXA EM PAZ.
9,878	00:36:46,170	00:36:47,904	1,734	[DIMAS] POR GENTILEZA, POLICIAL.
0,067	00:36:47,971	00:36:51,507	3,536	[POLICIAL] POIS NÃO?
0,068	00:36:51,575	00:36:57,080	5,505	[DIMAS] EU POSSO CONVERSAR COM O DELEGADO RESPONSÁVEL?
0,067	00:36:57,147	00:36:59,481	2,334	É SOBRE A SENHORA MELISSA SOBRAL.
0,068	00:36:59,549	00:37:01,184	1,635	¶

Hide:	B <i>I</i> <u>U</u>	Text (18/29/22=69 characters):
24,755	00:35:29,558	<p style="text-align: center;">VOCÊ TEM QUE FAZER EXATAMENTE O QUE A MAMÃE FALA E O QUE A MAMÃE MANDA.</p>
00,068	00:00:04,803	
#563 / 586 SubRip		

Fonte: Elaborada pelo autor

O grupo LEAD tem como padrão o uso de, no mínimo, 100 milissegundos para o intervalo entre legendas. Com respaldo na literatura da área (cf. Subseção 2.1), acreditamos que um número inferior a isso possa causar um desconforto na percepção do olhar do usuário ao fazer com que a legenda apenas “pisque” em tela, sendo 100 milissegundos tempo razoável para que o mesmo consiga diferenciar o desaparecimento de uma inserção de legenda e o aparecimento de outra. Sendo assim, utilizando um recurso do *Subtitle Workshop*, coloquei a LSE de partida em um intervalo mínimo de 100 milissegundos entre as legendas, fazendo, após isso, apenas pequenos ajustes em relação ao tempo inicial e final de cada legenda, alcançando assim, o sincronismo almejado pela legendagem. A segmentação retórica, que é alcançada através da divisão do texto em tela de acordo com o fluxo de fala e que também se relaciona à marcação dos tempos, será exemplificada a seguir.

O Quadro 19 traz um exemplo de segmentação retórica realizada em minha proposta de tradução. No exemplo em questão, a legenda, já inadequada do ponto de vista técnico dada a presença de 3 linhas ao invés de 2, fica em tela, de forma desnecessária, por cerca de 6,8 segundos, tempo esse que excede o limite de 4 segundos seguido pelo Grupo LEAD, sendo assim considerado deficiente para uma boa recepção da LSE.

Quadro 19 – Trecho da LSE de partida com problema de segmentação retórica

TCR	Duração	Nº de caracteres	Velocidade da legenda	Legenda
00:31:56,380 - -> 00:32:03,185	6,8s	62c	9,1cps (baixa)	E AÍ, ELE PASSOU POR AQUELA EXPERIÊNCIA DE QUASE MORTE... A EQM.

Fonte: Elaborado pelo autor

Isso acontece mesmo havendo uma pausa considerável antes que a personagem fale a expressão “EQM”, pausa essa explicitada pela presença das reticências. Esse tempo ainda é desnecessário devido a baixa média de caracteres por segundo (9,1), o que sustenta a minha proposta de relegendagem contida no Quadro 20, com duas legendas, ao invés de 1, resolvendo, assim, 3 problemas técnicos de uma só vez: o tempo que excede os 4 segundos em tela, a quantidade de 3 linhas na legenda e, por fim, a falta de uma segmentação retórica adequada, que siga o fluxo da fala da personagem, no trecho destacado.

Quadro 20 – Proposta de relegendagem de LSE com problema de segmentação retórica

TCR	Duração	Nº de caracteres	Velocidade da legenda	Legenda
00:05:02,125 - -> 00:05:05,254	3,1s	52c	16,7cps (média)	E aí ele passou por uma experiência de quase morte...
00:05:05,454 - -> 00:05:06,651	1,2s	6c	5cps (baixa)	A EQM.

Fonte: Elaborado pelo autor

As estratégias de redução e condensação aplicadas no processo de relegendagem foram aplicadas visando a boa segmentação linguística da LSE. Contudo, houve casos em que a necessidade de se reduzir/condensar o texto era motivado pela velocidade da legenda (mais caracteres do que o tempo disponível em tela para a velocidade máxima de 180ppm). Um exemplo disso está presente no Quadro 21, logo abaixo.

Quadro 21 – Exemplo de redução/condensação na proposta de relegendagem

TCR	Nº de caracteres	Velocidade da legenda	Legenda
3,9s	74c	19cps (alta)	[JULINHO] QUE É ISSO? NÃO FOI ISSO DAQUI QUE A GENTE COMBINOU, NÃO, BROTHER!
3,9s	55c	14,1cps (baixa)	[Julinho] Que é isso? Esse não foi o combinado, brother!

Fonte: Elaborado pelo autor

Como a legenda em questão possuiu mais caracteres (74) do que o tempo máximo de 3,9 segundos permitia na velocidade de 180ppm (cerca de 69), a oração ‘Não foi isso daqui que a gente combinou, não, brother!’, perfazendo um total de 54 caracteres, se transformou, através da condensação, em ‘Esse não foi o combinado, brother’, de 34 caracteres, sem que houvesse prejuízo do conteúdo informacional da legenda, ao se retirar uma dupla negação, assim como transformando uma oração complexa, com dois verbos, em uma oração simples de apenas um verbo. Dessa forma, a legenda que possuía 74 caracteres e uma média de 19 caracteres por segundo, passou a ter 55 caracteres em uma média de 14,1 caracteres por segundo. A condensação necessária à LSE em questão também tornou possível que as 3 linhas da legenda se transformassem em 2.

Outro exemplo de condensação e de redução está logo abaixo, no Quadro 22, em que há a presença de absurdos 84 caracteres na inserção. Essa legenda ainda apresenta uma média bem alta de caracteres por segundo para o tempo disponível em tela (3,8s) após os ajustes de sincronismo realizados, acima do limite de 180ppm. Nesse exemplo, o trecho ‘O quê que você acha de voltar’, de 29 caracteres, foi condensado e ficou como ‘Que tal voltar’, de 14 caracteres. A estratégia de redução adotada foi transformar o nome da cidade ‘Rio de Janeiro’ em um forma reduzida pela qual ela também é bastante conhecida³⁸ (‘Rio’). Essa redução só foi realizada após a constatação de que dificilmente haveria uma ambiguidade entre o nome reduzido da cidade e o nome do curso natural de água. Utilizando essas duas estratégias, o número de caracteres passou para 59, atingindo uma velocidade confortável de 15,5 caracteres por segundo.

³⁸ Inclusive, ao se buscar pela palavra ‘Rio’ no Google, a primeira página da busca só apresenta resultados relativos à cidade, ao filme homônimo que se passa na cidade e aos Jogos Olímpicos de 2016, a serem realizados também na cidade.

Quadro 22 – Exemplo de redução/condensação na proposta de relegendagem

TCR	Nº de caracteres	Velocidade da legenda	Legenda
3,8s	84c	22,1cps (alta)	O QUÊ QUE VOCE ACHA DE VOLTAR A CUIDAR DE BÚFALOS NUMA FAZENDA AQUI NO RIO DE JANEIRO?
3,8s	59c	15,5cps (baixa)	Que tal voltar a cuidar de búfalos numa fazenda aqui no Rio?

Fonte: Elaborado pelo autor

Outra estratégia de redução presente foi a utilização de uma sigla de modo a evitar que um sintagma fosse quebrado. Esse é o caso da legenda presente no Quadro 23, logo abaixo, no qual há a quebra indevida entre a preposição ‘sobre’ e o SN ‘a experiência de quase morte’. A redução aqui foi realizada para preservar esse texto em uma só legenda, utilizando uma sigla já utilizada anteriormente no mesmo diálogo (EQM) para se referir a esse fenômeno em questão. Essa redução foi utilizada de modo a preservar o sintagma e acomodar o texto a ser traduzido na LSE em uma única inserção de duas linhas, sem que houvesse a necessidade de se realizar uma ressegmentação retórica da legenda – leia-se dividir a mesma em duas inserções, apesar de a velocidade da legenda de partida estar abaixo do limite de 180ppm (16,5cps) e, desse modo, não haver necessidade de ser reduzida à priori. Contudo, não haveria a possibilidade de o trecho ‘sobre a experiência de quase morte em cegos de nascença’ estar todo na segunda linha da legenda, já que ele possui 55 caracteres, passando bastante do limite de 37 caracteres por linha de legenda comumente utilizado.

Quadro 23 – Exemplo de redução através do uso de sigla na proposta de relegendagem

TCR	Nº de caracteres	Velocidade da legenda	Legenda
3,8s	63c	16,5cps (média)	QUE FALA SOBRE A EXPERIÊNCIA DE QUASE MORTE EM CEGOS DE NASCENÇA.
3,8s	41c	10,7cps (baixa)	Que fala sobre a EQM em cegos de nascença.

Fonte: Elaborado pelo autor

Outra opção seria ressegmentar a fala em duas legendas para que fosse respeitada a segmentação linguística adequada do trecho e não houvesse a utilização de siglas para economia de caracteres na linha da legenda. Essa opção, contudo, necessitaria de ajustes de marcação e sincronismo do trecho legendado. A duração mínima que essa outra proposta de

relegendagem deveria ter e os seus respectivos números de caracteres estão presentes no Quadro 24.

Quadro 24 – Exemplo de redução e ressegmentação na proposta de relegendagem

Duração Mínima (180ppm)	Nº de caracteres	Legenda
2,4s	42c	Que fala sobre a Experiência de Quase Morte
1,1s	20c	em cegos de nascença.

Fonte: Elaborado pelo autor

O trecho abaixo, destacado no Quadro 25, contém muitos problemas, tanto em relação aos aspectos técnicos, quanto linguísticos. Do ponto de vista linguístico, logo à primeira vista, é possível observar, somente nesse pequeno trecho do *corpus* contendo 5 inserções de legendas, 6 problemas de segmentação linguística, um atrás do outro, inclusive uma inserção contendo 2 problemas de segmentação linguística (terceira legenda de cima para baixo). Do ponto de vista técnico, 5 das legendas contém 3 linhas de legenda, além de permanecerem muito tempo em tela (4 das 5 legendas acima do tempo máximo de 4 segundos). Essas 5 inserções perfazem um total de 24,5 segundos em tela, com uma média de 4,9 segundos por inserção de legenda. A fonte em caixa alta e alinhada à esquerda também são características que não fazem parte dos parâmetros preferidos pelos surdos brasileiros.

Quadro 25 – Trecho da LSE de partida com problemas técnicos e linguísticos

TCR	Duração	Nº de caracteres	Velocidade da legenda	Legenda
00:32:10,1 27 --> 00:32:15,6 98	5,5s	48c	8,7cps (baixa)	ELE ANDOU PELOS CORREDORES <SN_subst+SP> DO HOSPITAL. INCRÍVEL!
00:32:15,7 66 --> 00:32:20,4 70	4,7s	79c	16,8cps (média)	[GABRIEL] NÃO, FILHA, COINCIDENTEMENTE, EU ACABEI<SV_composto> DE LER UMA REPORTAGEM, ENFIM...
00:32:20,5 38 --> 00:32:25,2 07	4,6s	66c	14,3cps (baixa)	SOBRE UNS PSICÓLOGOS <SN_subst+SP> DE CAMBRIDGE QUE ESTAVAM <SV_composto> FALANDO DESSE ASSUNTO.
00:32:25,2 75 --> 00:32:31,5 14	6,2s	66c	10,6cps (baixa)	E ELES EXPLICAM QUE ESSE SUPOSTO FENÔMENO <SN_subst+SP> DA PESSOA SAIR DO CORPO...
00:32:31,5 82 --> 00:32:35,1 17	3,5s	62c	17,7cps (alta)	QUE ISSO NÃO TEM NADA <SN_especif+SN> DE PARA-NORMALIDADE. QUE É FRUTO DA MENTE.

Fonte: Elaborado pelo autor

Dessa forma, uma proposta de relegendagem do trecho em questão teria que, primeiramente, trabalhar a marcação e o sincronismo das legendas, já que a marcação das mesmas está problemática. Ao fazer isso, obtive 7 inserções de legenda, quebrando, em duas inserções, as legendas 1 e 4. Ao mesmo tempo em que consegui uma marcação adequada das legendas com as falas dos personagens, aliei isto a uma boa segmentação retórica, respeitando as hesitações e o ritmo de fala dos personagens da cena. Pela quantidade de reticências, percebe-se a quantidade de hesitações e, dessa forma, a necessidade de se separar em blocos de legenda a fala dos personagens conforme as possibilidades de tempo e espaço propiciadas pelo filme.

Após a marcação, sincronismo e consequente segmentação retórica adequada da LSE da cena em questão, além de tornar o texto da legenda centralizado e a fonte em caixa baixa, passemos aos aspectos linguísticos do texto. Nenhuma das legendas conteve uma velocidade maior do que 180ppm (a velocidade máxima considerada na proposta de relegendagem), sendo a mais veloz a última legenda do trecho, exatamente no limite de caracteres (62) permitido para o tempo de 3,5 segundos, conforme a tabela de velocidade da legenda de 180ppm de Diaz-Cintas e Remael (2007). Sendo assim, quaisquer modificações linguísticas que viessem a ocorrer seriam pautadas para se conseguir uma boa segmentação linguística da LSE, evitando os 6 problemas de quebras sintagmáticas na legenda de partida.

No Quadro 26, retirado do trecho em questão, pode-se ver uma das legendas originais do trecho e suas características técnicas e linguísticas, assim como a proposta de relegendagem.

Quadro 26 – LSE de partida com problemas técnicos e linguísticos e a proposta de relegendagem

TCR	Nº de caracteres	Velocidade da legenda	Legenda
4,7s	79c	16,8cps (média)	[GABRIEL] NÃO, FILHA, COINCIDENTEMENTE, EU ACABEI DE LER UMA REPORTAGEM, ENFIM...
3,9s	58c	14,8cps (baixa)	Filha, coincidentemente, eu acabei de ler uma reportagem...

Fonte: Elaborado pelo autor

Ao realizar o primeiro passo para adequar a legenda ao tempo máximo de 4 segundos e observar o sincronismo da fala do personagem, chegou-se aos 3,9 segundos, que, em uma velocidade de 180ppm, permitiria o uso de aproximadamente 69 caracteres. Contudo, a opção por não identificar o falante em uma cena com duas pessoas em que fosse visível

quem possui o turno de fala (conforme explicado anteriormente, configurando-se como uma escolha tradutória da proposta em questão), a menção a Gabriel foi retirada, assim como a palavra ‘não’, que, no trecho de partida, era mais uma marca de oralidade do que realmente uma negação do que havia sido dito anteriormente por Miriam, filha de Gabriel. A outra estratégia de redução utilizada teve como intuito retirar a palavra ‘enfim’ para que, dessa maneira, não se quebrasse o sintagma verbal composto ‘acabei de ler’, este deslocado para a segunda linha. A oração ‘eu acabei de ler uma reportagem, enfim...’, com 41 caracteres, passaria o limite de 37 caracteres por linha utilizado aqui. Dessa maneira, privilegiando uma segmentação adequada da legenda, optei por retirar a palavra ‘enfim’, assim deixando a segunda linha da legenda com 34 caracteres.

Presente no Quadro 27, a proposta de relegendagem não mudou tanto em relação ao trecho de partida, porém, os pequenos ajustes realizados conferiram adequação à LSE, segundo os parâmetros adotados.

Quadro 27 – LSE de partida com problemas técnicos e linguísticos e a proposta de relegendagem

TCR	Nº de caracteres	Velocidade da legenda	Legenda
4,6s	66c	14,3cps (baixa)	SOBRE UNS PSICÓLOGOS DE CAMBRIDGE QUE ESTAVAM FALANDO DESSE ASSUNTO.
3,9s	59c	15,1cps (baixa)	Sobre uns psicólogos de Cambridge que estavam falando desse assunto.

Fonte: Elaborado pelo autor

Sobre os parâmetros técnicos, assim como no exemplo anterior, foi feita uma redução do tempo em tela (de 4,6 segundos para 3,9 segundos). O texto que continha 3 linhas de legenda passou para 2 com a realização de ajustes linguísticos e espaciais. Para adequar a primeira quebra sintagmática entre o substantivo ‘psicólogos’ e o sintagma preposicional ‘de Cambridge’ apenas um “ajuste espacial” foi feito, de modo que havia espaço na primeira linha para que ‘Sobre uns psicólogos de Cambridge’ fosse possível, sem a necessidade de alguma modificação linguística no texto de modo a evitar a quebra do sintagma. O mesmo aconteceu para que fosse evitada a quebra do SV ‘estavam falando’.

O Quadro 28 traz esse trecho de 5 legendas em sequência com a proposta de relegendagem que engloba todas as questões técnicas, linguísticas e paralinguísticas discutidas aqui. As propostas de relegendagem de duas inserções do trecho em questão

comentadas logo acima exemplificam as ações tomadas de modo a alcançar uma LSE de acordo com a literatura da área.

Quadro 28 – Relegendagem de sequência de legendas problemáticas

TCR	Duração	Nº de caracteres	Velocidade da legenda	Legenda
00:05:15,942 - -> 00:05:18,544	3,6s	36c	10cps (baixa)	Ele andou pelos corredores do hospital.
00:05:18,644 - -> 00:05:19,741	1,1s	9c	8,1cps (baixa)	Incrível!
00:05:21,569 - -> 00:05:25,548	3,9s	58c	14,8cps (baixa)	Filha, coincidentemente, eu acabei de ler uma reportagem...
00:05:26,230 - -> 00:05:30,203	3,9s	59c	15,1cps (baixa)	Sobre uns psicólogos de Cambridge falando sobre esse assunto.
00:05:31,057 - -> 00:05:33,336	2,2s	18c	8,1cps (baixa)	E eles explicam...
00:05:33,436 - -> 00:05:37,080	3,6s	58c	16,1cps (média)	que esse fenômeno da pessoa, supostamente, sair do corpo...
00:05:37,180 - -> 00:05:40,635	3,4s	48c	14,1cps (baixa)	Não tem nada de paranormal, que é fruto da mente.

Fonte: Elaborado pelo autor

As 5 inserções viraram 7, respeitando, assim, as questões relativas à segmentação retórica. Todas as legendas passaram a ter tempo de 4 segundos como limite de exposição em tela. Nenhuma das legendas estava com uma velocidade acima dos 180ppm, desse modo, não houve a necessidade da utilização de recursos como redução e condensação da LSE para adequá-la ao tempo disponível em tela, assim como não havia a presença de fortes marcadores de oralidade que justificassem tais medidas de “transformação linguística”. Contudo, os parâmetros de redução e condensação foram aplicados sempre que foi necessário alcançar uma boa segmentação linguística da LSE, sendo este o parâmetro linguístico que deve ser priorizado pelos legendistas durante a realização da tradução.

No próximo capítulo, discorro sobre as considerações finais advindas da presente pesquisa.

5 CONSIDERAÇÕES FINAIS

Esta pesquisa teve como objetivos descrever e classificar os problemas de segmentação linguística da LSE do tipo *closed caption pop-on* de 5 capítulos da telenovela *Amor Eterno Amor* e propor a relegendagem da LSE de um trecho da mesma telenovela segundo os parâmetros paralinguísticos, técnicos e linguísticos apregoados pelos pesquisadores de TAV.

Para o cumprimento do primeiro objetivo da pesquisa, uma metodologia baseada em *corpus* foi utilizada a partir de anotação semiautomatizada de etiquetas próprias para a análise de problemas de segmentação linguística em Legendagem/LSE formuladas por Chaves (2012) e refinadas por Araújo e Assis (2014) e Araújo (2015) no escopo do Projeto CORSEL, projeto que visou descrever os problemas de segmentação linguística na legendagem *closed caption pop-on* da TV Brasileira a partir de telenovelas, documentários e seriados veiculados pela Rede Globo.

Através da quantificação descritiva dos dados categorizados em forma de etiquetas e evidenciados pela análise linguística feita pelo programa *Wordsmith Tools 5.0*, foi possível responder a pergunta de pesquisa aliada ao primeiro objetivo de pesquisa: ‘Quais problemas de segmentação linguística em LSE emergirão a partir da análise de 5 capítulos da telenovela *Amor Eterno Amor* e como eles classificam?’. A descrição, classificação e quantificação dos problemas de segmentação linguística em *Amor Eterno Amor* apresentaram resultados que reforçam aqueles presentes em Araújo e Assis (2014), pesquisa que analisou 1 capítulo de *Amor Eterno Amor* e foi a primeira a analisar problemas de segmentação linguística na LSE do tipo *closed caption* da TV Brasileira. O *corpus* de Araújo e Assis (2014) foi utilizado pela presente pesquisa, assim como teve seus resultados dispostos em meio a LSE dos 5 capítulos de *Amor Eterno Amor* aqui analisados, dado que se tratava de uma telenovela pertencente ao escopo do Projeto CORSEL.

A análise do *corpus* mostrou a proeminência dos problemas de segmentação linguística no nível do sintagma verbal na LSE de *Amor Eterno Amor* analisada, com 146 ocorrências (43,8%), e do sintagma nominal com 109 ocorrências (32,8%), confirmando o ranqueamento desses dois sintagmas na quantificação desses problemas na telenovela referida. Os problemas de segmentação linguística não foram apenas descritos, classificados e exemplificados, mas, ao propor ressegmentações para as legendas problemáticas identificadas no *corpus*, pude caracterizar como se daria o processo de aplicação de uma boa segmentação em um âmbito não apenas linguístico, mas que considerasse a legendagem como um todo,

inclusive os seus parâmetros técnicos, essenciais para a feitura de uma legendagem adequada segundo os parâmetros vigentes da área. Da mesma forma, ao inter-relacionar novamente, como em Araújo e Assis (2014), os parâmetros técnicos de velocidade da legenda e de número de linhas na análise dos problemas de segmentação em legenda, pude observar a recorrência vista em Araújo e Assis (2014): os problemas de segmentação linguística em *Amor Eterno Amor*, em relação à porção analisada de 5 capítulos na presente pesquisa, tendem a ocorrer em legendas de velocidade alta (180ppm), com 176 ocorrências (52,9%) e em legendas de 3 linhas, com 183 ocorrências (55%). Em relação à maior recorrência dos problemas de segmentação nos níveis dos sintagmas verbal e nominal e em legendas de velocidade alta (180ppm), a pesquisa ora relatada, além de confirmar os dados presentes em Araújo e Assis (2014), confirmou, também, os dados de Chaves (2012) em um filme em DVD, Gabriel (2015) em uma telenovela e Arraes (2015) em um filme veiculado pela TV, assim como reforçou os resultados de Gabriel (2015) no que diz respeito à recorrência desses problemas de segmentação linguística em legendas de 3 linhas.

Para cumprir o segundo objetivo da pesquisa e responder a pergunta de pesquisa ‘Como seria a relegendagem de um trecho de *Amor Eterno Amor* com problemas de segmentação, segundo os parâmetros paralinguísticos, técnicos e linguísticos apregoados pelos pesquisadores de TAV?’, relacionada a esse segundo objetivo, o programa *freeware* de legendagem *Subtitle Workshop* em sua versão 6.0B foi utilizado para a marcação, tradução e pré-visualização do trecho de pouco mais de 10 minutos de um dos capítulos de *Amor Eterno Amor* que fez parte do *corpus* da análise descritiva. Ao relegendar o trecho em questão seguindo os parâmetros técnicos, linguísticos e paralinguísticos concernentes à Legendagem para Surdos e Ensurdecidos, pude exemplificar algumas ferramentas de edição e visualização de erros presentes no software, que, a meu ver, são de grande valia para a prática profissional da legendagem.

Do mesmo modo, creio que pude explicitar o *modus operandi* do trabalho realizado pelo Grupo LEAD-UECE no âmbito da prática da LSE, resultado das pesquisas descritivas e exploratórias realizadas por pesquisadores do Grupo desde 2002 e da maneira como enxergamos a Tradução como um todo. Quase todas as decisões realizadas no momento da relegendagem descrita no capítulo de resultados foram abalizadas por essas pesquisas e são compartilhadas entre os pesquisadores do grupo no que diz respeito aos parâmetros técnicos, linguísticos e paralinguísticos e como eles interagem no momento da legendagem, com exceção, creio, do caso da não identificação dos falantes em cenas com 2 falantes em tela em que, a partir de um enquadramento próximo da câmera, seja claro quem detém o turno de fala

(cf. Subseção 4.2.1), proposição esta, contudo, que precisa ser confirmada ou refutada por pesquisas de recepção de modo a verificar sua plausibilidade e aceitação entre participantes surdos.

Considero que uma visão holística da legendagem/LSE enquanto modalidade de Tradução Audiovisual que engloba e se junta a meio semióticos distintos – a tradução de sons na LSE e as imagens em vídeo atreladas à legenda – seja necessária para que se alcance a recepção desejada a esse tipo específico de tradução. No caso da LSE em específico e da relegendagem aqui realizada, passando pela marcação, sincronismo, velocidade da legenda, segmentação, organização espacial da legenda, fonte, etc., todos os parâmetros devem estar em consonância para que haja uma boa recepção do telespectador surdo ao produto audiovisual legendado, para que ele consiga compreender o conteúdo geral, os detalhes e acompanhar a imagem na tela sem desconforto. Entretanto, segundo a visão aqui adotada, certa ênfase deve ser dada ao parâmetro da segmentação linguística nesse processo tradutório.

O processo tradutório da legendagem não deve ser visto como algo prescritivo, mas que abarca determinados parâmetros que devem ser apropriados pelo tradutor quando se está legendando, ao mesmo tempo em que há espaço para as idiossincrasias do legendista. Dessa forma, penso que a descrição do processo de relegendagem realizada por essa pesquisa a partir de parâmetros abalizados pelo Grupo LEAD-UECE possa, também, auxiliar na formação de legendistas. A análise realizada durante essa atividade de relegendagem configurou-se como uma mescla das questões teóricas e práticas retiradas da literatura e de pesquisas da área – tanto descritivas, quanto exploratórias, questões estas que devem ser utilizadas de forma consciente pelo profissional legendista durante o seu ofício.

Os resultados da descrição-classificação dos problemas de segmentação linguística na LSE de 5 capítulos de *Amor Eterno Amor* contribuem para a caracterização desse tipo de problema na LSE *closed caption* do tipo *pop-on* veiculada pela TV do Brasil e, dessa forma, contribuem para os objetivos do Projeto CORSEL. De forma mais específica, os resultados elencados pelo CORSEL apontam para uma regularidade em gêneros de ficção (filme e novela) no que diz respeito aos problemas de segmentação linguística em si e em relação com os parâmetros técnicos de velocidade da legenda e de número de linhas.

Essa regularidade sobre os problemas de segmentação linguística apontada pela presente pesquisa e por outras no âmbito do CORSEL pode auxiliar na formação de tradutores através da elaboração de cursos e treinamentos em legendagem que, utilizando os resultados do Projeto em questão, possam focar nos aspectos problemáticos da Legendagem/LSE, sobretudo o parâmetro da segmentação linguística. Uma possibilidade para o uso desses dados

é a compilação e posterior disponibilização de um *corpus* eletrônico de LSE acessível para consulta por legendistas na plataforma do Projeto CoMET³⁹ – *Corpus* Multilíngue para o Ensino e Tradução, este um *corpus* monolíngue e paralelo cujo objetivo é servir pesquisadores em Legendagem e legendistas em geral.

A partir dos resultados aqui elencados, vislumbro, de forma mais direta, duas possibilidades de desdobramentos desses resultados em pesquisas futuras. Uma dessas possibilidades seria uma pesquisa de recepção da relegendagem aqui proposta seguindo os parâmetros apregoados pela literatura em TAV acessível/LSE, de modo a verificar a sua eficácia em proporcionar uma boa recepção do produto audiovisual legendado para os surdos. Outra possibilidade, com os padrões vistos na descrição-classificação dos problemas de segmentação linguística, seria investigar experimentalmente, através de uma metodologia baseada no uso de rastreamento ocular, em que medida legendas bem ou mal segmentadas influenciam na recepção de surdos e ensurdecidos, algo ainda pouco investigado por pesquisas de recepção, sendo a relevância desse parâmetro uma recomendação prescritiva da literatura em Legendagem que necessita ser testada. Apesar de não ter feito uma pesquisa com rastreamento ocular nesse contexto supracitado ao término desta pesquisa de mestrado, como era intencionado, intenção esta explicitada no capítulo introdutório, considero os resultados aqui elencados proveitosos para que esse objetivo seja realizado no futuro.

³⁹ <http://www.fflch.usp.br/dlm/comet/>

REFERÊNCIAS

AMOR Eterno Amor. Direção: Pedro Vasconcelos. Brasil: TV Globo, 2012, 161 capítulos. sonor. color. Legenda CC disponível.

ARAÚJO, V. L. S. Closed subtitling in Brazil. In: **Topics in audiovisual translation**. Amsterdam: John Benjamins Publishing Company, 2004, v.1, 199-212.

_____. Subtitling for the deaf and hard-of-hearing in Brazil In: **Media for All: Subtitling for the Deaf, Audio Description and Sign Language**. Kenilworth:, Nova Jersey, EUA: Rodopi, 2007, v.30, 99-107.

_____. Por um modelo de legendagem para Surdos no Brasil. In VERAS, V. (org.). **Tradução e Comunicação**, Revista Brasileira de Tradutores, São Paulo: UN BERO, n. 17, p. 59–76, 2008.

_____. In Search of SDH Parameters for Brazilian Party Political Broadcasts. In: **The Sign Language Translator and Interpreter**, Manchester: St. Jerome Publishing Company, v. 3, n. 2, p. 157-167, 2009.

_____. **Legendagem para surdos**: em busca de um modelo para o Brasil. Relatório Técnico n. 306948/2008-7. Fortaleza: CNPq. Fev/2012.

_____; **A segmentação na legendagem para surdos e ensurdecidos (LSE)**: um estudo baseado em corpus. Relatório Técnico n. 306441/2011-0. Fortaleza: CNPq. Fev/2015.

_____; ASSIS, Í. A. P. A segmentação linguística na legendagem para surdos e ensurdecidos (LSE) de ‘Amor Eterno Amor’: uma análise baseada em corpus. In: NOVODVORSKI, A.; FINATTO, M. J. B. (orgs.) **Letras e Letras**, Uberlândia, v. 30, n. 2, p. 156-184, jul/dez. 2014.

ARRAES, D. A. **Legendagem para surdos e ensurdecidos**: análise baseada em *corpus* da segmentação linguística do filme “Virada Radical”. 70 f. 2015. Monografia (Bacharelado em Letras Inglês) – Universidade Estadual do Ceará, Fortaleza, 2015.

ASSIS, Í. A. P.. **A segmentação na LSE de amor eterno amor: uma análise baseada em corpus**. 2015 (Bacharelado em Letras Inglês). 71 f. Universidade Estadual do Ceará, Fortaleza- CE, 2013.

BAKER, M. Corpus Linguistics and Translation Studies: Implications and Applications. In: Baker, M.; Francis, G.; Tognini - Bonelli, E. (orgs.). **Text and technology**: In honour of John Sinclair. Philadelphia, Amsterdam: John Benjamins, 1993, p. 233- 250.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri, São Paulo: Manole, 2004, 410 p. ISBN: 85- 204-1676- 4.

CASTILHO, A. de. **Nova Gramática do Português Brasileiro**. São Paulo: Contexto, 2012. ISBN: 978-85-7244-462-0

CHAVES, É. G. **Legendagem para Surdos e Ensurdidos**: um estudo baseado em corpus da segmentação nas legendas de filmes brasileiros em DVD. 2012. 126 f. Dissertação (Mestrado em Linguística Aplicada). Universidade Estadual do Ceará, Fortaleza-CE, 2012.

CHEIAS de Charme. Direção: Carlos Araújo. Brasil: TV Globo, 2012, 143 capítulos. sonor. color. Legenda CC disponível.

DIAZ-CINTAS, J.; REMAEL, A. **Audiovisual Translation: Subtitling**. Manchester, UK, Kinderhook, N Y, UK : St. Jerome Publishing, 2007. ISBN : 978- 1900650- 95- 3/1- 900650 - 95- 9.

DINIZ, N. S. L. **A Segmentação em Legendagem para Surdos e Ensurdidos**: um Estudo Baseado em Corpus. 2012. 149 f. Dissertação (Mestrado em Linguística). Universidade Federal de Minas Gerais, Belo Horizonte-MG, 2012.

D'YDEWALLE, G. et al. Reading a message when the same message is available auditorily in another language: the case of subtitling. In: O'REGAN, J.K.; LÉVY-SCHOEN, A. (orgs.). **Eye Movements: From Physiology to Cognition**. Amsterdã e Nova York: Elsevier Science Publishers, 1987, p. 313-321.

D'YDEWALLE, G; DE BRUYCKER, W. Reading native and foreign language television subtitles in children and adults. In: HYONA, J., RADACH, R., DEUBEL, H. (orgs.), **The Mind's Eye: Cognitive and Applied Aspects of eye Movements**. Oxford, UK:Research Lirias, 2003, p. 671-684.

FRANCO, E.; ARAÚJO, V. L. S. Reading Television: Checking deaf people's Reactions to Closed Subtitling in Fortaleza, Brazil. In: GAMBIER, Y. (org.). **The Translator**, v. 9, n. 2, p.249- 267, 2003.

GABRIEL, M. H. C. Problemas de segmentação linguística na legendagem para surdos e ensurdidos (LSE) de “cheias de charme”: uma análise baseada em corpus. 2015. 108 f. Dissertação (Mestrado em Linguística Aplicada). Universidade Estadual do Ceará, Fortaleza-CE, 2015.

GOTTLIEB, Henrik. Subtitling: diagonal translation. **Perspectives: Studies in Translatology**, v. 2, issue 1, 1994.

HALLIDAY, M.A.K. **An introduction to functional grammar**. 2. ed. London: Edward Arnold, 1994.

IRMÃOS de fé. Direção: Moacyr Góes. Brasil: Columbia, 2004. 1DVD (105 min), região 4, NTSC, color., legendas (para surdos em português), janela de LIBRAS, audiodescrição e audionavegação.

IVARSSON, J.; CARROLL, M.; **Subtitling**. Simrishamm, Suécia: TransEditHB, 1998.

KALANTZI, D. **Subtitling for the Deaf and Hard of Hearing** : A corpus- based methodology for the analysis of subtitles with a focus on segmentation and deletion. 366 f.

Tese (Doutorado): School of Languages, Linguistics and Cultures of the University of Manchester, UK, 2008.

KARAMITROGLOU, F. **A Proposed Set of Subtitling Standards in Europe.** In: **Translation Journal**, v. 2, n. 2, p. 1- 15, 1998. Disponível em: <<http://translationjournal.net/journal//04stndrd.htm> > Acesso em: 15 de Março de 2011.

NASCIMENTO, A. K. P. **Linguística de corpus e legendagem para surdos e ensurdecidos (LSE): uma análise baseada em corpus da tradução de efeitos sonoros na legendagem de filmes brasileiros em DVD.** 2013. 109 f. Dissertação (Mestrado em Linguística Aplicada). Universidade Estadual do Ceará, Fortaleza-CE. 2013.

NOSSO lar. Direção: Wagner de Assis. Brasil: Fox do Brasil, 2010. 1 DVD (102min), região 4, NTSC, color., legendas (para surdos em português) e audiodescrição.

PEREGO, E. What Would We Read Best? Hypotheses and Suggestions for the Location of Line Breaks in Film Subtitles. In: **The Sign Language Translator and Interpreter.** Manchester, UK : St. Jerome Publishing, p. 35- 63, 2008. ISSN 1750- 3981.

PRAXEDES FILHO, P. H. L.; MAGALHÃES, C. M. A neutralidade em audiodescrições de pinturas. In: Org. ARAUJO, V. L. S; ADERALDO, M. F. (Orgs). **Os novos rumos da pesquisa em audiodescrição no Brasil.** 1 ed. Editora CRV. Curitiba, PR. 2013a, p. 73-87, ISBN: 978-85-8042-592-5.

PERINI, M. A. **Gramática do português brasileiro.** 366f. São Paulo: Parábola Editorial, 2010. ISBN : 978- 85- 7934- 004 - 8.

REID, H. Literature on the screen: subtitle translation for public broadcasting. In: BART, W.; D'HAEN, T. (Eds.). **Something understood.** Studies in Anglo- Dutch literary translation. Amsterdam: Rodopi, p. 97- 107, 1990.

SANTOS, D. Corporizando algumas questões. In: TAGNIN, S.E.O.; ARAÚJO, O. V. (orgs.). **Avanços da Linguística de Corpus no Brasil.** São Paulo: Humanitas, p. 41-55, 2008. ISBN: 978-85-7732-095-0.

VIRADA Radical. Direção: Jessica Bendinger. [s.l] : Touchstone Pictures e Spyglass Entertainment, 2006. 103 min. Son, Color.

WORDSMITH Tools: sítio do programa disponível em: <<http://www.lexically.net/wordsmith/index.html>>. Acesso em: 17 de março de 2013.