

UNIVERSIDADE ESTADUAL DO CEARÁ
ÉLIDA GAMA CHAVES

**LEGENDAGEM PARA SURDOS E
ENSURDECIDOS: UM ESTUDO BASEADO EM
CORPUS DA SEGMENTAÇÃO NAS LEGENDAS DE
FILMES BRASILEIROS EM DVD.**

FORTALEZA
2012

Élida Gama Chaves

LEGENDAGEM PARA SURDOS E ENSURDECIDOS: UM ESTUDO BASEADO EM *CORPUS* DA SEGMENTAÇÃO NAS LEGENDAS DE FILMES BRASILEIROS EM DVD.

Dissertação apresentada ao Programa de Pós-Graduação em Linguística Aplicada da Universidade Estadual do Ceará, como requisito parcial para obtenção do grau de mestre em Linguística Aplicada. Área de Concentração: Estudos da Linguagem.

Orientador: Prof. Dr. Antonio Luciano Pontes
Co-orientadora: Profa. Dra. Vera Lúcia Santiago Araújo

FORTALEZA - CEARÁ
2012

C5121

Chaves, Élide Gama

Legendagem para Surdos e Ensurdidos: um estudo baseado em *corpus* da segmentação nas legendas de filmes brasileiros em DVD / Élide Gama Chaves. – 2012.

126f. : il. color., enc. ; 30cm.

Dissertação (Mestrado) – Universidade Estadual do Ceará, Centro de Humanidades, Curso de Mestrado Acadêmico em Linguística Aplicada, Fortaleza, 2012.

Área de concentração: Estudos da Linguagem

Orientação: Prof. Dr. Antonio Luciano Pontes.

Co-orientação: Profa. Dra. Vera Lúcia Santiago Araújo.

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

CDD: 410

Universidade Estadual do Ceará
Programa de Pós-Graduação em Linguística Aplicada

LEGENDAGEM PARA SURDOS E ENSURDECIDOS: UM ESTUDO BASEADO
EM *CORPUS* DA SEGMENTAÇÃO NAS LEGENDAS DE FILMES
BRASILEIROS EM DVD.

Autora: Élide Gama Chaves

Defesa em: 27 / 04 / 2012

Conceito obtido: Satisfatório

Nota obtida: 10,0 (com louvor)

BANCA EXAMINADORA

Prof. Dr. Antonio Luciano Pontes (Orientador)
Universidade Estadual do Ceará - UECE

Profa. Dra. Vera Lúcia Santiago Araújo (Co-orientadora)
Universidade Estadual do Ceará - UECE

Profa. Dra. Stella Esther Ortweiler Tagnin
Universidade de São Paulo - USP

Profa. Dra. Cibele Gadelha Bernardino
Universidade Estadual do Ceará - UECE

À Vera Lúcia Santiago Araújo

DEDICO

AGRADECIMENTOS

À Cristo, que me dá luz e força para lutar pelos meus objetivos.

A minha querida família: meus amados pais, Irajá e Irene Chaves, pelo amor, apoio, dedicação e compreensão que me dão sempre, aos meus queridos irmãos, Átila e Nélida Chaves e aos meus fofinhos João Billy, Brad, Cauê, Rex, Elvis e Priscila que são a leveza do meu peso.

Ao meu namorado Douglas, que acompanhou de perto todas as etapas desta luta e que sempre se mostrou compreensivo com minha ‘ausência’, por mais perto que estivéssemos.

Aos meus queridos orientadores e amigos: Luciano Pontes e Vera Santiago por toda paciência e dedicação.

Aos meus queridos mestres, que me ajudaram, desde o começo, a crescer intelectual e profissionalmente, em especial Profa. Marisa Aderaldo.

Às Professoras Stella Tagnin, Cibele Gadelha e Dilamar Araújo, pelas contribuições no Projeto desta Dissertação.

Aos amigos que de alguma forma me ajudaram, indicando bibliografia e/ou refletindo sobre esta pesquisa, em especial Prof. Pedro Praxedes e Robson Luís.

Ao Grupo LEAD, em especial aos meus amigos e colegas de trabalho: Alexandra, Bruna, João Francisco, Juarez, Katarinna, Klístenes, Renata e Walquiria, que compartilham comigo as experiências profissionais e a amizade.

À Coordenação de Aperfeiçoamento de Pessoal de nível Superior pela bolsa de Demanda Social, no período de 2010 a 2012.

À Universidade Estadual do Ceará, que me acolheu pela segunda vez, agora em nível de mestrado.

A todos que fizeram parte desta luta.

RESUMO

Este trabalho se insere no campo dos estudos da Tradução Audiovisual, mais especificamente na área de Legendagem para Surdos e Ensurdecidos (LSE) e tem como foco de investigação a segmentação e a natureza de seus problemas. A segmentação diz respeito ao parâmetro da legendagem relacionado à distribuição de texto e divisão de legendas, que ocorre tanto entre duas legendas diferentes, quanto dentro da mesma legenda. Neste último, é chamada de quebra de linha. Para se segmentar uma fala em legendas, pode-se adotar três critérios: 1) linguístico - pautado pela sintaxe, ou seja, cada linha de legenda deve conter um pensamento o mais completo possível; 2) retórico - pelo fluxo da fala, ou seja, após uma pausa, uma nova legenda deve ser produzida; 3) visual - pautado pelo corte de cena, ou seja, sempre que a mudança de cena coincidir com o tempo de saída de uma legenda, esta não deve durar até a cena seguinte. Algumas pesquisas sugerem que se a segmentação na legendagem não for cuidadosa, os espectadores farão esforço para decodificar o texto e, conseqüentemente, poderão se cansar mais rapidamente e perder o prazer proporcionado pelo produto audiovisual. Partindo da observação que a segmentação é fator determinante no processamento de legendas e que não se encontram parâmetros para análise e produção, apenas regras, a presente pesquisa de mestrado, que consiste num estudo descritivo e baseado em *corpus* objetiva estabelecer parâmetros para análise da segmentação na LSE, além de descrever por meio de *corpus* quais são e como se caracterizam os problemas de segmentação na LSE, em língua portuguesa, de filmes. Esses objetivos foram viabilizados a partir de uma proposta de etiquetagem para análise eletrônica da segmentação na legendagem e pela análise eletrônica da LSE do filme em DVD, *Nosso Lar* (2010), com auxílio do programa de análise linguística *WordSmith Tools 5.0*. Após a análise foram identificadas inúmeras categorias de problemas de segmentação linguística (gramatical e retórica) no *corpus*: problemas nos níveis dos sintagmas nominal, preposicionado, verbal, adverbial e adjetivo; problemas nos níveis das orações coordenada e subordinada; e problemas nos níveis da retórica, porém, os resultados sugeriram que os problemas de segmentação estão concentrados em sua maioria nos níveis dos sintagmas, em especial no sintagma verbal, quando há quebra da estrutura verbo + verbo. Esses resultados foram relacionados a alguns parâmetros técnicos (nº de linhas e velocidade da legenda) e observou-se que esses problemas acontecem em legendas de duas linhas e com velocidade alta (a partir de 16 caracteres por segundo). Tendo em vista esses resultados foi possível vislumbrar que os problemas de segmentação podem ser sanados a partir do desenvolvimento de estratégias de segmentação fundamentadas nas categorias linguísticas encontradas no *corpus*, que por sua vez podem servir de parâmetros para pesquisadores e legendistas realizarem análises mais conscientes.

Palavras chave: Legendagem para surdos e ensurdecidos; Segmentação; Linguística de *Corpus*; Análise baseada em *corpus*.

ABSTRACT

This work fits into the field of Audiovisual Translation, more specifically in the area of Subtitling for the Deaf and the Hard of Hearing (SDH), focusing on segmentation and the nature of its problems. Segmentation is a subtitling parameter related to the text distribution and subdivision of subtitles that can be seen into two or more subtitles (across subtitles) and in two available lines of a subtitle (within subtitles), that are called line breaks. There are three ways of segmenting a dialogue into subtitles: 1) linguistically - on the basis of syntax, that means each subtitle line must contain a complete sentence, 2) rhetorically - on the basis of speech rhythms, that is, after a break a new subtitle must be inserted, 3) or visually - on the basis of shot cuts, that means every time a shot change coincide with a subtitle time out, this subtitle must not last until the following scene. Previous findings suggests that the bad segmentation forces the viewers to decode the text and thus they may get tired more quickly and lose the pleasure afforded by audiovisual product. Based on the observation that segmentation is an important aspect in the processing of subtitles and that there are some rules, but no parameters for its analysis and production, this present research, which is a descriptive and a corpus-based study, aims to suggest parameters for the analysis of segmentation in SDH, and describe via corpus, what are the segmentation problems in Brazilian Portuguese SDH of films. These goals were made possible from a proposed tagging meant for electronic analysis of segmentation in subtitling and from an electronic analysis of the SDH of the Brazilian DVD film, *Nosso Lar* (2010), through the software *WordSmith Tools 5.0*. The results showed several categories of linguistic segmentation problems (grammatical and rhetorical) in the corpus: problems in the levels of noun, prepositional, verb, adverbial and adjective phrases; problems at levels of coordinated and subordinated clauses, and problems in the levels of rhetorics. However, the problems of segmentation are concentrated mostly in the levels of phrases, especially in the verb phrase, where breaks in the verb + verb structure occur. These findings were related to some technical parameters (number of lines and subtitle speed) and it was observed that these problems arise in two line subtitles with high speed (16 characters per second). Given these results it was possible to suggest that the problems of segmentation can be solved by the development of segmentation strategies based on linguistic categories found in the corpus which can serve as parameters for researchers and professionals perform mindful analysis.

Keywords: Subtitling for the Deaf and the Hard of Hearing, Segmentation, Corpus Linguistics, Corpus-based analysis.

SUMÁRIO

LISTA DE ABREVIATURAS	10
LISTA DE FIGURAS	12
LISTA DE GRÁFICOS	13
LISTA DE QUADROS	13
LISTA DE TABELAS	14
1 INTRODUÇÃO	15
2 REVISÃO DE LITERATURA	22
2.1 Tradução Audiovisual e Linguística de <i>Corpus</i>	22
2.2 Legendagem para Surdos e Ensurdidos.....	32
3 METODOLOGIA	43
3.1 Tipo de pesquisa.....	43
3.2 Contexto da pesquisa.....	43
3.3 <i>Corpus</i>	44
3.3.1 Tamanho do <i>corpus</i> e representatividade	46
3.4 Instrumentos.....	47
3.5 Tratamento do <i>corpus</i>	47
3.5.1 Extração das legendas.....	47
3.6 Anotação do <i>corpus</i>	49
3.7 Análise dos Dados.....	59
4 ANÁLISE E DISCUSSÃO DOS DADOS	69
4.1 Dados gerados pela <i>Word List</i>	69
4.2 Parâmetros Técnicos da LSE do filme <i>Nosso Lar</i>	70
4.2.1 Número de linhas por legenda	70
4.2.2 Número de caracteres por linha	71
4.2.3 Velocidade da legenda.....	73
4.3 Parâmetros Linguísticos da LSE do filme <i>Nosso Lar</i>	75
4.3.1 Problema de Segmentação Retórica (PROSEGR).....	77
4.3.2 Problema de Segmentação Gramatical (Sintagma e Oração)	79
4.3.2.1 Sintagma Nominal (SN).....	80
4.3.2.2 Sintagma Preposicionado (SP).....	84
4.3.2.3 Sintagma Verbal (SV).....	85
4.3.2.4 Sintagma Adverbial (SAdv).....	89
4.3.2.5 Sintagma Adjetivo (SAdj).....	91
4.3.2.6 Oração Coordenada (COORD)	94
4.3.2.7 Oração Subordinada (SUBORD)	95
4.4 Discussão dos Dados.....	97

5 CONSIDERAÇÕES FINAIS	103
REFERÊNCIAS	107
APÊNDICES	114
APÊNDICE A – Conjunto de etiquetas	115
APÊNDICE B – Legendas etiquetadas com problemas de Segmentação	116

LISTA DE ABREVIATURAS

ABREVIATURA/SIGLA	SIGNIFICADO
LSE	– Legendagem para Surdos e Ensurdidos
BNDES	– Banco Nacional do Desenvolvimento
MPF	– Ministério Público Federal
SDH	– Subtitling for the deaf and the Hard of Hearing
TAV	– Tradução Audiovisual
CoMET	– <i>Corpus</i> Multilíngue para Ensino e Tradução
BNC	– <i>British National Corpus</i>
COCA	– <i>Corpus of Contemporary American English</i>
AD	– Audiodescrição
LATAV	– Laboratório de Tradução Audiovisual
LEAD	– Legendagem e Audiodescrição
UECE	– Universidade Estadual do Ceará
DVD	– <i>Digital Versatile Disc</i> , em português, Disco Digital Versátil
PosLA	– Programa de Pós-Graduação em Linguística Aplicada
MOLES	– Modelo de Legendagem para Surdos
CNPq	– Conselho Nacional de Desenvolvimento Científico e Tecnológico
UNB	– Universidade de Brasília
CPS	– Caracteres por segundo
PPM	– Palavras por minuto
CAPES	– Coordenação de Aperfeiçoamento de Pessoal de nível Superior
LBRAS	– Língua Brasileira de Sinais
OCR	– Reconhecimento de caractere (<i>Optical Character Recognition</i>)
BNB	– Banco do Nordeste do Brasil
HTML	– Linguagem de Marcação de Hipertexto (<i>Hyper Text Markup Language</i>)
sub	– Legenda (<i>subtitle</i>)
L	– Linha (de legenda)
<t>	– Tempo inicial
</t>	– Tempo final
PROSEGG	– Problema de Segmentação Gramatical

PROSEGR	– Problema de Segmentação Retórica
PROSEGV	– Problema de Segmentação Visual
SN	– Sintagma Nominal
SP	– Sintagma Preposicionado
SV	– Sintagma Verbal
SAdv	– Sintagma Adverbial
SAdj	– Sintagma Adjetivo
COORD	– Oração Coordenada
SUBORD	– Oração Subordinada

LISTA DE FIGURAS, GRÁFICOS, QUADROS E TABELAS

LISTA DE FIGURAS

Figura 1: Identificação de efeito sonoro entre colchetes no modelo de legendagem do Brasil.	23
Figura 2: Identificação do falante por cores no modelo de legendagem da Espanha.	24
Figura 3: Árvore Sintática proposta por Karamitroglou (1998)	40
Figura 4: Arquivo de legenda em formato “.srt” aberto com Bloco de Notas	48
Figura 5: Interface do programa SubRip	48
Figura 6: Interface do <i>WordSmith Tools 5.0 – Settings</i>	60
Figura 7: Interface do <i>WordSmith Tools 5.0 – Tags</i>	61
Figura 8: Interface do <i>WordSmith Tools 5.0 – Word List</i>	61
Figura 9: Interface do <i>WordSmith Tools 5.0 – Word List-Choose Texts Now</i>	62
Figura 10: Interface do <i>WordSmith Tools 5.0 – Choose Texts</i>	62
Figura 11: Interface do <i>WordSmith Tools 5.0 – Make a Word List Now</i>	63
Figura 12: Interface do <i>WordSmith Tools 5.0 – Listas de Frequência, Alfabética e Estatística</i>	63
Figura 13: Interface do <i>WordSmith Tools 5.0 – Salvar WordList</i>	64
Figura 14: Interface do <i>WordSmith Tools 5.0 – Menu Tags-Mark up to ignore</i>	65
Figura 15: Interface do <i>WordSmith Tools 5.0 – Concord</i>	66
Figura 16: Interface do <i>WordSmith Tools 5.0 – Concord-Choose Texts Now</i>	66
Figura 17: Interface do <i>WordSmith Tools 5.0 – Concord-Search Word</i>	67
Figura 18: Interface do <i>WordSmith Tools 5.0 – Palavra de busca em concordância</i>	67
Figura 19: Interface do <i>WordSmith Tools 5.0 – Salvar Lista de Concordância</i>	68
Figura 20: Resultado estatístico gerado pela <i>Word List</i>	69
Figura 21: Tela da <i>Concord</i> mostrando resultado de 574 legendas de 1 linha.....	70
Figura 22: Tela da <i>Concord</i> mostrando resultado de 558 legendas de 2 linhas	71
Figura 23: Tela da <i>Concord</i> exibindo o número de caracteres e o tempo de cada legenda	73
Figura 24: Tela da <i>Concord</i> mostrando as 3 ocorrências de PROSEGR	78
Figura 25: Problema de segmentação no SN pela quebra de pre-nucleares+substantivo	80
Figura 26: Tela da <i>Concord</i> mostrando as 23 ocorrências de problema de segmentação no nível do SN.	81

Figura 27: Tela da <i>Concord</i> mostrando as 2 ocorrências de problema de segmentação no nível do SP.	84
Figura 28: Tela da <i>Concord</i> mostrando as 36 ocorrências de problema de segmentação no nível do SV.	86
Figura 29: Tela da <i>Concord</i> mostrando as 6 ocorrências de problema de segmentação no nível do SAdv.	90
Figura 30: Tela da <i>Concord</i> mostrando as 9 ocorrências de problema de segmentação no nível do SAdj.	92
Figura 31: Problema de segmentação na Oração Subordinada pela quebra de conjunção+oração	93
Figura 32: Tela da <i>Concord</i> mostrando as 2 ocorrências de problema de segmentação no nível da oração coordenada.	94
Figura 33: Tela da <i>Concord</i> mostrando as 7 ocorrências de problema de segmentação no nível da oração subordinada.	96

LISTA DE GRÁFICOS

Gráfico 1: Velocidades das legendas com problemas de segmentação	75
Gráfico 2: Resultado geral da Segmentação na legendagem do filme <i>Nosso Lar</i>	97

LISTA DE QUADROS

Quadro 1: Quadro sinótico das etiquetas propostas	58
Quadro 2: Exemplo de trecho sem etiquetas e com etiquetas do filme <i>Nosso Lar</i>	59
Quadro 3: Legenda 713 etiquetada – Exemplo de identificação de falante e de efeito sonoro	72
Quadro 4: Legenda 914 etiquetada – Exemplo de identificação de falante e de efeito sonoro	72
Quadro 5: Problemas de segmentação gramatical – SN_pre-nucleares+subst	81
Quadro 6: Problemas de segmentação gramatical – SN_nominal+modif/modif+nominal	82
Quadro 7: Problemas de segmentação gramatical – SN_superlativo+adj	82
Quadro 8: Problemas de segmentação gramatical – SN_relativo+oração incompleta	83
Quadro 9: Problemas de segmentação gramatical – SN_título+nome próprio	83

Quadro 10: Problemas de segmentação gramatical – SN_colocações/idiom/conv.....	83
Quadro 11: Problemas de segmentação gramatical – SP_prep+subst	85
Quadro 12: Problemas de segmentação gramatical – SV_verbo+verbo	87
Quadro 13: Problemas de segmentação gramatical – SV_verbo+adv	87
Quadro 14: Problemas de segmentação gramatical – SV_negação+verbo	88
Quadro 15: Problemas de segmentação gramatical – SV_(verbo)+oblíquo+verbo.....	88
Quadro 16: Problemas de segmentação gramatical – Sadv.....	90
Quadro 17: Problemas de segmentação gramatical – SAdj_subst+adj	92
Quadro 18: Problemas de segmentação gramatical – COORD_coordenador+oração.....	95
Quadro 19: Problemas de segmentação gramatical – SUBORD_conj+oração.....	96

LISTA DE TABELAS

Tabela 1: Velocidade de leitura da legenda a 145ppm detalhada em cps	36
Tabela 2: Velocidade de leitura da legenda a 160ppm detalhada em cps	36
Tabela 3: Velocidade de leitura da legenda a 180ppm detalhada em cps	37
Tabela 4: Velocidade das legendas do filme <i>Nosso Lar</i>	74

1 INTRODUÇÃO

A legendagem para surdos e ensurdecidos¹ (LSE) é um recurso de acessibilidade e uma modalidade de tradução que difere da legendagem para ouvintes² no que diz respeito às informações adicionais de identificação de falante e de efeito sonoro contidas na LSE, traduzidas para que o surdo possa ter acesso à trilha sonora do filme e para que não confunda quem está com o turno. Esse recurso começou a ser transmitido na televisão aberta em 1997, quando foi ao ar pela primeira vez no *Jornal Nacional* transmitido pela *Rede Globo de Televisão* (SELVATICI, 2010, p. 27), porém, no que diz respeito aos filmes nacionais essa realidade é outra. Desde 2004, surdos brasileiros brigam na justiça pela inclusão de legendas em língua portuguesa em filmes nacionais, mas só agora, fevereiro de 2012, a Procuradoria da República em São Paulo entrou na justiça com uma ação civil pública para regulamentar a obrigatoriedade das legendas em filmes nacionais realizados com patrocínio público. Os réus, Petrobrás e BNDES - Banco Nacional do Desenvolvimento, principais patrocinadores e financiadores das obras audiovisuais nacionais, devem adequar seus editais e contratos num prazo de 40 dias, a fim de que todas as cópias de produções audiovisuais por eles financiadas e/ou patrocinadas, destinadas ao mercado nacional, contemplem legendas abertas descritivas em língua portuguesa com o fim de proporcionar acessibilidade das pessoas com deficiência auditiva a seu conteúdo. O descumprimento pode acarretar em multa diária no valor mínimo de 100 mil Reais. Segundo o Ministério Público Federal - MPF, a ação pode beneficiar mais de cinco milhões de brasileiros que sofrem algum tipo de deficiência auditiva.³

As legendas abertas descritivas em língua portuguesa são consideradas abertas porque não precisam ser acionadas pelo telespectador para serem visualizadas, ao contrário das legendas fechadas ou ocultas que precisam ser acionadas através de algum dispositivo para aparecerem. Também são consideradas descritivas porque traduzem, além das falas, os sons dos programas, como latido de cachorro, campainha, choro de bebê, mesmo quando não

¹ Segundo Neves (2005, p. 83) Ensurdido é a pessoa que tem de leve a moderada perda auditiva (algo em torno de 15 a 60db).

² Para este trabalho, ouvinte é aquele com audição normal, que não é surdo ou ensurdido.

³ Fonte: sítio da Procuradoria da República de São Paulo. Disponível em: <http://www.prsp.mpf.gov.br/sala-de-imprensa/noticias_prsp/14-02-12-2013-prdc-move-acao-para-que-filmes-nacionais-tenham-legendas-para-surdos>. Ação completa disponível em: <http://www.prsp.mpf.gov.br/sala-de-imprensa/pdfs-das-noticias/PRDC_ACP_Legenda_0002444-97.2012.4.03.6100.PDF/at_download/file>. Acesso em: 08 de março de 2012.

aparecem em cena⁴. O termo ‘legenda descritiva’ utilizado na Ação do MPF não é contemplado na literatura de Tradução Audiovisual ou Legendagem, o termo estabelecido é o mesmo utilizado neste trabalho ‘Legendagem para Surdos e Ensurdidos’ ou LSE, que provém do inglês *Subtitling for the deaf and the Hard of Hearing* (SDH). Independente da concepção do MPF com relação à legendagem para surdos, essa Ação se configura como um primeiro passo na inclusão de surdos e ensurdidos à indústria de filmes nacionais, mas há muito a fazer em prol dessa inclusão, tendo em vista as reais necessidades dos surdos brasileiros relatadas em pesquisas exploratórias, as quais definiram um modelo de LSE que melhor atende as necessidades da comunidade surda brasileira: as legendas devem ser condensadas e de cor amarela ou branca; deve haver identificação de falante e de efeito sonoro por meio de colchetes, sempre que relevante para a narrativa; a voz em *off* (aquela vinda de alguém que não aparece na tela, do rádio, da televisão, do telefone etc.) deve estar em itálico; a identificação de música deve ser representada apenas por uma clave (nota musical), não sendo necessário discriminar os ritmos nem traduzir as letras, a não ser que se trate de uma música tema, ou que seja fundamental à narrativa, nesse caso a letra deve estar em itálico (ARAÚJO, 2008; ARAÚJO; NASCIMENTO, 2011).

Com a regulamentação dessa Ação, haverá um aumento crescente na demanda por profissionais legendistas na indústria do cinema nacional. Essa ação do MPF, dentre outros fatores, faz com que essa pesquisa se torne cada vez mais necessária à sociedade em geral, à comunidade surda, à academia e aos profissionais legendistas, que terão suporte teórico metodológico para oferecerem legendas de qualidade aos mais de nove milhões de brasileiros que declaram ter algum tipo de deficiência auditiva, segundo dados do mais recente CENSO demográfico (IBGE, 2010)

Sobre a motivação para investigar a LSE neste trabalho, posso dizer que ela surgiu da inquietação, enquanto espectadora, legendista e pesquisadora da área, causada pelos frequentes problemas de segmentação encontrados em vídeos legendados, que ocasionavam a quebra de raciocínio e demandavam maior esforço para recobrar as ideias perdidas. Além do desconforto, essa quebra de raciocínio causava prejuízos à compreensão do filme e ao prazer por ele proporcionado. O problema se agravou quando me deparei com a falta de parâmetros

⁴ Definição de legenda descritiva obtida no Portal Acessibilidade total, disponível em: <<http://www.acessibilidadetotal.com.br/legenda-ou-janela-de-interpretres/>>. Acesso em: 13 de março de 2012.

tanto para o exercício quanto para a análise da segmentação nas legendas para surdos. Então, partindo desse contexto, me propus a investigar a natureza desses problemas de segmentação, tendo em vista que nenhum trabalho no campo da Tradução Audiovisual (TAV), com exceção de Perego (2008), detalhado mais adiante, havia investigado a fundo esse parâmetro. O que havia disponível na literatura, que fosse do meu conhecimento, eram regras de segmentação gramatical que prescreviam o certo e o errado, sem justificar o que de fato, determinava aquele fenômeno, ou seja, a natureza dos problemas. Essa situação me motivou a investigar a natureza de tais problemas com o intuito de entender como eles se apresentam nas legendas e quais são as características de cada um. Para isso, tive que buscar fundamentos teóricos que amparassem minhas inquietações, além de uma metodologia que me desse todo o suporte para realizar a pesquisa e alcançar os objetivos pretendidos.

Os primeiros fundamentos que me nortearam pelo caminho da investigação foram: a afirmação de que uma segmentação criteriosa aumenta a capacidade de processamento da legendagem (PEREGO, 2008, p. 35), e o relato de pesquisas exploratórias com surdos brasileiros que vêm sugerindo que se a legenda for bem segmentada os surdos têm uma boa recepção mesmo que para isso haja o descumprimento de outros parâmetros (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004; ARAÚJO, 2005; ARAÚJO, 2007; ARAÚJO, 2008; ARAÚJO; NASCIMENTO, 2011). Corroborando as afirmações de que uma boa segmentação influencia positivamente na recepção e no processamento das legendas, é que dou início à investigação dos problemas de segmentação na LSE com o intuito de que os resultados dessa pesquisa, ao serem colocados em prática, possam influenciar no modo de pensar a tradução para legendas, em especial a questão da segmentação, foco dessa investigação, bem como, contribuir para o desenvolvimento das estratégias de legendagem.

No tocante à metodologia, a escolha para esta pesquisa teve como ponto de partida uma pesquisa descritiva, da presente autora, que contemplou o uso de *corpora* como metodologia de análise e descrição de legendas (CHAVES, 2009)⁵. Essa pesquisa, detalhada mais adiante, teve como objetivo verificar a viabilidade do uso de *corpora* na análise e descrição de LSE a partir de uma análise baseada em *corpus* de 6 variados trechos da

⁵ Monografia de conclusão de curso aprovada em 30 de setembro de 2009, na Universidade Estadual do Ceará para a aquisição do grau de Bacharel em Letras em inglês, com o tema: Legenda para Surdos no Brasil: uma análise baseada em *corpus*.

programação da *Rede Globo* com auxílio do programa de análise linguística *WordSmith Tools 5.0*. Os parâmetros analisados na legendagem desses programas foram: densidade lexical, nível de condensação, segmentação e explicitação por adição e por especificação. Os resultados sugeriram que a análise baseada em *corpus* não só é viável, como também imprescindível para agilizar a análise de legendas em grande quantidade. Além disso, a análise baseada em *corpus* pôde revelar características do *corpus* e gerar parâmetros de análise, que não seriam viáveis por meio de análise manual (CHAVES, 2009).

O uso de *corpora* nos estudos da tradução foi introduzido por Baker (1993), quando, a partir de então, a metodologia baseada em *corpus* passou a ser utilizada para investigar supostos universais de tradução (BAKER, 1996), formar tradutores (ZANETTIN, et al. 2003) e disponibilizar *corpora* para serem consultados por tradutores como é o caso do projeto CoMET – *Corpus Multilíngue para Ensino e Tradução*, disponível em: <<http://www.fflch.usp.br/dlm/comet/>> (TAGNIN, 2004). Além do CoMET existem outros *corpora* que permitem a investigação de questões da tradução em várias línguas, como o COMPARA (Português/Inglês), *Lácio-Web* (Português), BNC – *British National Corpus* (Inglês), *Cobuild* (Inglês), COCA – *Corpus of Contemporary American English* (Inglês), *Corpus del Español* (Espanha), *WebCorp* (Multilíngue), dentre outros. Embora os *corpora* sejam comuns para outras modalidades de tradução, só recentemente foram introduzidos nos estudos da TAV, mais especificamente nos estudos em LSE e Audiodescrição (AD), mas por questões metodológicas nem sempre os *corpora on line* são a melhor escolha. Segundo Berber Sardinha (2004, p. 143) embora haja vários *corpora* em existência aos quais o interessado pode ter acesso, seja localmente, ou por acesso remoto via *www*, em muitos casos torna-se necessário coletar e organizar o seu próprio *corpus*, como é o caso da pesquisa proposta e de outras pesquisas em TAV, especificamente em LSE e AD.

No que diz respeito à audiodescrição é possível citar os trabalhos de Hurtado (2007); Hurtado, et al. (2010) e Salway (2007). Hurtado (2007; 2010) analisou um *corpus* de mais de 300 filmes audiodescritos com objetivo de descrever e delinear os parâmetros utilizados em roteiros de AD em espanhol, inglês, francês e alemão. Salway (2007) analisou um *corpus* de 91 roteiros de filmes audiodescritos com o objetivo de categorizar as singularidades linguísticas dos roteiros de AD.

Quanto à legendagem, é possível citar as pesquisas de Perego (2003; 2009), Perego (2008), Kalantzi (2008), Chaves (2009) e Feitosa (2009). Kalantzi (2008) compilou um *corpus* de legendas para surdos de três programas da BBC e teve como foco a análise e descrição de dois parâmetros linguísticos da legendagem: segmentação e redução, e teve como objetivos descrever, além de parâmetros linguísticos, parâmetros técnicos da LSE, e por fim, também objetivou propor regras de segmentação gramatical para LSE. Perego (2003; 2009), por sua vez, realizou análise baseada em *corpus* para descrever um parâmetro da legendagem que também se configura como um ‘universal da tradução’ proposto por Baker (1996): a explicitação por adição e por especificação. No que diz respeito à segmentação, Perego (2008) analisou a segmentação entre linhas de legendas (*line breaks*) de um *corpus* heterogêneo de legendas para filmes em DVD e cinema com o intuito de investigar e descrever como funciona essa segmentação, do ponto de vista linguístico. Tanto Kalantzi quanto Perego não fizeram uso de *corpora* eletrônicos, desta forma, as análises foram realizadas manualmente.

Ao contrário de Perego (2003; 2009) e Kalantzi (2008), tanto Chaves (2009) quanto Feitosa (2009) realizaram análise eletrônica através do programa de análise linguística *Wordsmith Tools 5.0*. Chaves (2009) teve como objetivo verificar a viabilidade do uso de *corpora* na análise e descrição de parâmetros da LSE em trechos de gêneros variados da *TV Globo*. Já Feitosa (2009) utilizou a metodologia baseada em *corpus* para comparar *fansubs* (legendas de fãs ou piratas) com as legendas comerciais de trechos de 10 filmes da década de 2000 com o objetivo de correlacionar os resultados das análises buscando uma caracterização da legendagem pirata.

Esses estudos, detalhados no capítulo teórico, mostraram a potencialidade da metodologia baseada em *corpus* na análise e descrição de LSE e AD. A pesquisa de Chaves (2009) e a pesquisa proposta têm seu diferencial em relação às pesquisas supracitadas: somente as duas fizeram análise eletrônica de LSE utilizando o *WordSmith Tools 5.0* para analisar e descrever parâmetros da LSE.

É importante contextualizar aqui as raízes desta pesquisa. Trata-se de uma pesquisa em nível de mestrado, inserida na linha *Tradução, Lexicologia e Processos Cognitivos*, vinculada ao PosLA - Programa de Pós-Graduação em Linguística Aplicada da UECE. Também está

inserida no projeto MOLES - Modelo de Legendagem para Surdos, que visa definir um modelo de LSE que atenda às necessidades da comunidade surda brasileira e garanta acesso desse público aos meios audiovisuais. O MOLES é desenvolvido no âmbito do LATAV, nas dependências do Centro de Humanidades da UECE, onde acontecem pesquisas em LSE e AD vinculadas ao grupo de pesquisa registrado no CNPq *Tradução e Semiótica* liderado pelas Profas. Dras. Vera Lúcia Santiago Araújo (UECE) e Soraya Ferreira Alves (UNB). Esta pesquisa conta com a orientação do Prof. Dr. Antonio Luciano Pontes e com a co-orientação da Profa. Dra. Vera Lúcia Santiago Araújo, pesquisadores atuantes na área de Lexicologia e Terminologia, e Tradução Audiovisual e Acessibilidade, respectivamente.

Esta pesquisa tem relevância social porque gera expectativas quanto à inclusão de surdos e ensurdecidos aos meios audiovisuais uma vez que os resultados obtidos oferecem estratégias de tradução para legendas que podem contribuir para melhorar a qualidade dos produtos audiovisuais. E quanto melhor (mais acessível) for o produto audiovisual maior será a inclusão e adesão desse público a esse produto. Também tem relevância acadêmica porque amplia as discussões nos estudos baseados em *corpora*, com relação aos estudos da TAV e serve como base para tradutores em formação, tradutores profissionais e pesquisadores. Tem relevância tecnológica porque lida com as novas tecnologias oferecidas pela Linguística de *Corpus* que por sua vez, possibilitam o rápido processamento de grandes quantidades de dados e isso atribui mais validade e representatividade à pesquisa. Também tem relevância institucional, pois contribui para a continuidade do ciclo de estudos em acessibilidade audiovisual que vem sendo desenvolvido pelo LEAD, além de contribuir com a visibilidade do PosLA.

A partir desse suporte teórico metodológico e de toda relevância desta pesquisa para a sociedade e para a academia foi possível dar início à investigação dos parâmetros de legendagem, com foco na segmentação, por meio da análise eletrônica da LSE do filme brasileiro *Nosso Lar* (2010). Essa análise eletrônica foi viabilizada por duas das três ferramentas do *WordSmith Tools 5.0*⁶: o gerador de lista de palavras *Word List* e o concordanciador *Concord* que estão detalhados no capítulo de metodologia.

⁶ Software desenvolvido por Mike Scott através da *Lexical Analysis Software Ltd.* e *Oxford University Press*. Disponível em: <<http://www.lexically.net/wordsmith/index.html>>. Acesso em: 10 de março de 2012.

Esta dissertação está dividida em cinco capítulos. Além dessa introdução, temos: 2- Revisão de Literatura, 3- Metodologia, 4- Análise e Discussão dos dados, e 5 - Considerações Finais. O capítulo 2 aborda as principais fontes teóricas da Tradução Audiovisual, da Legendagem e da Linguística de *Corpus* através de uma revisão de literatura dos aspectos mais importante para o desenvolvimento deste trabalho. O capítulo 3 expõe o desenvolvimento dessa pesquisa, a fim de alcançar os objetivos pretendidos, ressaltando cada passo e cada ferramenta, bem como a utilização das mesmas e os procedimentos necessários para a análise dos dados. O capítulo 4 consiste numa análise detalhada do *corpus* e na interpretação dessa análise em relação aos dados da pesquisa, além de mostrar as implicações desse estudo para a academia, para os tradutores e para a sociedade. E por fim, o capítulo 5, que recapitula os objetivos, bem como os principais pontos dessa pesquisa, enfatiza os resultados obtidos e traça perspectivas para novas pesquisas na área.

São objetivos gerais desta pesquisa:

- Estabelecer parâmetros para análise da segmentação na LSE utilizando a Linguística de *Corpus*;
- Descrever por meio de *corpus* quais são e como se caracterizam os problemas de segmentação na LSE, em língua portuguesa, de filmes.

São objetivos específicos desta pesquisa:

- Propor uma etiquetagem para análise eletrônica da segmentação na legendagem;
- Realizar uma análise eletrônica da LSE, do filme em DVD *Nosso Lar* (2010).

2 REVISÃO DE LITERATURA

Este capítulo mostra os fundamentos teóricos que embasam esta pesquisa: de um lado a Tradução Audiovisual e seu relacionamento com as inovações tecnológicas e metodológicas da Linguística de *Corpus*, e de outro a Legendagem para Surdos e Ensurdecidos e seus parâmetros, ambas adotadas de forma complementar. No que concerne à TAV, foram apresentados os principais trabalhos da área que se aliaram à Linguística de *Corpus*, que de uma forma ou de outra dialogam com esta pesquisa, seja através do tema, da metodologia ou de outro ponto. E no que concerne à LSE foram discutidos os parâmetros e as técnicas utilizadas nessa pesquisa, na visão de grandes estudiosos da área.

2.1 Tradução Audiovisual e Linguística de *Corpus*

A tradução audiovisual é um campo de estudo relativamente recente que vem crescendo conforme as demandas da sociedade por produtos audiovisuais. Sobre a TAV, Diaz Cintas (2009, p. 1) comenta que nos últimos 20 anos a indústria audiovisual gerou um campo fértil de atividade para as pesquisas acadêmicas envolvendo a tradução. Comenta também que além de crescer enquanto atividade profissional, graças, principalmente, à revolução digital, a TAV tornou-se uma área estabelecida e de destaque da pesquisa acadêmica.

As principais modalidades de tradução audiovisual são: legendagem, dublagem, *voice over* e mais recentemente, audiodescrição e a legendagem para surdos e ensurdecidos. A legendagem consiste em texto escrito, exibido em produtos audiovisuais, traduzido a partir das informações verbais (diálogos, músicas etc.) e não verbais (fachadas de prédios, letreiros, bilhetes, toque de telefone etc.). Dublagem, adaptado do conceito encontrado em Diaz Cintas (2009, p 4) consiste na substituição da trilha sonora fonte pela trilha sonora alvo contendo todos os diálogos gravados de modo que os sons e o movimento dos lábios dos atores estejam sincronizados, dando a aparência de que os diálogos acontecem realmente na língua alvo. E *voice over* consiste num texto oral, traduzido a partir de um texto oral na língua fonte que deve ser sobreposto à trilha de áudio somente segundos depois de começar e durar até segundos antes de terminar a fala. O áudio fonte deve estar com volume mais baixo que o áudio traduzido apenas no momento da sobreposição para garantir que o espectador possa acompanhar ao menos parte do áudio fonte com volume normal.

No que tange à acessibilidade audiovisual, a AD e a LSE são modalidades que crescem a passos largos, além de gerarem expectativas e aos poucos atenderem às demandas sociais por produtos acessíveis especialmente, mas não exclusivamente, a pessoas com deficiência visual e a surdos e ensurdecidos. Sobre as expectativas da LSE, Neves (2007) acredita que quando há o envolvimento das pesquisas com as emissoras de TV, produtoras, legendistas e espectadores, pode haver melhoria significativa dos padrões de qualidade e nas ferramentas de trabalho dos profissionais, que podem ser usadas na formação do legendista, uma vez que irão se beneficiar de programas de ensino que desenvolvem as habilidades necessárias para a produção de boas legendas.

Voltando às definições, a AD é um recurso de acessibilidade audiovisual e uma modalidade da tradução audiovisual que consiste na tradução de elementos não verbais em texto oral inserido de modo sobreposto à trilha sonora do produto audiovisual, preferencialmente nos intervalos em que não há diálogos. A LSE, detalhada no próximo capítulo, é objeto de investigação desta pesquisa e consiste num tipo de tradução que se assemelha em grande parte à legendagem para ouvintes, com a diferença que a LSE traduz em texto escrito os efeitos sonoros e identifica os falantes com o intuito de que surdos e ensurdecidos possam construir a narrativa do produto audiovisual através dos elementos sonoros e da distinção dos falantes. Em alguns países essa identificação é feita por meio de colchetes [], como no Brasil, e em outros é feita através das cores e da posição da legenda na tela.



Figura 1: Identificação de efeito sonoro entre colchetes no modelo de legendagem do Brasil. Quadro do filme brasileiro *Nosso Lar* (2010).



Figura 2: Identificação do falante por cores no modelo de legendagem da Espanha. Quadro do filme espanhol *Torrente 3: El Protector* (2005).

Para desenvolver essa pesquisa no campo da tradução audiovisual foi necessário buscar um suporte teórico-metodológico que servisse aos objetivos pretendidos. Esse suporte foi dado pela Linguística de *Corpus*, especialmente pela metodologia baseada em *corpus* (*Corpus Based Research*), que segundo o glossário de Linguística de *Corpus* é utilizada para comprovar (ou não) uma hipótese ou para extrair exemplos, diferentemente da pesquisa direcionada pelo *corpus* (*Data Driven Research*) que consiste num estudo desenvolvido conforme dados apresentados pelo *corpus*, sem pressuposições teóricas (TAGNIN, 2011).

A Linguística de *Corpus* tem sido suporte metodológico muito usado pelos Estudos da Tradução para gerar dados e resultados mais confiáveis. Pode-se considerar que a precursora dos estudos de *corpora* vinculados aos estudos da tradução foi Mona Baker, que publicou em 1993 um artigo seminal, importantíssimo para o desenvolvimento dos estudos descritivos da tradução, no qual previu que tanto a técnica quanto a metodologia, desenvolvidas no campo da Linguística de *Corpus*, causariam impacto direto sobre os estudos da tradução que já sofriam transformações nos campos teórico e descritivo. Essas transformações ajudaram a reformular a visão tradicionalista predominante nos estudos da tradução, que entendia a tradução como uma atividade de segunda classe, desmerecedora de investigação científica, e que tinha como foco a investigação da correspondência, da equivalência e da questão de escolha, mas esse tipo de investigação não gerava resultados conclusivos e acabava suprimindo questões práticas, como o processo tradutório e a formação de tradutores (BAKER, 1993).

O fato pelo qual muitos compiladores de *corpora*, especialmente na Europa, excluía textos traduzidos de seus *corpora* fundamentou a ideia de que textos traduzidos não eram representativos e poderiam distorcer a visão real de língua sob investigação. Entretanto, Baker (1993) defende que textos traduzidos relatam eventos comunicativos genuínos que nem são inferiores nem superiores a outros eventos comunicativos em qualquer língua, são apenas diferentes, e conscientiza que a natureza dessa diferença precisa ser explorada e descrita.

Após tamanha revolução no escopo da disciplina de estudos da tradução, Baker inaugurou seus pressupostos teóricos propondo uma investigação baseada em *corpus* dos aspectos recorrentes em textos traduzidos, e os resultados mostraram quatro aspectos que ocorreram em todos os textos do *corpus*, conhecidos, *a posteriori* como *universais da tradução*, que Segundo Baker (1996) incluem: simplificação (a ideia inconsciente que o tradutor tem de simplificar tanto a linguagem quanto a mensagem), explicitação (tendência de explicar ou adicionar informações detalhadas ao texto traduzido), normalização (tendência de adaptar (domesticar) o texto fonte a padrões e práticas típicas da língua alvo, mesmo a ponto de parecer exagerado) e estabilização (consiste na tendência que os textos traduzidos têm de manterem sempre o mesmo padrão textual, padrão este não encontrado em textos não traduzidos).

Essa discussão impulsiona a atual pesquisa, na qual a LSE é entendida como texto autêntico, cheio de características que precisam ser exploradas e descritas. Assim como a presente pesquisa, outras no âmbito da tradução audiovisual tiveram suporte metodológico da Linguística de *Corpus* para explorarem suas características particulares, como é o caso das pesquisas em AD (HURTADO, 2007; HURTADO, et al. 2010; SALWAY, 2007;) e das pesquisas em legendagem (PEREGO, 2003; 2008; 2009; 2010; KALANTZI, 2008; FEITOSA, 2009; CHAVES, 2009).

No que tange a AD, Hurtado (2007; 2010) analisou um *corpus* de mais de 300 filmes audiodescritos com objetivo de descrever e delinear os parâmetros utilizados em roteiros audiodescritos em espanhol, inglês, francês e alemão. Foram analisados parâmetros narratológicos, cinematográficos e linguísticos. A análise foi realizada por meio de um *corpus* anotado, etiquetado com o auxílio de um programa etiquetador (*Tagetti*) e processado pelo *Wordsmiths Tools*. Já Salway (2008), por meio também de uma análise automatizada,

trabalhou com aspectos linguísticos de 91 roteiros de audiodescrição com o objetivo de investigar as singularidades da linguagem utilizada nesses roteiros.

As pesquisas em audiodescrição acima mencionadas se relacionam com o estudo em questão porque utilizam a pesquisa baseada em *corpus* para investigar padrões e regularidades presentes no *corpus*. Já as pesquisas em legendagem mencionadas estão mais próximas ainda porque compartilham o mesmo tema e investigam padrões e regularidades em parâmetros da legendagem como a explicitação e a segmentação, esta última é foco de investigação da pesquisa proposta.

No âmbito das pesquisas em legendagem começo por Perego (2003), que teve como objetivo identificar os casos de explicitação na legendagem interlinguística e propor uma categorização para as mesmas, partindo do pressuposto que o fenômeno da explicitação ocorre na legendagem tanto quanto nas outras modalidades de tradução. Para identificar os casos de explicitação na legendagem, Perego compilou um *corpus* com legendas interlinguísticas (Húngaro/Italiano) de 2 filmes húngaros lançados ambos em 1970 no cinema e realizou uma análise manual, ou seja, sem auxílio de ferramentas computacionais, na qual as legendas na língua alvo (húngaro) e os diálogos na língua fonte (italiano) foram comparados por alinhamento a fim de, pelo contraste, identificar como a explicitação acontecia nas legendas e a partir daí categorizar os tipos encontrados. Após análise, tal categorização foi sugerida para os tipos de explicitação encontrados no *corpus*: Cultural; Baseada no canal (semiótico); e Baseada na redução, que poderiam acontecer por adição e por especificação. Os resultados da pesquisa mostraram que a explicitação tem função compensatória por equilibrar importantes perdas do texto fonte, aproximando seus efeitos na língua alvo, também tem função facilitadora, que torna o produto alvo mais fácil, simples, dando informações mais detalhadas para que o espectador possa entender sem dificuldades.

A investigação do parâmetro de explicitação foi ampliada na pesquisa de Perego (2009) que teve como objetivo observar a natureza das explicitações que ocorriam nas informações não verbais das legendas interlinguísticas. A informação não verbal consiste em elementos não orais categorizados como: não verbais sonoros e não verbais não sonoros, que incluem elementos essenciais na comunicação: prosódia e entonação, bem como cinésica e proxêmica. Ao investigar esses elementos, Perego (2009, p. 59) definiu outra tipologia,

diferente daquela definida em sua pesquisa anterior, para os casos de explicitação que participam ativamente na codificação das informações não verbais, a saber: adição (inserção de elementos linguísticos ausentes no original); especificação (opera no nível lexical da linguagem e envolve a substituição de uma unidade lexical no texto fonte por outra mais precisa e específica no texto alvo); e reformulação (que opera no nível textual e envolve a substituição de uma oração ou frase com menos informação por outras com mais informações). Acredito que essa tipologia da explicitação encontrada em Perego (2009) se adéqua melhor ao tema da legendagem, tanto por ser mais intuitiva, quanto por estar mais próxima dos termos encontrados na literatura de legendagem.

No que tange a segmentação na legendagem, Perego desenvolveu duas pesquisas correlatas que investigaram a segmentação sob uma ótica descritiva (PEREGO, 2008) e sob uma ótica experimental (PEREGO, 2010). A primeira investigou a segmentação linguística entre linhas de legendas (*line breaks*) de um *corpus* heterogêneo de legendas de filmes para DVD e para cinema, composto por diferentes gêneros fílmicos, diferentes anos de lançamento e direcionado para todos os públicos. Tal análise focou casos de má segmentação baseando-se no estudo de Karamitroglou (1998) que propõe que a segmentação deve ser pautada nos mais altos níveis sintáticos possíveis. Isso significa que, idealmente, uma oração deve compreender uma linha de legenda, mas nem sempre a oração completa cabe em uma linha só, e quando isso acontece o texto da oração deve ser distribuído em duas linhas. A partir desse ponto, Perego (2008) começou a investigar a natureza dos problemas de segmentação na legendagem de modo a investigar em que parte do texto se deu a quebra da linha e qual foi o tipo de problema encontrado. Em consequência dessa investigação foram encontrados problemas de segmentação nos sintagmas nominal, preposicionado, verbal, e problemas nas orações coordenadas e subordinadas.

A pesquisa acima mencionada impulsionou bastante o atual estudo, pois tomando como base as categorias de má segmentação propostas por Perego (2008) foi possível criar etiquetas para ampliar a análise de segmentação em LSE. As etiquetas propostas nesse trabalho, explicadas mais adiante no capítulo de metodologia, possibilitaram encontrar a frequência com que cada categoria apareceu no *corpus* de LSE do filme *Nosso Lar*, em DVD e a partir dessa frequência foi possível descrever e caracterizar como os problemas de segmentação acontecem nas legendas de filmes.

A segunda pesquisa (PEREGO, 2010) continuou investigando problemas de segmentação na legendagem, desta vez, sob um viés experimental, no qual o processamento de legendas bem e mal segmentadas foi verificado, dentre outros aspectos, através de rastreamento ocular. A triangulação dos dados contou com dados do rastreador ocular, dados de questionários aplicados e dados do desempenho dos participantes (obtidos através do reconhecimento das cenas e das palavras). O resultado relacionado ao desempenho dos participantes foi o mesmo, tanto quando expostos a legendas bem segmentadas, quanto quando expostos a legendas mal segmentadas. Esse resultado vai de encontro às pesquisas desenvolvidas no âmbito do processamento de legendas verificado via rastreamento ocular, que comprovaram que o desempenho dos participantes era influenciado por várias questões, dentre elas, a segmentação (D'Ydewalle, et al. 1985, 1987). Um dos fatores que pode ter contribuído para esses resultados consiste na homogeneidade dos participantes: homens e mulheres com média de 25 anos, pertencentes ao contexto universitário. Talvez se esse estudo tivesse investigado o desempenho de sujeitos em diferentes faixas etárias (crianças, jovens, adultos e idosos) em contextos socioculturais diferentes, os resultados teriam se configurado de modo diferente. Outro fator coadjuvante pode ter sido a escolha do trecho do filme legendado que não apresentava grandes problemas de segmentação linguística.

Outro estudo na esfera da legendagem que dialoga com a pesquisa proposta é a tese de doutorado de Feitosa (2009) que realizou pesquisa baseada em *corpus* para comparar legendas de fãs ou piratas (conhecidas como *fansubs*) com legendas comerciais, para isso analisou trechos de 10 filmes da década de 2000 e correlacionou os resultados buscando uma caracterização da legendagem sob os parâmetros técnicos (número de linhas, localização das legendas na tela, número de caracteres por linha, tipo de letra, cor das legendas, duração das legendas, tipo de alinhamento das legendas, marcação), e os parâmetros linguísticos (redução, condensação, omissão, segmentação) (DÍAZ-CINTAS, 2003; DÍAZ-CINTAS; REMAEL, 2007). Para essa análise descritiva o pesquisador fez uma comparação por alinhamento das legendas piratas e comerciais. Também foi analisado o conceito de explicitação (por adição e especificação), observado nas diferenças culturais, na mudança de canal semiótico e na redução, baseando-se no conceito encontrado em Perego (2003). Para tal análise o pesquisador fez um procedimento de anotação do *corpus* com etiquetas para cada tipo de explicitação e contabilizou os dados com auxílio do concordanciador do programa de análise linguística *WordSmith Tools*. Por último analisou o fluxo da informação ao longo de cada

legenda, observando aspectos relacionados ao Método de Desenvolvimento (FRIES, 1995; 2002; HALLIDAY; MATTHIESSEN, 2004; THOMPSON, 2004; 2007), também por meio de anotação do *corpus*, utilizando Códigos de Rotulação Sistemico Funcional.

Os resultados encontrados em Feitosa (2009) mostraram certas diferenças entre os dois tipos de legendagem: sobre os parâmetros técnicos foram encontrados maior número de caracteres por legenda nas legendagens piratas, a localização das legendas na tela e o tipo de letra nas legendas piratas são escolhidos pelo usuário espectador, enquanto nas legendas comerciais esses aspectos já vêm predeterminados sem possibilidade de mudança, e sobre os parâmetros linguísticos houve maior redução, condensação e omissão do texto nas legendas comerciais. Sobre o fenômeno da explicitação os resultados mostraram números próximos de ocorrências nos dois tipos de legendagem, sendo a categoria de explicitação mais frequente aquela motivada por aspectos culturais, realizada por meio de adição. E por último, o padrão de Método de Desenvolvimento mais frequente no *corpus* foi a configuração baseada em Progressão Temática derivada do Tema, ou seja, quando a legenda é informação retomada daquilo que se apresenta na tela do filme.

Outra pesquisa importante na fundamentação do atual estudo foi a da presente autora, Chaves (2009) que teve como objetivo verificar a viabilidade do uso de *corpora* na análise e descrição de alguns parâmetros da LSE. Para isso, foi montado um *corpus* com legendas de programas da TV Globo, as mesmas utilizadas na segunda pesquisa realizada pelo grupo LEAD (ARAÚJO, 2008) que propôs um modelo de legendagem para surdos baseado nas sugestões de um grupo de consultores surdos do CAS (Centro de Atendimento ao Surdo) de Fortaleza - Ceará. Os parâmetros analisados por Chaves (2009) foram densidade lexical (relacionada à quantidade de palavras por minuto de uma legendagem), nível de condensação (relacionado ao percentual de fala traduzido), segmentação (relacionada à divisão das legendas e à distribuição do texto nas legendas) e explicitação por adição e redução (adição de informações entre colchetes, como efeitos sonoros e identificação de falantes, e redução de elementos linguísticos com fins de respeitar os parâmetros de caracteres por minuto, ou com fins de facilitar a compreensão do produto audiovisual traduzido). Os procedimentos que antecederam a análise eletrônica do *corpus* foram: anotação e alinhamento. A anotação do *corpus* foi manual e feita através de etiquetas específicas para esse fim: i) etiqueta de ordem numérica da legenda, legenda 1 = <L1>, legenda 2 = <L2>, assim por diante; ii) etiqueta de

identificação do falante 1 = <Fal1>, do falante 2 = <Fal2>; iii) e etiqueta de identificação de efeitos sonoros = <ES>. O alinhamento que consistiu na comparação do texto fonte (transcrição dos diálogos) com o texto alvo (legendas para surdos da TV Globo ou legendas para surdos propostas pela equipe) através do Utilitário *Viewer and Aligner* do programa *Wordsmith Tools 5.0*.

Os resultados da pesquisa mostraram que as legendas da Rede Globo são mais densas e menos condensadas que as legendas feitas pelo LEAD. A segmentação só pôde ser analisada no que diz respeito ao número de legendas e os resultados dessa segmentação mostraram que a equipe divide o texto em mais legendas que a Globo. Com esse trabalho pôde-se concluir que o uso de *corpora* na análise e descrição de legendas não só é viável como possibilita uma descrição sistematizada dos parâmetros. Este trabalho foi um marco importante nos estudos desenvolvidos pelo LEAD, pois foi o primeiro a desenvolver uma pesquisa baseada em *corpus*.

Outra pesquisa importante para a fundamentação deste trabalho foi a tese de doutorado de Kalantzi (2008) que realizou pesquisa baseada em *corpus* ao compilar um *corpus* com 360 legendas de dois documentários e um programa de atualidades da BBC e teve como foco analisar e descrever dois parâmetros linguísticos da LSE (segmentação e redução), além de descrever os parâmetros técnicos (estratégia de tradução, tempos inicial e final, tempo de exibição, caracteres por legenda, velocidade da legenda, restrição de tempo, intervalo entre legendas, duração dos intervalos entre legendas, número de palavras por legenda, natureza das exclusões, número de linhas por legenda e erros de segmentação). Para a análise do parâmetro linguístico de segmentação a pesquisadora procedeu com anotação automática do *corpus* se valendo do etiquetador *Claws5* e do conjunto de etiquetas morfossintáticas desenvolvidos pelo Centro de Pesquisa da Universidade de Lancaster UCREL⁷. Para análise do parâmetro de redução fez comparação por alinhamento entre as legendas e as respectivas transcrições das falas dos programas, utilizando o comparador de documentos do *Word*. Para os demais parâmetros técnicos procedeu com anotação manual do *corpus* criando as próprias etiquetas. Além de objetivar descrever parâmetros da LSE, Kalantzi (2008) teve como meta prescrever regras linguísticas para a segmentação gramatical de legendas.

⁷ Sigla da *University Centre for Computer Corpus Research on Language*.

Sobre os resultados relacionados à segmentação verificou-se que das 360 legendas, 89 compreendiam um total de 92 erros de segmentação gramatical e a partir desses erros, Kalantzi propõe regras de segmentação gramatical: 1) Não separar sujeito de predicado; 2) Não separar um verbo do seu objeto direto; 3) Não separar o determinante do nome; Não separar o adjetivo do nome que ele modifica; Não separar nomes próprios, nem nome de sobrenomes, nem título do nome; 4) Não separar nomes numa sequência nome-nome incluindo colocações; 5) Não separar possessivo/genitivo do nome que ele modifica; 6) Não separar o advérbio do adjetivo que ele modifica; 7) Não separar preposição de uma frase preposicionada; 8) Não separar preposições complexas; 9) Não separar o advérbio do verbo que ele modifica; 10) Não separar o marcador de infinitivo do seu infinitivo; 11) Não separar formas verbais complexas; 12) Não separar verbos frasais; 13) Não separar conjunção subordinada do resto da oração subordinada; 14) Não separar pronome relativo/advérbio do resto da oração relativa; 15) Não separar uma conjunção coordenada da oração que ela inicia; 16) Não separar colocações; 17) Não hifenizar as palavras.

Os resultados sobre o parâmetro de redução mostraram que Kalantzi identificou e categorizou 18 tipos de redução nas legendas e para essas 18 categorias propostas preferiu dividir em duas grandes categorias como propostas no trabalho até então não publicado de Eugeni (2008): 1) Redução semântica (quando diz respeito à exclusão de lexemas ou frases), 2) Redução não semântica (quando envolve a exclusão de palavras de função, marcas de oralidade e lexemas que não sejam de conteúdo semântico).

Sobre os parâmetros técnicos seria pouco interessante apontar os resultados de cada um, visto que a pesquisadora realizou uma descrição pormenorizada de cada parâmetro sempre comparando aos parâmetros vigentes na Europa e avaliando se os parâmetros encontrados ou os vigentes eram ideais ou não para os espectadores. Para obter a opinião dos espectadores Kalantzi não realizou pesquisa exploratória, mas sim um apanhado dessas opiniões na literatura da área.

Sobre a pesquisa acima resenhada (KALANTZI, 2008) considero que o tema Legendagem para Surdos e Ensurdidos, o foco na segmentação e a proposta de utilizar a metodologia baseada em *corpus* dialogam muito bem com a presente pesquisa, porém as estratégias utilizadas para realizar a pesquisa bem como os vieses de análise descritivo,

prescritivo e avaliativo, utilizados ao mesmo tempo ao longo de toda a análise, fogem à proposta desta pesquisa que tem caráter predominantemente descritivo.

Os estudos resenhados e os resultados da atual pesquisa (disponíveis no capítulo 4) comprovam aquilo que foi previsto por Baker em 1993 de que os estudos com *corpora* dariam um avanço nos estudos da tradução, contribuindo com a investigação de questões importantes na tradução como a formação de tradutores.

Diante desse panorama sobre os estudos de TAV baseados em *corpora* é possível traçar uma perspectiva positiva de crescimento para esse tipo de pesquisa, tendo em vista que a pesquisa baseada em *corpus* está intimamente ligada aos estudos descritivos da tradução, e é merecido informar que os estudos descritivos da tradução estão fazendo toda a diferença nas esferas profissional e acadêmica porque dão a oportunidade de pesquisadores conhecerem bem seu objeto de estudo a ponto de contribuir com o aperfeiçoamento das práticas e treinamentos de recursos humanos.

Essa revisão de literatura mostrou que o ponto forte das pesquisas é a metodologia e essa tendência é observada como característica da pesquisa baseada em *corpus*. Portanto, isso só aumenta a certeza de que a união entre TAV e Linguística de *Corpus* veio pra ficar e que a metodologia baseada em *corpus* já está se estabelecendo como preferência dentre os pesquisadores da área e por essa razão ganha cada vez mais forças em pesquisas no âmbito da tradução audiovisual e da acessibilidade audiovisual.

2.2 Legendagem para Surdos e Ensurdidos

Como definida no capítulo anterior, a LSE consiste num tipo de tradução que se assemelha em grande parte à legendagem para ouvintes, com a diferença que a LSE traduz em texto escrito os efeitos sonoros e identifica os falantes com o intuito de que surdos e ensurdidos possam construir a narrativa do produto audiovisual através dos elementos sonoros e da distinção dos falantes.

Sobre as questões técnicas da LSE pode-se dizer que possuem no máximo duas linhas; são normalmente de cor branca ou amarela, duram no mínimo 1 e no máximo 6 segundos

(D'YDEWALLE, et al. 1987), e possuem velocidades diferentes. Contudo, de acordo com pesquisas experimentais com rastreador ocular, observou-se que há três níveis de velocidade de leitura que permitem que o espectador harmonize a leitura das legendas com o áudio e as imagens do filme: 145 palavras por minuto (ppm), 160ppm e 180ppm (D'YDEWALLE, et al. 1987; IVARSSON; CARROLL, 1998; DIAZ CINTAS; REMAEL, 2007).

Essa harmonização entre imagem, áudio e legendas é conquistada com a redução do texto para que o espectador tenha tempo de ler as legendas e desfrutar das imagens, além de escutar o áudio, no caso dos ouvintes. A redução é necessária porque o tempo que se gasta para ler é maior que o tempo que se gasta para escutar, portanto se as legendas estiverem muito rápidas o entendimento das legendas por parte dos espectadores pode ficar comprometido sendo necessário se valer de redução para que essa harmonização aconteça.

Para garantir a harmonização entre produto audiovisual, legenda e áudio é necessário seguir os parâmetros técnicos e linguísticos da legendagem, além das convenções e pontuações linguísticas vividas em cada contexto. De antemão, exponho quão difícil é para o pesquisador se responsabilizar pela boa compreensão das legendas além de se responsabilizar pela harmonização entre legenda e imagem, pois há vários fatores que podem influenciar no desempenho dos espectadores, como o cultural, o conhecimento de mundo, a familiaridade dos surdos com o tema, dentre outros, portanto é necessário ressaltar que essa garantia está relacionada a um contexto em que os espectadores tenham níveis de leitura normais, como observado nas pesquisas de processamento de legendas (D'YDEWALLE, et al. 1987).

Voltando à questão dos parâmetros, corroboro a classificação proposta em Diaz Cintas e Remael (2007) que estabelece que os parâmetros técnicos podem ser divididos em duas dimensões: espacial e temporal. A dimensão espacial envolve fatores relacionados à disposição e aparência das legendas na tela, são elas: número de linhas e posição na tela, tipo de fonte e número de caracteres por linha, enquanto a temporal envolve fatores como o tempo corrente das legendas em relação ao tempo corrente das produções audiovisuais, que consistem em: marcação e duração das legendas, sincronização, intervalo entre legendas consecutivas, velocidade de leitura e tempo de exposição das legendas.

Os parâmetros linguísticos podem ser elencados da seguinte forma: redução do texto por condensação ou reformulação, omissões ou exclusões, coesão e coerência, segmentação e quebra de linhas, segmentação retórica e segmentação visual.

E por fim, sem esgotar o assunto, Diaz Cintas e Remael (2007) elencam algumas convenções linguísticas observadas na legendagem, no âmbito das pontuações: vírgula e ponto e vírgula, ponto final, dois pontos, parênteses e colchetes, pontos de exclamação e interrogação, travessão e hífen, reticências, asteriscos, barras, outros símbolos, letra maiúscula e aspas em citações; e no âmbito da formatação: Itálico em música, cartas e documentos escritos traduzidos em produtos audiovisuais legendados, cores, abreviações, números cardinais e ordinais, horário e unidades de peso e medida.

As convenções linguísticas não serão levadas em conta na análise desse *corpus*, apenas alguns dos parâmetros técnicos e linguísticos, como número de linhas, caracteres por linha, velocidade das legendas e segmentação gramatical e retórica. Isso porque não haveria tempo suficiente para investigar e descrever todos os parâmetros, depois, descrever tudo fugiria completamente à proposta, pois o foco desta pesquisa é na segmentação e em menor escala nos outros parâmetros.

Sobre o número de linhas é possível relatar que, em geral, possuem no máximo duas linhas, mas o mesmo não acontece na TV, especialmente no sistema estadunidense de *closed caption*⁸ adotado pela Rede Globo, no qual é comum encontrar legendas de até três linhas. Diaz Cintas e Remael (2007, p. 86) explicam que a escolha entre uma ou duas linhas depende do aspecto linguístico da legenda, que deve sempre estar de acordo com a sintaxe e a semântica no intuito de contribuir para a boa leitura e aspectos estéticos da legenda. Na literatura é possível encontrar, no âmbito das pesquisas experimentais com rastreador ocular, outra discussão sobre o processamento de legendas de uma ou duas linhas, desta vez, com foco na preferência por parte dos espectadores. Alguns pesquisadores defendem que legendas de duas linhas são mais fáceis de interpretar porque contém mais informação distribuída no texto (IVARSSON; CARROLL, 1998; DÍAZ CINTAS, 2003), enquanto outros defendem que, por serem menores e compreenderem menos informação no texto, legendas de uma linha

⁸ O sistema norte americano de legenda para surdos é o *closed caption*, adotado no Brasil pela Rede Globo, que consiste na transcrição dos diálogos sem redução do texto e na tradução das informações adicionais de identificação de falante e de efeito sonoro.

são mais fáceis de entender porque sua semântica e estrutura contém menos informação e são mais redundantes por conta das imagens (D'YDEWALLE, et al. 1990).

Sobre o número de caracteres por linha não há regras que determinem um número ideal, porém costuma-se encontrar algo em torno de 35. Isso não significa que numa legenda de 40 caracteres deve-se encher a primeira linha com 35 caracteres e a segunda com apenas 5. Para que isso não ocorra deve haver uma preocupação, por parte do tradutor, com a segmentação linguística. Além disso, a quantidade de caracteres por linha de legenda também varia de acordo com o cliente e o meio audiovisual no qual será exibida. Segundo Diaz Cintas e Remael (2007, p. 84) a TV exibe uma média de 37 caracteres por linha, alguns clientes exigem de 33 a 35 ou 39 a 41 caracteres por linha dependendo das imposições. Ocasionalmente, alguns festivais de filme chegam a ultrapassar 43 caracteres por linha e para o cinema e o DVD é normal encontrar até 40 caracteres por linha. O número máximo de caracteres por linha tomado como referencia para esta pesquisa foi de 35, visto que esse valor é praticado pelos integrantes do LEAD.

Sobre a velocidade das legendas pode-se expor, a partir dos resultados de pesquisas experimentais sobre o processamento de legendas feitas por rastreamento ocular (D'YDEWALLE, et al. 1987), três velocidades recorrentes na leitura de legendas: 145, 160 e 180ppm, que foram observadas em legendas de duração mínima de 1 segundo e máxima de 6 segundos. Diaz Cintas e Remael (2007), por sua vez, a partir dos dados encontrados em D'Ydewalle, et al. (1987) fizeram uma tabela equivalente com as velocidades apresentadas em caracteres por segundo (cps) ao invés de ppm. Isso porque no contexto dos programas de legendagem as medidas utilizadas para a tradução para legendas são dadas em caracteres por segundo e não em palavras por minuto. O programa de legendagem utilizado como referência no trabalho de Diaz Cintas e Remael (2007) foi o programa profissional utilizado por emissoras de televisão do Reino Unido, *WinCAPS* da Companhia *SysMedia*⁹. E o programa de legendagem utilizado como referência para esta pesquisa foi o *Subtitle Workshop*, da *UruSoft*¹⁰, programa gratuito amplamente utilizado por amadores, profissionais e empresas de legendagem.

⁹ Sítio da SysMedia. Disponível em: <www.sysmedia.com>. Acesso em: 10 de fevereiro de 2012.

¹⁰ Sítio da UruSoft. Disponível em: <www.urusoft.net>. Acesso em: 10 de fevereiro de 2012.

Abaixo é possível visualizar as tabelas de velocidade de leitura de legendas adaptadas para a unidade de medida, caracteres por segundo, propostas em Diaz Cintas e Remael (2007), com base nas velocidades de leitura de 145, 160 e 180ppm encontradas em D'Ydewalle, et al. (1987).

Tabela 1: Velocidade de leitura da legenda a 145ppm detalhada em cps.

145 palavras por minuto		Seconds:	Spaces	Seconds:	Spaces
		frames		Frames	
		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		
Seconds:	Spaces	Seconds:	Spaces	Seconds:	Spaces
Frames		Frames		frames	
03:00	44	04:00	58	05:00	70
03:04	46	04:04	60	05:04	71
03:08	48	04:08	62	05:08	72
03:12	50	04:12	64	05:12	73
03:16	52	04:16	65	05:16	73
03:20	54	04:20	67	05:20	74
				06:00	74

Fonte: Adaptada de Diaz Cintas e Remael (2007, p. 97)

Tabela 2: Velocidade de leitura da legenda a 160ppm detalhada em cps.

160 palavras por minuto		Seconds:	Spaces	Seconds:	Spaces
		frames		Frames	
		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		
Seconds:	Spaces	Seconds:	Spaces	Seconds:	Spaces
Frames		Frames		frames	
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: Adaptada de Diaz Cintas e Remael (2007, p. 98)

Tabela 3: Velocidade de leitura da legenda a 180ppm detalhada em cps.

180 palavras por minuto		Seconds:	Spaces	Seconds:	Spaces
		frames		Frames	
		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		
Seconds:	Spaces	Seconds:	Spaces	Seconds:	Spaces
Frames		Frames		frames	
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: Adaptada de Diaz Cintas e Remael (2007, p. 99)

Para melhor entender as unidades de medida da velocidade de leitura é necessário compreender as diferenças entre os dois programas de legendagem, visto que ambos adotam unidades de medidas um pouco distintas. O *WinCAPS* marca o tempo das legendas no formato (00:00:00:00) hora: minuto: segundo: **frames**, e o *Subtitle Workshop* marca o tempo das legendas no formato (00:00:00:000) hora: minuto: segundo: **milissegundos**. A diferença entre os dois programas acontece apenas na menor unidade, na qual o *WinCAPS* trabalha com segundo:frames (00:00) e o *Subtitle Workshop* com segundo:milissegundos (00:000).

Esses estudos mostraram que a velocidade de leitura serviu de parâmetro para a velocidade das legendas, pois a partir dos padrões de leitura de legendas encontrados em espectadores europeus (145, 160 e 180ppm) foi possível parametrizar a velocidade das legendas na mesma escala. Portanto a velocidade da legenda consiste na razão entre a quantidade de palavras ou caracteres pelo tempo de exibição das legendas em minuto ou segundo respectivamente.

Para esta pesquisa, as velocidades das legendas foram caracterizadas como baixa, ideal e alta, mas como explicar esses conceitos em palavras por minuto ou caracteres por segundo? Essa explicação está presente na pesquisa de recepção de LSE, por surdos cearenses, que definiu um modelo de legendagem, como descrito no capítulo de introdução, testando as três velocidades de legendas: 145, 160 e 180ppm. Os resultados mostraram que a velocidade ideal, que permite harmonização entre legenda e imagem, é a de 145ppm (entre 14 e 15cps) (ARAÚJO, 2008). Essa velocidade, eleita como ideal, serve de parâmetro para a atual

pesquisa, desta forma, toda legenda com velocidade abaixo de 14cps é considerada baixa e toda legenda com velocidade a partir de 16cps é considerada alta. Em resumo, as legendas podem assumir velocidades baixa (até 13cps), ideal (entre 14 e 15cps) e alta (a partir de 16cps). Mas a recepção das três velocidades continua sendo testada por surdos de todo o Brasil e os resultados estão sugerindo que legendas bem segmentadas são compreendidas em qualquer uma das três velocidades. (ARAÚJO; NASCIMENTO, 2011). Porém, mais uma vez reforço que o parâmetro de velocidade tido como ideal para esta pesquisa é o de 145ppm.

A segmentação é um parâmetro fundamental da legendagem que está relacionado à divisão do texto e das legendas. Sobre a segmentação Ivarsson e Carroll (1998, p. 74) afirmam: “Qualquer irritação ou inconsistência experienciada pelos espectadores atrasa o processo de compreensão e faz dispersar do foco principal, o filme em si.”¹¹

Segundo Reid (1996, p.100) a segmentação pode acontecer de três formas: pela gramática (pautada pela semântica), pela retórica (pautada pelo ritmo da fala), ou pelo visual (pautada pelo que acontece na cena, como cortes, mudanças de ângulo etc.). Porém, esta pesquisa se deteve em explorar apenas problemas de segmentação nos níveis linguísticos e retóricos da legendagem para surdos do filme brasileiro *Nosso Lar* lançada para DVD.

A segmentação visual não entrou para a análise porque não houve tempo suficiente para expandir esta pesquisa e também pelo fato de que este tipo de segmentação está mais situado na dimensão espacial da legendagem que na dimensão linguística, mas é importante explicar que o problema de segmentação visual acontece quando a legenda não acompanha a mudança de cena e acaba se estendendo ao quadro seguinte, ou na linguagem dos legendistas: quando a legenda “vaza”. Quando isso acontece, os profissionais legendistas dizem que a legenda “vazou” para a outra cena. Isso não torna a segmentação visual menor, do ponto de vista científico, mas a desvia dos objetivos desta pesquisa, muito embora uma etiqueta de análise de problema de segmentação visual tenha sido criada com vistas a ampliar a investigação da segmentação na LSE em futuras pesquisas.

Em Diaz Cintas e Remael (2007, p. 172) pode-se entender a segmentação em dois níveis: o primeiro diz respeito à divisão dentro da mesma legenda, chamado "quebra de linha"

¹¹ Minha tradução de: *Each irritation or inconsistency experienced by viewers slows down the comprehension process and distracts from the real issue at hand, the film itself.*

(*line break*) e o segundo diz respeito à divisão ou segmentação entre duas ou mais legendas. Sobre a segmentação retórica Diaz Cintas e Remael (2007, p. 179) expõem que ela tenta marcar as características da linguagem oral nas legendas levando em consideração hesitações e pausas e afirmam que boa segmentação retórica ajuda a manter surpresa, suspense, ironia, hesitação etc. Essas características prosódicas da linguagem oral têm como finalidade garantir a qualidade da mensagem, portanto, os problemas de segmentação retórica acontecem quando a surpresa, o suspense etc. não são mantidos nas legendas.

Para que a legibilidade da legenda seja garantida, alguns aspectos precisam ser observados com relação à segmentação, para isso, ponho em foco a proposta de Karamitroglou (1998), que sugere que a divisão do texto seja feita no mais alto nível sintático possível.

Quando segmentamos uma sentença, forçamos o cérebro a pausar seu processo linguístico por um instante, até os olhos localizarem a próxima informação linguística. Nos casos em que a segmentação é inevitável devemos tentar levar o cérebro a fazer essa pausa num ponto onde a carga semântica se configure numa informação completa.¹²

Essa ideia representa muito bem os objetivos desta pesquisa e significa que, idealmente, cada legenda deve conter um pensamento completo, ou seja, quando uma construção linguística é muito grande e precisa ser distribuída em mais de uma linha, cada linha deve procurar conter uma informação mais completa possível. Para Diaz Cintas e Remael (2007, p. 173) o proposto por Karamitroglou (1998) diz em outras palavras que se a segmentação não for cuidadosa os leitores farão esforço para decodificar o texto e consequentemente poderão se cansar mais rapidamente e perder o prazer proporcionado pelo produto audiovisual. As legendas precisam ter um pensamento completo para serem entendidas no curto espaço de tempo em que são exibidas (DIAZ CINTAS; REMAEL, 2007, p. 172). A proposta de Karamitroglou (1998) pode ser visualizada com mais clareza pela árvore sintática também proposta pelo autor.

¹² 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. In cases where segmentation is inevitable, therefore, we should try to force this pause on the brain at a point where the semantic load has already managed to convey a satisfactorily complete piece of information.*

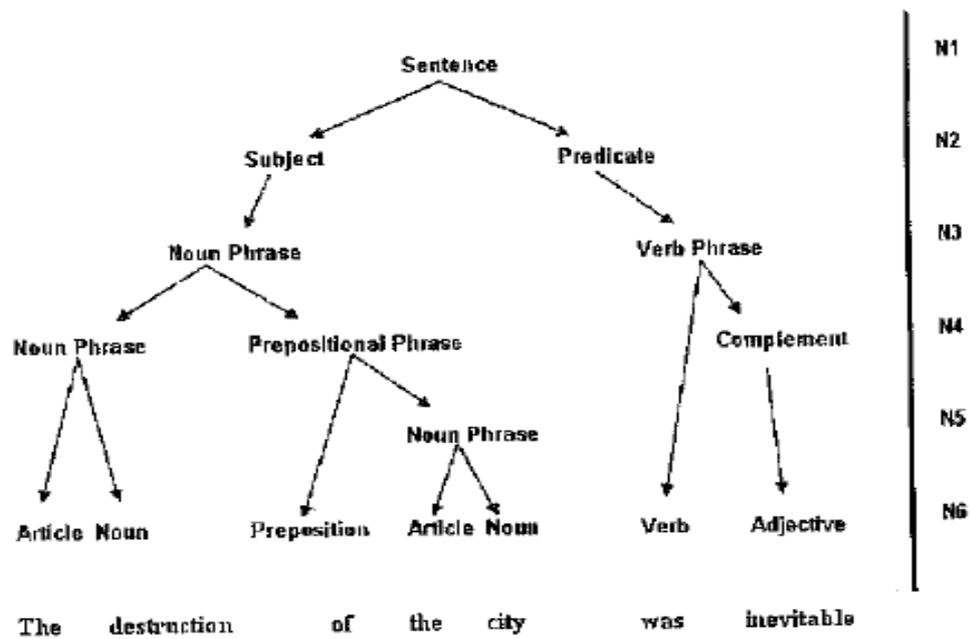


Figura 3: Árvore Sintática proposta por Karamitroglou (1998)
 Fonte: Karamitroglou (1998)

Essa árvore sintática é constituída de seis nós (construções sintáticas) sendo o 1º representado pela oração completa e o 6º representado pelas palavras ou morfemas. Alguns exemplos ajudam a entender melhor o que significa o mais alto nível sintático possível.

A construção analisada na árvore é *The destruction of the city was inevitable* (Em português: *A destruição da cidade foi inevitável*). Essa construção linguística contém 43 caracteres e deve ser distribuída numa legenda. Como não é possível alocar essa construção em apenas uma linha de legenda ela deve ser dividida em duas linhas.

A segmentação no primeiro nó (N1) faria apenas uma linha de legenda, mas essa não seria indicada porque ultrapassa o número máximo de 35 caracteres por linha

The destruction of the city was inevitable

A segmentação no segundo nó (N2) criaria duas linhas de legenda nos mais altos níveis sintáticos possíveis

*The destruction of the city
 was inevitable*

A segmentação no terceiro nó (N3) seria igual a do segundo nó (N2), por isso também se configura nos mais altos níveis sintáticos

*The destruction of the city
was inevitable*

A segmentação no quarto nó (N4) quebraria a estrutura interna do sintagma preposicionado (*of the city*)

*The destruction of
the city was inevitable*

A segmentação no quinto nó (N5), assim como a do quarto nó, também quebraria a estrutura interna do sintagma preposicionado (*of the city*)

*The destruction of **the**
city was inevitable*

E por fim, a segmentação no sexto nó (N6) não pode ser ilustrada, porque se encontra no nível da palavra permitindo que a legenda seja dividida em qualquer ponto da construção e isso diminui as chances de a legenda conter um pensamento completo.

The/ destruction/ of/ the/ city/ was/ inevitable

Dos exemplos dados, a segmentação igualmente representada em (N2) e (N3) se configura como a que promove melhor leitura, pois a linha de cima contém um nível sintático completo (sintagma nominal) e a linha de baixo contém outro nível sintático completo (sintagma verbal). Na árvore, a nomenclatura do (N2) consiste em sujeito e predicado e a do (N3) consiste em sintagmas nominal e verbal.

Além dos 3 critérios de segmentação, há outro pouco privilegiado por profissionais e pesquisadores que trata da estética da legenda. Esse tipo de segmentação recebe críticas porque obedece a um padrão bastante formalista de segmentação pautado pela perfeita geometria da legenda, em que as linhas superior e inferior se encontram com a mesma

extensão, mesmo quando isso é inviável do ponto de vista linguístico e retórico. Karamitroglou (1998) acredita que deve haver um compromisso entre sintaxe e geometria, mas quando é necessário sacrificar um dos critérios de segmentação em prol de outro, deve-se sacrificar a geometria.

Embora tenha adotado o termo encontrado na literatura, ‘segmentação gramatical’, prefiro o termo ‘segmentação linguística’ (não encontrado na literatura), visto que a acepção de gramática está arraigada de preconceito e por isso acaba conduzindo a discussão por um viés normativo da língua. Entretanto gostaria de esclarecer que a gramática está presente neste trabalho não para dizer o que é certo ou errado na segmentação das legendas, mas para explicar o seu funcionamento. Este trabalho abre espaço para uma discussão sobre a segmentação das legendas do ponto de vista descritivo, com fins de compreender como essa segmentação é e como ela está representada na LSE, além de descrever, através de padrões e regularidades encontrados no *corpus*, como esse fenômeno acontece na legendagem.

Esta revisão de literatura pretendeu reunir os trabalhos e pesquisas mais relevantes da área que fundamentam teórica e metodologicamente o escopo dessa pesquisa. As visões, concepções, sugestões, ideias, dentre outras contribuições dessas pesquisas, com certeza aguçaram e muito o poder de reflexão da presente autora, que também pretende, ao final desse árduo, mas prazeroso trabalho, causar o mesmo naqueles a quem o tema interessar. Os próximos capítulos mostram o desenvolvimento da pesquisa e a análise e discussão dos dados.

3 METODOLOGIA

Este capítulo expõe a metodologia necessária para alcançar os objetivos desta pesquisa. O passo a passo pode ser acompanhado pelos sub-capítulos: tipo de pesquisa, contexto da pesquisa, *corpus*, instrumentos, tratamento do *corpus*, anotação do *corpus* e análise dos dados.

3.1 Tipo de pesquisa

A presente pesquisa pode ser definida como descritiva, de natureza quali-quantitativa, por procurar observar, classificar e interpretar o fenômeno da segmentação na LSE de DVDs sem interferir nas suas características com fins de modificá-lo. Os dados foram analisados e interpretados de forma mista, utilizando dados quantitativos e qualitativos para descrever o fenômeno. Com relação aos procedimentos esta pesquisa se caracteriza como baseada em *corpus* por envolver procedimentos característicos como: anotação do *corpus* (etiquetagem) e uso de programas de análise linguística, como o *WordSmith Tools*. Além disso, também se caracteriza como tal porque lida com texto autêntico, nesse caso a LSE, que é texto traduzido escrito, representativo da linguagem oral, em língua portuguesa. A pesquisa baseada em *corpus* requer duas etapas de análise, a primeira diz respeito à análise do *corpus*, viabilizada pela etiquetagem, e a última diz respeito à análise dos dados da pesquisa.

3.2 Contexto da pesquisa

Esta pesquisa está inserida num contexto acadêmico, como pesquisa de mestrado do Curso de Mestrado Acadêmico do PosLA - Programa de Pós-Graduação em Linguística Aplicada da UECE. O estudo em questão teve origem nas pesquisas exploratórias sobre LSE (explicitadas anteriormente), coordenadas pela Profa. Dra. Vera Lúcia Santiago Araújo, que visaram definir um modelo de legendagem para surdos no Brasil a partir da avaliação dos parâmetros da legendagem por parte dos surdos brasileiros. Diferentemente das pesquisas exploratórias, esta não envolveu seres humanos e é uma ampliação de outra pesquisa desenvolvida em nível de graduação que visou verificar a viabilidade do uso de *corpora* na análise e descrição dos parâmetros da LSE. Esse dado confere maior experiência ao pesquisador e atribui maior validade e consistência à pesquisa. Nesse contexto, essa pesquisa

amplia as discussões no âmbito da LSE além de incentivar e vincular ainda mais a metodologia baseada em *corpus* aos estudos da tradução.

O *corpus* e os resultados dessa pesquisa estão disponíveis no LATAV com vistas a ampliar as discussões e metodologias dos estudos em Tradução Audiovisual, Acessibilidade Audiovisual e Linguística de *Corpus* que acontecem no LATAV, desde 2002, no Centro de Humanidades da Universidade Estadual do Ceará.

3.3 *Corpus*

O *corpus* é do tipo especializado, ou seja, que visa atender às necessidades de um trabalho de pesquisa em particular, de acordo com os objetivos propostos e é composto das legendas para surdos e ensurdecidos em língua portuguesa do filme brasileiro *Nosso Lar* (2010) lançado comercialmente em DVD com os recursos acessíveis de LSE e AD. Há vários tipos de *corpora* além do especializado: Comparável bi- ou multilíngue - *corpus* composto por dois ou mais *subcorpora* com textos originais nas respectivas línguas, Comparável monolíngue - *corpus* composto por textos originais numa língua e traduções nessa mesma língua, *Corpus* de estudo - o *corpus* em que se baseia a pesquisa a ser desenvolvida, *Corpus* de referência - *corpus* que serve de termo de comparação para o *corpus* de estudo, em geral, deve ter três a cinco vezes o tamanho do *corpus* de estudo, *Corpus* monitor - *corpus* que é constantemente atualizado a fim de representar a evolução da língua, *Corpus* paralelo - *corpus* constituído de originais e suas respectivas traduções, *Corpus* estático - *corpus* que não permite acrescentar material novo, *Corpus* dinâmico - *corpus* que permite o acréscimo de material novo (TAGNIN, 2010).

No Brasil, são poucos os DVDs com algum recurso de acessibilidade, precisamente onze, porém, apenas sete contém legendagem para surdos, número pequeno se comparado ao número de DVDs lançados comercialmente com recursos acessíveis a surdos e pessoas com deficiência visual em países como Espanha, Portugal, Inglaterra, Estados Unidos e Austrália. O acervo de DVDs acessíveis no Brasil ainda é pequeno. Até o presente momento, os filmes lançados em DVD no Brasil com algum recurso de acessibilidade audiovisual como LSE, AD, janela de LIBRAS, menus com audionavegação e título do filme em BRAILLE, que seja do meu conhecimento, são: *Irmãos de Fé*, *O Signo da Cidade*, *Fronteira*, *Ensaio sobre a*

*Cegueira, Bezerra de Menezes: O Diário de um Espírito, Chico Xavier, Nosso Lar, Turma da Mônica em: Cine Gibi 5*¹³. Além dos filmes lançados comercialmente com recursos acessíveis, há também os filmes do projeto *DVD Acessível*¹⁴, projeto financiado pelo edital *Programa BNB de Cultura* em parceria com o BNDES em 2009, que corresponde à produção de DVDs acessíveis de dois filmes de longa metragem: *Corisco e Dadá, O Grão* e quatro filmes de curta metragem: *Adorável Rosa, Águas de Romanza, Reisado Miudim e Capistrano no quilo*¹⁵.

Nem todos os filmes com recursos acessíveis oferecem todos os recursos juntos. Alguns filmes oferecem AD, mas não oferecem LSE, outros oferecem AD e LSE, e apenas os dois filmes de longa metragem do projeto *DVD Acessível* oferecem AD, LSE, janela de LIBRAS e menus com audionavegação.

Mesmo sendo pequena a quantidade de DVDs com recursos acessíveis no Brasil, não seria possível analisar todos por causa do fator tempo, pois o processo de etiquetagem de um único filme, explicado mais adiante, foi árduo e teve duração de quase seis meses. Dentre as sete opções de DVD com legendagem para surdos apenas uma foi escolhida para análise, então optei pelo filme mais recente, *Nosso Lar* (2010) adaptado do livro homônimo do espírito André Luís, psicografado pelo médium Chico Xavier. O filme retrata a trajetória do médico André Luiz pelo mundo espiritual. A trama começa quando o médico acorda num lugar sujo e sombrio e descobre que não pertence mais ao mundo dos vivos. Após tempos de sofrimento nessas zonas purgatórias, ele é socorrido por espíritos de luz que o levam para uma cidade em outra dimensão, chamada *Nosso Lar*, aonde aprende novas lições marcadas por momentos de dor e sofrimento.

¹³ Ficha técnica dos DVDs com recursos de acessibilidade lançados comercialmente no Brasil: *Irmãos de Fé* (Dir: Moacyr Góes, Drama, 105', Sony Pictures, 2004, Brasil); *O Signo da Cidade* (Dir: Carlos Alberto Riccelli, Drama, 95', Europa Filmes, 2008, Brasil); *Ensaio sobre a Cegueira* (Dir: Fernando Meirelles, Drama, 121', Fox, 2008, Brasil); *Fronteira* (Dir: Rafael Conde, Drama, 85', Filmegraph e Camisa Listrada, 2008, Brasil); *Bezerra de Menezes: O Diário de um Espírito* (Dir: Glauber Filho e Joe Pimentel, Drama, 75', Fox, 2008, Brasil); *Chico Xavier* (Dir: Daniel Filho, Drama, 125', Sony Pictures, 2010, Brasil); *Nosso Lar* (Dir: Wagner de Assis, Drama, 102', Fox, 2010, Brasil); *Turma da Mônica em: Cine Gibi 5* (Dir: Maurício de Sousa, Infantil, 71', 2010).

¹⁴ Ficha técnica dos DVDs com recursos de acessibilidade lançados pelo projeto *DVD Acessível*: *Corisco e Dadá* (Dir: Rosemberg Cariry, 110', Colorido, 1996, Brasil); *O Grão* (Dir: Petrus Cariry, Drama, 88', Colorido, 2010, Brasil); *Adorável Rosa* (Dir: Aurora Miranda Leão, Documentário, 19', Colorido, 2008, CE); *Águas de Romanza* (Dir: Gláucia Soares e Patrícia Baía, Ficção, 15', Colorido, 2002, CE); *Capistrano no Quilo* (Dir: Firmino Holanda, Documentário, 21', Colorido, 2006, CE); *Reisado Miudim* (Dir: Petrus Cariry, Ficção, 13', Colorido, 2008, CE).

¹⁵ Os filmes do Projeto *DVD Acessível* contém os recursos de LSE, AD, janela de LIBRAS e menus com audionavegação. Toda a produção, bem como as traduções foram feitas pelo Grupo LEAD – UECE.

A questão do gênero fílmico não será levada em conta na análise dos dados, uma vez que não foi possível compilar um *corpus* com gêneros fílmicos diversos devido ao curto prazo de entrega da pesquisa e à escassez de filmes lançados em DVD com o recurso de LSE. O item que entrou na discussão desta pesquisa foi a legendagem para surdos e ensurdecidos lançada para o suporte DVD, independente do seu gênero fílmico. Essa delimitação ajuda a eliminar possíveis variáveis e confere maior rigor metodológico à pesquisa. No contexto dessa pesquisa a delimitação do *corpus* foi determinada pelo seu tamanho, pois quanto maior o *corpus*, maior o tempo de processamento e análise dos dados. Seria inviável etiquetar manualmente e analisar eletronicamente legendas para surdos dos seis filmes de longa metragem e dos quatro filmes de curta metragem no âmbito de uma pesquisa de mestrado.

3.3.1 Tamanho do *corpus* e Representatividade

O tamanho do *corpus* analisado foi de 6126 itens lexicais (*tokens*)¹⁶, ou palavras. Esse total de palavras equivale a 1132 legendas¹⁷ num filme de 102 min. Para a representatividade do *corpus* esse valor de 6126 palavras pode ser considerado pequeno, entretanto se trata de um *corpus* especializado que atende às necessidades dessa pesquisa sendo, portanto ideal para o que se propõe. Esses dados serão aprofundados no capítulo de análise e discussão dos dados. Como os *corpora* especializados não seguem um tamanho padrão não será possível classificar esse *corpus* pelo seu tamanho. A ideia de tamanho de *corpora* pode ser pensada a partir da compreensão de Bowker e Pearson (2002, p. 45):

Infelizmente, não há regras consistentes e seguras que possam ser seguidas para determinar o tamanho ideal de um *corpus*. Ao invés disso, você deverá tomar decisões baseadas em fatores como as necessidades de seu projeto, a disponibilidade de dados e o tempo que você tem. É muito importante, no entanto, não considerar o maior sempre melhor. Você pode descobrir que consegue obter mais informações úteis de um *corpus* pequeno e bem planejado, do que de um *corpus* maior que não atende as suas necessidades.¹⁸

¹⁶ Dado obtido através do gerador de lista de palavras *Word List* do conjunto de ferramentas do programa *WordSmith Tools 5.0*.

¹⁷ A saber: legenda é cada porção de texto exibida no mesmo instante e legendagem é um conjunto de legendas.

¹⁸ Minha tradução de: *Unfortunately, there are no hard and fast rules that can be followed to determine the ideal size of a corpus. Instead, you will have to make this decision based on factors such as the needs of your Project, the availability of data and the amount of time that you have. It is very important, however, not to assume that bigger is always better. You may find that you can get more useful information from a corpus that is small but well designed than from one that is larger but is not customized to meet your needs.*

3.4 Instrumentos

Foram instrumentos da pesquisa os programas de computador: *SubRip* – para extração das legendas dos filmes em DVD, *Word* – para edição de texto e anotação do *corpus*, além do gerador de lista de palavras *Word List* e do concordanciador *Concord*, ambos do conjunto de ferramentas do programa de análise linguística *WordSmith Tools 5.0* para auxiliar na análise dos dados.

3.5 Tratamento do *corpus*

O tratamento do *corpus* diz respeito aos procedimentos que antecedem e possibilitam a análise dos dados, são eles: compilação do *corpus*, por meio da extração das legendas, etiquetagem ou anotação do *corpus*, e por último, análise dos dados a partir de interpretação e utilização das ferramentas do programa de análise linguística *WordSmith Tools 5.0*.

3.5.1 Extração das Legendas

Primeiramente as legendas foram extraídas via OCR (*Optical Character Recognition*) com auxílio do programa *SubRip*, para o formato “.srt”. Dentre inúmeros formatos existentes de legenda o “.srt” é o que apresenta, ao meu ver, a melhor estrutura textual, no sentido de reunir todas as informações típicas de uma legenda (posição da legenda no vídeo, tempos de entrada e saída de cada legenda, e legenda propriamente dita) de modo claro e sucinto, permitindo rápida identificação de todas as informações da legenda quando aberto com o *Bloco de Notas* do *Windows*, editor e visualizador de textos sem formatação. Logo abaixo uma figura ilustra o arquivo “.srt” aberto com o *Bloco de Notas*.

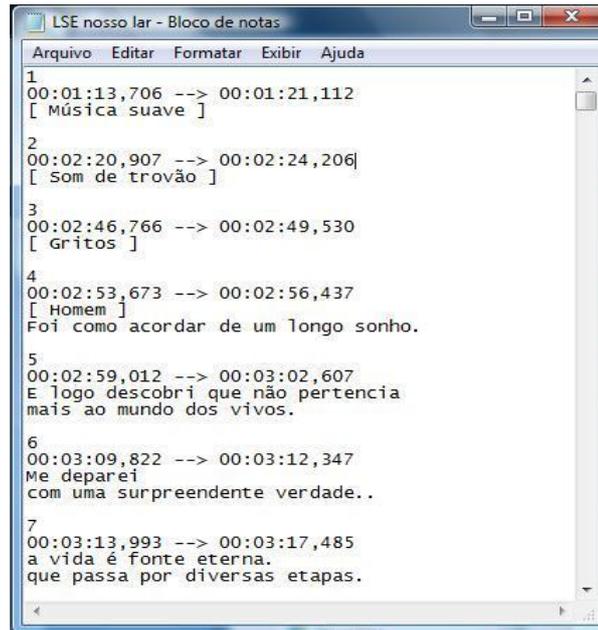


Figura 4: Arquivo de legenda em formato “.srt” aberto com Bloco de Notas.

Vale informar que os programas de reconhecimento de caracteres são passíveis de erros do tipo indistinção entre as letras “l” maiúsculo e “L” minúsculo, entre as letras “ri” e “n”, portanto faz-se necessário uma revisão das legendas extraídas, a fim de consertar esses erros. A partir dos arquivos de legendas em “.srt” é possível fazer a anotação do *corpus*, procedimento explicado mais adiante. Abaixo uma figura ilustra a interface do programa *SubRip*.

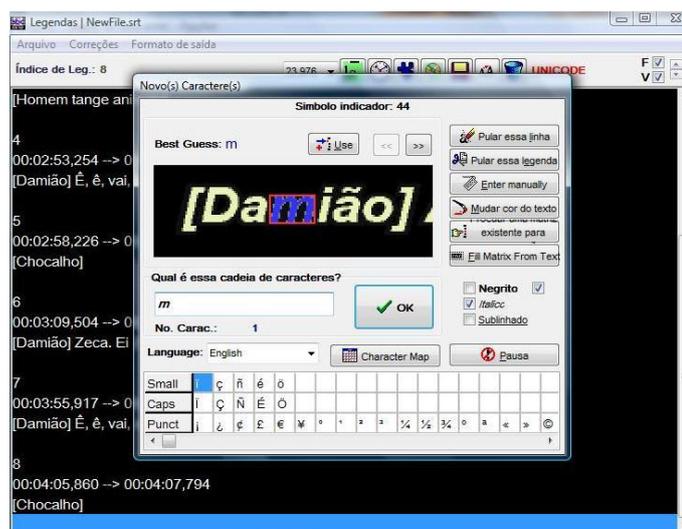


Figura 5: Interface do programa *SubRip*

3.6 Anotação do *corpus*

Após a extração das legendas foi feito o processo de anotação, considerado por O'Donnell (1999) a análise do *corpus*. A análise ou categorização do *corpus* é um procedimento típico da metodologia baseada em *corpus* que contempla duas fases de análise, a do *corpus* propriamente dita e a dos dados ou resultados da pesquisa. Ainda sobre anotação, Leech (2005) esclarece que é a prática de adicionar informações linguísticas interpretativas ao *corpus*.¹⁹ De acordo com McEnery, et al. (2006) essas informações adicionais podem ocorrer num nível meta linguístico, como por exemplo, informações sobre autor, público leitor ou data de publicação para compor um cabeçalho, como também podem codificar uma análise de aspectos nos níveis discursivo, semântico, sintático, lexical, morfológico ou fonético²⁰.

A anotação do presente *corpus* tomou como base categorias linguísticas propostas por Perego (2008), mostradas adiante, e a partir dessas categorias foi possível definir um conjunto de etiquetas para analisar a segmentação em legendas para surdos e ensurdecidos.

A etiquetagem, processo autoral deste trabalho, é um dos procedimentos mais importantes desta pesquisa, pois faz parte do processo de anotação do *corpus* que serviu para a análise da segmentação na LSE. Segundo Berber Sardinha (2004, p. 145) a etiquetagem auxilia na desambiguação lexical e permite a descrição de padrões léxico-gramaticais. Em outras palavras a etiquetagem consiste numa codificação do *corpus* passível ou não de análise eletrônica através de um concordanciador. Para esta pesquisa o concordanciador utilizado foi o *Concord* do conjunto de ferramentas do *WordSmith Tools 5.0*.

Tomando como base o glossário de Linguística de *Corpus* elaborado por Tagnin (2010) os níveis de etiquetagem linguística são definidos como:

- Etiquetagem morfossintática (em inglês *pos-tagging*): processo pelo qual o conteúdo do *corpus* recebe etiquetas morfológicas, isto é, de categorias gramaticais (por exemplo, *substantivo, adjetivo, verbo* etc.).

¹⁹ Minha tradução de: *the practice of adding interpretative linguistic information to a corpus*.

²⁰ Minha tradução de: *Annotation can also occur at the meta-linguistic level – for example, by adding information such as author, level of readership or date of publication to a text's header – or it can encode an analysis of some feature at the discourse, semantic, grammatical, lexical, morphological or phonetic level.*

- Etiquetação sintática: processo pelo qual o conteúdo do *corpus* recebe etiquetas sintáticas (por exemplo, *sujeito, predicado, objeto direto* etc.).
- Etiquetação semântica: processo pelo qual o conteúdo do *corpus* recebe etiquetas semânticas (por exemplo, *cor, roupa, tempero, utensílio* etc.).
- Etiquetação discursiva: processo pelo qual o conteúdo do *corpus* recebe etiquetas que demarcam as partes de um texto (por exemplo, *ingredientes, modo de fazer, resumo, introdução, materiais e métodos* etc.)

As etiquetas podem ser automáticas, sem intervenção do pesquisador, feitas por programas etiquetadores que já compreendem um conjunto de etiquetas predefinidas para os mais variados tipos de análises, ou podem ser manuais, com intervenção e de autoria do pesquisador. O nível de etiquetação condizente com esta pesquisa foi ora o discursivo, ora o sintático. A etiquetação foi feita manualmente, sem auxílio de etiquetadores, já que se trata de um modelo personalizado que pretende servir para ampliar a investigação da segmentação na LSE. A anotação manual deve ser feita com auxílio de editores de texto e salva em formato “.txt” (texto sem formatação), podendo desta forma ser compreendida pelos programas de análise linguística, como o escolhido para esta pesquisa, *Wordsmith Tools 5.0*. Como o *corpus* desta pesquisa são legendas em formato “.srt” tive que abrir os arquivos “.srt” com o editor de textos *Word* para começar a etiquetação, neste caso utilizei o editor de textos *Word* por considerá-lo de fácil e prático manuseio, mas poderia ter optado pelo *Bloco de Notas* ou qualquer outro editor de textos.

As etiquetas foram criadas a partir das categorias de análise da má segmentação linguística de legendas, desenvolvidas por Perego (2008), que analisou a má segmentação, nesse caso a “quebra de linhas” (*linebreaks*), num *corpus* de legendas em italiano e em húngaro de quinze filmes de gêneros diversos para os suportes DVD e cinema. As categorias de má segmentação encontradas nas legendas dos quinze filmes foram as seguintes: quebra da estrutura interna do Sintagma Nominal, do Sintagma Preposicionado, do Sintagma Verbal, da Oração Coordenada e da Oração Subordinada. Perego (2008) observou que os casos de má segmentação nas legendas aconteciam devido à quebra das estruturas internas dos sintagmas ou das orações e isso comprometia o entendimento do filme.

Perego (2008) encontrou casos em que a quebra das estruturas linguísticas acontecia nas situações, a saber:

No Sintagma Nominal

- Pre-modificadores (determinantes) + substantivo,
ex: *a / four-hour documentary, some / life*
- Nomes próprios,
ex: *the "Susan / Marie"*
- Título+nome próprio,
ex: *Miss / Sherwood*
- Sequência substantivo + substantivo (in English),
ex: *construction / work*
- Nominais:
 - 1- Formas cristalizadas (ex: *credit / card*),
 - 2- Colocações (ex: *animal / urges*),
 - 3- Expressões idiomáticas (ex: *dato / i numeri 'go mad'*)

Vale explicar que o sinal de mais (+) representa a quebra indevida da estrutura linguística e a barra (/) indica onde se deu a quebra da linha de legenda e conseqüentemente da estrutura linguística.

No Sintagma Preposicionado

- Preposição + substantivo,
ex: *della / barca, from / France*
- Estruturas ambíguas,
ex: *The night before / he left* when intended meaning is [The night] [before he left]

No Sintagma Verbal

- Verbo principal+auxiliar, exceto construções menos deflexivas
ex: *we could/talk e I just want/to talk*
- Padrões verbais fixos como a estrutura causativa
ex: *fare/infinitive*

- Combinações específicas verbo-partícula (phrasal verbs)

ex: *get along/with*

- Construções com marcadores de infinitivo

ex: *try to/figure out*

Na Oração Coordenada

- Conjunção coordenada + oração

ex: *Come puoi fare la marmellata e/fingere che non sia successo niente?*

- Construções negativas

ex: *If the boys are half English it does not / mean they're bad Pakistanis.*

Na Oração Subordinada

- Conjunção subordinada + oração

ex: *I felt that/you just wanted*

Exceção: o grau de aceitação aumenta se parte da oração subordinada estiver na parte de cima da legenda²¹

- Se

ex: *If / you agree*

- Construções comparativas

ex: *similar to the / wife*

A partir das categorias propostas por Perego (2008) foi possível analisar as legendas do filme *Nosso Lar* e propor etiquetas de análise de segmentação linguística, pautada pela sintaxe, tendo como base o presente *corpus*. As etiquetas foram direcionadas à análise do português brasileiro e para tanto utilizei a tipologia presente na Gramática Descritiva do Português Brasileiro (PERINI, 2010) que tem por objetivo descrever o português brasileiro no contexto falado informal com ênfase na sintaxe e na semântica da oração.

Nem todas as categorias encontradas em Perego (2008) foram encontradas no presente *corpus* e vice versa. Em resumo: 18 categorias de problema de segmentação gramatical foram encontradas em Perego (2008) e 19 categorias de problema de segmentação gramatical no

²¹ Minha tradução de: *the degree of acceptability increases if also part of the subordinate clause is retained in the upper line.*

presente *corpus*, entretanto, apenas 10 delas foram encontradas em ambos os *corpora*. O fato de apenas 10 das 18 categorias definidas por Perego (2008) serem comum a ambos os *corpora* é justificado pela diferença na estrutura linguística dos idiomas analisados, a saber, construções específicas da língua inglesa como *phrasal verbs*, (constituídos de verbo + partícula) não são encontradas na língua portuguesa. Além das 19 etiquetas referentes às categorias de problemas de segmentação gramatical encontradas no presente *corpus* foram definidas mais 11 etiquetas: 4 indicativas de problemas de segmentação (gramatical, retórica e visual) e 7 para análise de parâmetros técnicos da LSE, como velocidade da legenda, que serão mostradas adiante.

Antes de elencar as etiquetas propostas, algumas informações são importantes: todos os exemplos foram extraídos do *corpus*, contudo, aqueles cuja categoria não teve ocorrência no *corpus* foram extraídos de outras fontes. Os recursos gráficos utilizados na análise do *corpus* significam o mesmo ao longo de todo o trabalho. Na Linguística de *Corpus* os parênteses angulares, com ou sem barra, caracterizam as etiquetas e delimitam o conteúdo analisado. Os parênteses angulares vazios <> representam o início daquilo que se pretende marcar e os parênteses angulares com uma barra </> representam o fim daquilo que se pretende marcar²², o sinal de mais (+) representa a quebra indevida da estrutura linguística, a barra (/) indica aonde se deu a quebra da linha na legenda e conseqüentemente da estrutura linguística e o sublinhado corresponde ao bloco de palavras que deveria estar junto na mesma linha de legenda. Abaixo estão todas as etiquetas do *corpus* da presente pesquisa.

ETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA NOMINAL (SN)

- Quando há quebra entre pre-nucleares e substantivo <SN_pre-nucleares+subst>
ex: <sub142>Pelo menos tira aquela/sensação de fome.
- Quando há quebra entre nominal e modificador, ou na ordem inversa, modificador e nominal <SN_nominal+modif/modif+nominal>
ex: <sub227>Seu aparelho/gastrointestinal estava
- Quando há quebra entre superlativo e adjetivo <SN_superlativo+adj>
ex: <sub145>Muito mais/bem disposto, pelo visto.

²² O uso de parênteses angulares provém de uma linguagem computacional de marcação de hipertexto para produzir páginas da WEB, conhecida por HTML (*Hyper Text Markup Language*). Programas de análise linguística como o utilizado nesta pesquisa, *WordSmith Tools*, interpretam a linguagem HTML, isso justifica o uso de parênteses angulares na etiquetagem.

- Quando há quebra entre relativo e oração incompleta <SN_relativo+oração incompleta>

ex: <sub288>A vida na Terra é que/é uma cópia daqui, André.

- Quando há quebra entre nome próprio <SN_nome próprio>

ex: Ela vai se chamar Ana/Rosa Belo Fernandes

- Quando há quebra entre título e nome próprio <SN_título+nome próprio>

ex: <sub389>Você pode procurar o irmão/Genésio no Ministério do Auxílio,

- Quando há quebra da estrutura interna de colocações, expressões idiomáticas e convencionais <SN_colocações/idiom/conv>

ex: <sub98>que vinha do fundo/da minha alma, mas fui ouvido.

ETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA PREPOSICIONADO (SP)

- Quando há quebra entre preposição e substantivo <SP_prep+subst>

ex: <sub523>O que sabe sobre/a medicina espiritual?

ETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA VERBAL (SV)

- Quando há quebra entre dois ou mais verbos, quer sejam auxiliares, modais ou principais <SV_verbo+verbo>

ex: <sub197>O amigo parece ter/compreendido o sentido da água,

- Quando há quebra entre verbo e advérbio <SV_verbo+adv>

ex: <sub354>Já me perdi/muito por essas trilhas.

- Quando há quebra entre colocações verbais <SV_colocações>

ex: Vamos tomar/providências sobre o assalto

- Quando há quebra entre partícula de negação (não, nem etc.) e verbo <SV_negação+verbo>

ex: <sub781>E se eu não/quiser entender, vó?

- Quando há quebra entre pronome oblíquo (precedido ou não de verbo) e verbo <SV_(verbo)+oblíquo+verbo>

ex: <sub731>O ministro vai/nos receber em breve.

ETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA ADVERBIAL (SAdv)

- Quando há quebra da estrutura interna de um advérbio <SAdv>

ex: <sub230>Um ato realizado durante/longos e longos anos,

ETIQUETAS DE ANÁLISE DE QUEBRA DE SINTAGMA ADJETIVO (SAdj)

- Quando há quebra entre substantivo e adjetivo <SAdj_subst+adj>

ex: <sub411>com a separação/temporária da morte,

ETIQUETAS DE ANÁLISE DE QUEBRA DE ORAÇÃO COORDENADA (COORD)

- Quando há quebra entre coordenador (*e, mas, logo etc.*) e oração coordenada <COORD_coordenador+oração>

ex: <sub585> vamos entrar e você/fala com o governador.

- Quando há quebra entre uma partícula negativa da oração e o restante da oração <COORD_negativa>

ex: vamos conversar, mas não/quero brigas.

ETIQUETAS DE ANÁLISE DE QUEBRA DE ORAÇÃO SUBORDINADA (SUBORD)

- Quando há quebra entre conjunção subordinada (*quando, enquanto, que, porque etc.*) e oração <SUBORD_conj+oração>

ex: <sub624>Tudo perde o sentido quando/a gente acorda depois de morrer.

- Quando há quebra entre partícula *se* e oração <SUBORD_se>

ex: Vou comprar um carro se/conseguir o novo emprego.

Além das etiquetas de análise de segmentação pautada pela sintaxe foram propostas etiquetas de análise de segmentação pautada pela retórica e pautada pelo corte da cena (segmentação visual) e etiquetas de análise dos parâmetros técnicos da legendagem. Esses resultados serão discutidos nos próximos capítulos.

ETIQUETA INDICATIVA DE PROBLEMAS DE SEGMENTAÇÃO GRAMATICAL

- Quando há problema de segmentação gramatical (linguística) <PROSEGG>

ETIQUETAS INDICATIVAS DE PROBLEMA DE SEGMENTAÇÃO RETÓRICA

- Quando antecipa informações da legenda subsequente
<PROSEGR_antecipouinformação>
- Quando atrasa informações da legenda antecedente
<PROSEGR_atrasouinformação>

ETIQUETA INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO VISUAL

- Quando a legenda ‘vaza’ (continua) na cena subsequente <PROSEGV_vazou>.

Esta categoria, em especial, não foi analisada no *corpus*, pois, como explicado anteriormente, foge aos objetivos deste trabalho, entretanto a etiqueta foi criada porque há interesse por parte da pesquisadora de ampliar a investigação da segmentação na LSE em futuros trabalhos. As etiquetas definidas neste trabalho não servem apenas à análise das legendas para DVD do filme *Nosso Lar*, pelo contrário, essas etiquetas foram pensadas para que essa metodologia seja facilmente replicada por outros pesquisadores com o intuito de analisar legendas em geral, fortalecendo, dessa forma, os estudos de legendagem baseados em *corpora*.

ETIQUETA DE IDENTIFICAÇÃO DA LEGENDA

- A etiqueta de identificação da legenda ajuda a encontrar uma legenda específica no *corpus*, pela sua numeração, e indica onde começa e termina determinada legenda
<sub1>legenda1</sub1>

A palavra “sub” é a abreviação da palavra *subtitle* (legenda, em português). Optei pelo uso de ‘sub’ pelo fato de ser um termo reconhecido internacionalmente.

ETIQUETA INDICATIVA DE NÚMEROS DE LINHAS POR LEGENDA

- Informa se a legenda contém uma, duas ou mais linhas <1L>, <2L>

Isso ajuda, dentre outras coisas, a identificar se as legendas de uma linha são diferentes das legendas de duas linhas, e se sim, o que caracteriza cada uma.

ETIQUETA IDICATIVA DE TEMPOS INICIAL E FINAL

- Indica o tempo inicial pelo parêntese angular vazio e o tempo final de cada legenda pelo parêntese angular com barra <t>início --> final</t>

Essa marcação ajuda a identificar rapidamente a legenda em seu contexto e posição exata no vídeo.

ETIQUETA DE NÚMERO DE CARACTERES POR LINHA DE LEGENDA

- Apresenta a quantidade de caracteres de cada linha de legenda e só é utilizada em legendas de duas linhas <cpl>

Saber a quantidade de caracteres por linha de uma legenda ajuda a refletir se a má segmentação pode ter sido decorrente da restrição de no máximo 35 caracteres por linha, ou se a má segmentação era inconsciente ou mesmo, se era decorrente de uma decisão pessoal do tradutor.

ETIQUETAS DE VELOCIDADE DA LEGENDA

- <velocidade da legenda_baixa>

- <velocidade da legenda_ideal>

- <velocidade da legenda_alta>

Como mencionado no capítulo 2 as velocidades podem ser classificadas, para fins desta pesquisa, como baixa (até 13cps), ideal (entre 14 e 15cps) e alta (a partir de 16cps). Nas etiquetas, logo após a classificação baixa, ideal ou alta, é necessário apresentar a quantidade total de caracteres seguida do tempo de exibição total da legenda analisada. Ex: <velocidade da legenda_ideal 39c/2,8s>. Abaixo, é possível visualizar um quadro sinótico com as 30 etiquetas definidas.²³

²³ Este conjunto de etiquetas também pode ser visualizado no *Apêndice A*, na seção de Apêndices.

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>	
ETIQUETAS 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>	

Quadro 1: Quadro sinótico das etiquetas propostas

Em seguida, um trecho do *corpus* sem etiqueta e com etiqueta ilustra o processo de etiquetagem:

TRECHO SEM ETIQUETAGEM	TRECHO COM ETIQUETAGEM
1 00:01:13,706 --> 00:01:21,112 [Música suave]	<sub1><1L>1 <t>00:01:13,706 --> 00:01:21,112</t> [Música suave] <velocidade da legenda_baixa 16c/7,4s></sub1>
2 00:02:20,907 --> 00:02:24,206 [Som de trovão]	<sub2><1L>2 <t>00:02:20,907 --> 00:02:24,206</t> [Som de trovão] <velocidade da legenda_baixa 17c/3,3s></sub2>
3 00:02:46,766 --> 00:02:49,530 [Gritos]	<sub3><1L>3 <t>00:02:46,766 --> 00:02:49,530</t> [Gritos] <velocidade da legenda_baixa 10c/2,7s></sub3>
4 00:02:53,673 --> 00:02:56,437 [Homem] Foi como acordar de um longo sonho.	<sub4><2L>4 <t>00:02:53,673 --> 00:02:56,437</t> <cpl9>[Homem] <cpl35>Foi como acordar de um longo sonho. <velocidade da legenda_alta 44c/2,7s></sub4>
5 00:02:59,012 --> 00:03:02,607 E logo descobri que não pertencia mais ao mundo dos vivos.	<sub5><2L>5 <t>00:02:59,012 --> 00:03:02,607</t> <cpl33>E logo descobri que não pertencia<PROSEGG><SV_verbo+adv> <cpl24>mais ao mundo dos vivos. <velocidade da legenda_alta 57c/3,6s></sub5>

Quadro 2: Exemplo de trecho sem etiquetas e com etiquetas do filme *Nosso Lar*

Após o processo de anotação do *corpus* foi possível realizar a análise eletrônica com auxílio da *Word List* e da *Concord*. Além da análise eletrônica foi realizada uma análise qualitativa a partir de observação e interpretação dos dados. A análise está detalhada no próximo capítulo.

3.7 Análise dos Dados

A análise dos dados foi realizada de forma eletrônica e de forma interpretativa. Essa análise eletrônica foi viabilizada por duas das três ferramentas do programa *Wordsmith Tools 5.0*, o gerador de lista de palavras *Word List*, que fornece uma lista de todas as palavras do *corpus* por ordem de frequência, ordem alfabética e por estatística e o concordanciador *Concord*, definido no *Glossário de Linguística de Corpus* (TAGNIN, 2010) como programa que extrai todas as ocorrências de uma palavra de busca num *corpus* juntamente com seu

contexto, apresentando-as na forma de uma concordância, em posição central, junta com seu contexto. Para fins desta análise, não foi necessário utilizar o extrator de palavra chave, *KeyWord*, pois exige um *corpus* de estudo e outro de referência para extrair as palavras chave ou termos com frequência estatisticamente significativa no *corpus* de estudo em relação ao *corpus* de referência, portanto a busca de palavras chave foge aos objetivos desta pesquisa.

A *WordList* viabilizou a análise da ocorrência de todas as palavras do *corpus* por uma lista em ordem de frequência, alfabética e estatística. Para utilizar a *Word List*, basta proceder conforme mostrado do passo 1 ao 8:

1) Abrir o programa *WordSmith Tools 5.0* e ajustar as configurações do menu *Tags* clicando na opção *Adjust Settings* do menu *Settings* que aparece na tela inicial do programa.

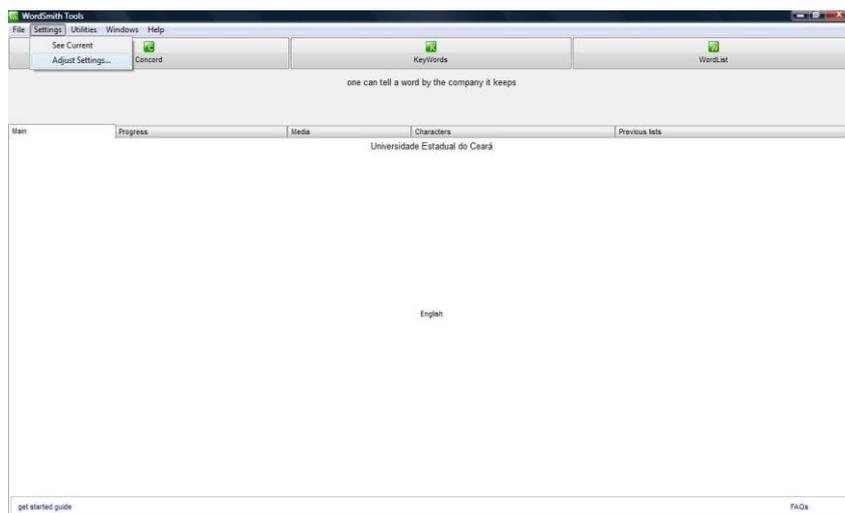


Figura 6: Interface do *WordSmith Tools 5.0 - Settings*

2) Clicar no menu *Tags* e verificar se a opção *Mark-up to ignore* (ignorar marcação, em português) está preenchida com um asterisco entre parênteses angulares <*>, que significa que toda e qualquer etiqueta que estiver entre parênteses angulares será ignorada na análise, por isso é fundamental verificar sempre esta opção. Nesse caso, os parênteses angulares devem permanecer para que as etiquetas sejam ignoradas na análise, pois a inclusão das etiquetas aumentaria o número de *tokens* do *corpus* e daria diferença nos resultados. Depois disso, basta marcar a opção *save* e clicar em *ok*.

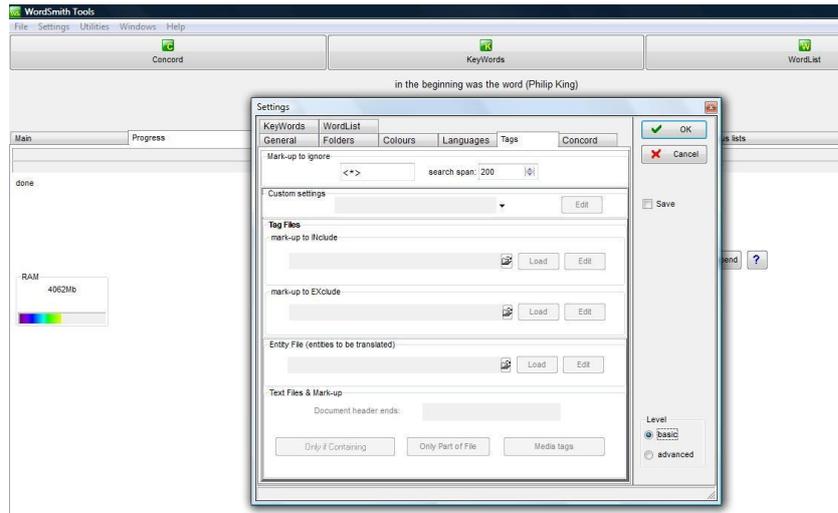


Figura 7: Interface do *WordSmith Tools 5.0 – Tags*

- 3) Retornar à tela principal do programa e escolher a ferramenta *Word List*. Nela é só abrir um novo arquivo para escolher o *corpus* clicando na opção *New* do menu *File*

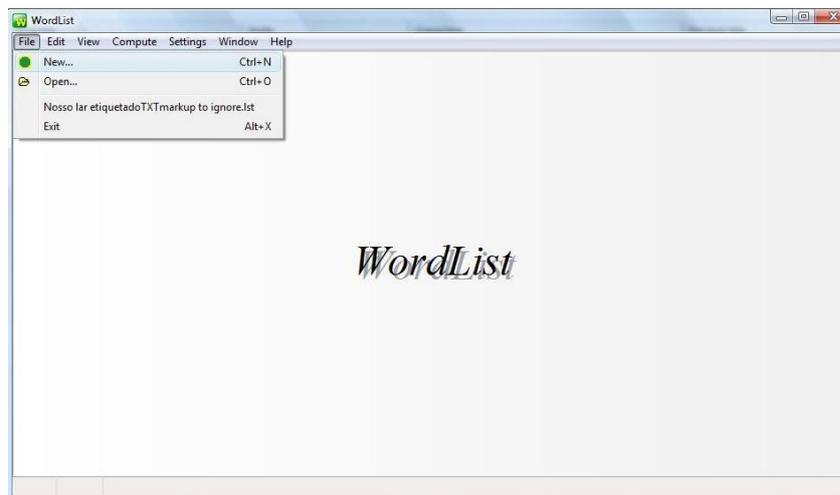


Figura 8: Interface do *WordSmith Tools 5.0 – Word List*

- 4) Para escolher os textos do *corpus* basta clicar em *Choose Texts Now* no menu *Main*.

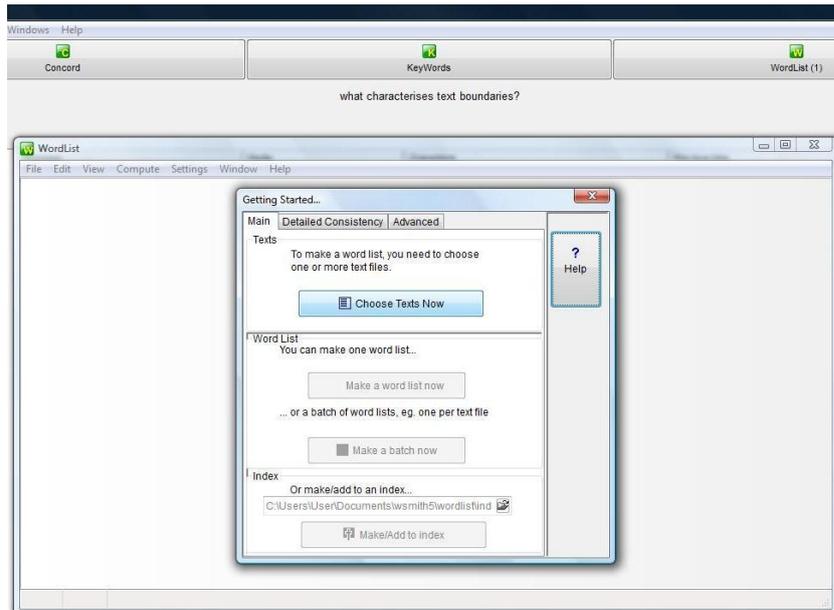


Figura 9: Interface do *WordSmith Tools 5.0 – Word List-Choose Texts Now*

5) Na janela *Choose Texts*, os textos aparecem do lado esquerdo e precisam ser arrastados para o lado direito da caixa, quando isso for feito basta clicar em *ok* para prosseguir com a análise

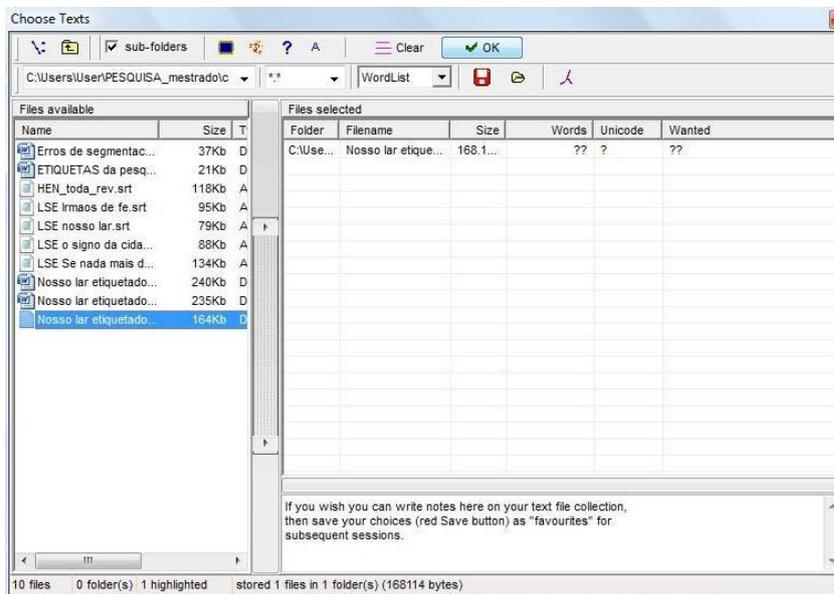


Figura 10: Interface do *WordSmith Tools 5.0 – Choose Texts*

6) Ao clicar na opção *Make a Word List Now* uma lista de palavras aparece podendo ser visualizada de três formas, a saber: por ordem de frequência, alfabética e por estatística. A forma que melhor reúne as informações do *corpus* é a estatística.



Figura 11: Interface do *WordSmith Tools 5.0* – *Make a Word List Now*

- 7) As formas podem ser selecionadas nas abas no canto inferior esquerdo da tela.

WordList		
File	Edit	View
Compute	Settings	Window
Help		
N	Overall	1
text file	Overall	TXT.txt
file size	68,114	68,114
tokens (running words) in text	16,313	16,313
tokens used for word list	6,122	6,122
sum of entries	0	0
types (distinct words)	1,587	1,587
type/token ratio (TTR)	25.92	25.92
standardised TTR	22.06	22.06
standardised TTR std.dev.	71.57	71.57
standardised TTR basis	1,000	1,000
mean word length (in characters)	3.18	3.18
word length std.dev.	1.89	1.89
sentences	7,042	924
mean (in words)	6.63	6.63
std.dev.	1.79	4.95
paragraphs	4,426	3
mean (in words)	211.38	040.67
std.dev.	27.66	284.06
headings	0	0
mean (in words)		
std.dev.		
sections	1	1
mean (in words)	122.00	122.00
std.dev.		
numbers removed	10,191	10,191
stoplist tokens removed	0	0
stoplist types removed	0	0
1-letter words	474	474
2-letter words	7,685	7,685
3-letter words	4,304	4,304
4-letter words	915	915
5-letter words	1,059	1,059
6-letter words	647	647
frequency	alphabetical	statistics
filenames	notes	

Figura 12: Interface do *WordSmith Tools 5.0* – Listas de Estatística

8) Para encerrar a análise do *corpus* pela *Word List* é só salvar os arquivos em formato padrão *Statistics List*, que poderão ser visualizados sempre que necessário pelo *WordSmith Tools 5.0*.

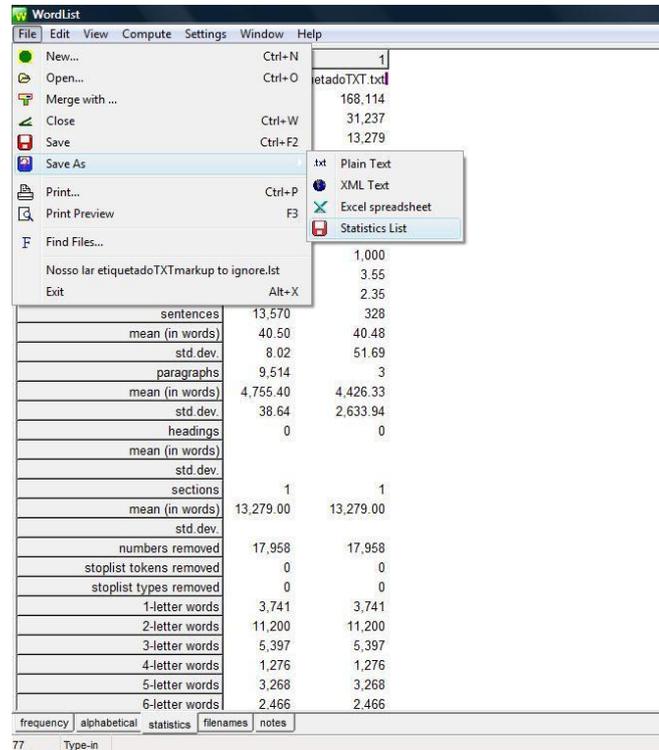


Figura 13: Interface do *WordSmith Tools 5.0* – Salvar *WordList*

Antes de partir para a análise do *corpus*, em concordância, é interessante conhecer primeiro o funcionamento da ferramenta *Concord*. A *Concord* oferece uma opção de busca de palavras chamada *search word* que ajuda a visualizar a palavra de busca em contexto, destacada no centro da linha. A partir da observação da palavra de busca em contexto foi possível fazer interpretações sobre os padrões encontrados. Além disso, foi possível contabilizar todas as categorias (etiquetas) recorrentes no *corpus*. Por exemplo, para saber a quantidade de erros de segmentação gramatical basta escrever no campo de busca a etiqueta PROSEGG, sem os parênteses angulares, e automaticamente o programa disponibilizará uma lista com todas as ocorrências de problema de segmentação gramatical dentro da mesma legenda.

Para realizar a análise eletrônica do *corpus* através da *Concord* foi necessário seguir os passos detalhados do 9 ao 16:

9) Abrir o programa *WordSmith Tools 5.0* e ajustar as configurações do menu *Tags* conforme mencionado no passo 1 de utilização da *Word List*.

10) Clicar no menu *Tags* e certificar que a opção *Mark-up to ignore* (ignorar marcação, em português) está em branco. Caso esteja preenchida com um asterisco entre parênteses angulares <*> significa que toda e qualquer etiqueta que estiver entre parênteses angulares será ignorada na análise, por isso é fundamental deixar essa opção em branco, pois o que se pretende analisar nesse trabalho são as etiquetas. Depois disso basta marcar a opção *save* e clicar em *ok*.

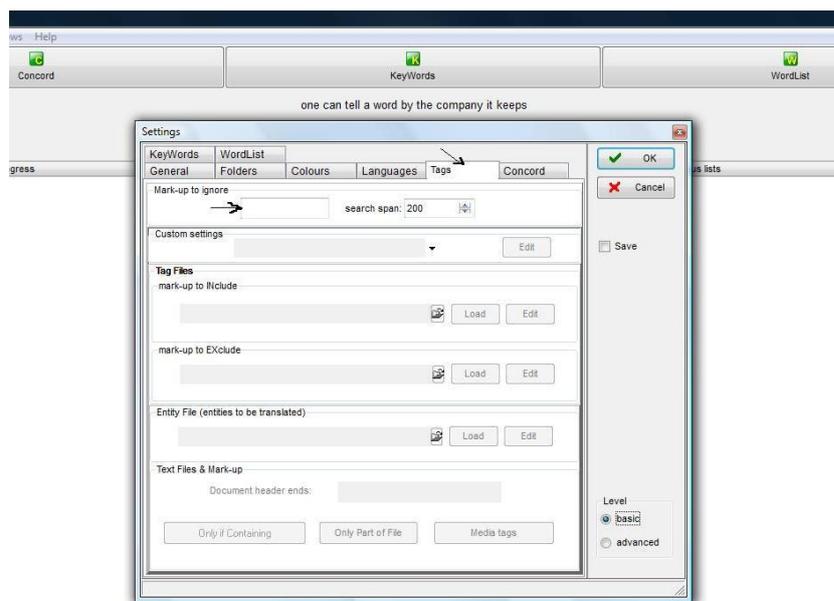


Figura 14: Interface do *WordSmith Tools 5.0* – Menu *Tags*-*Mark up to ignore*

11) Retornar à tela principal do programa e escolher a ferramenta *Concord*. Nela, é só abrir um novo arquivo para escolher o *corpus* clicando na opção *New* do menu *File*.

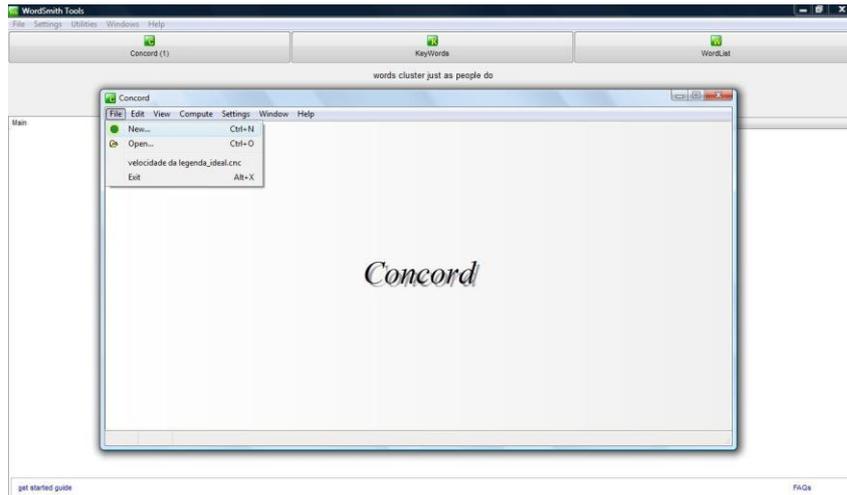


Figura 15: Interface do *WordSmith Tools 5.0 – Concord*

- 12) Para escolher os textos do *corpus* basta clicar em *Choose Texts Now* no menu *Texts*.

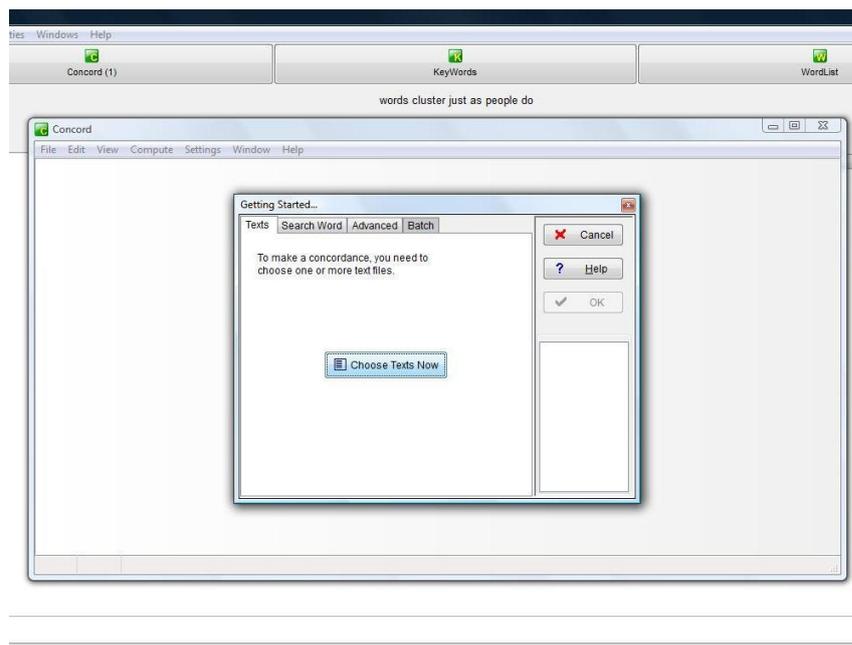


Figura 16: Interface do *WordSmith Tools 5.0 – Concord-Choose Texts Now*

- 13) Seguir o passo 6 para seleccionar o texto do *corpus*
- 14) Uma caixa de diálogos se abrirá mostrando a lacuna *Search Word* a ser preenchida com a palavra ou etiqueta de busca.

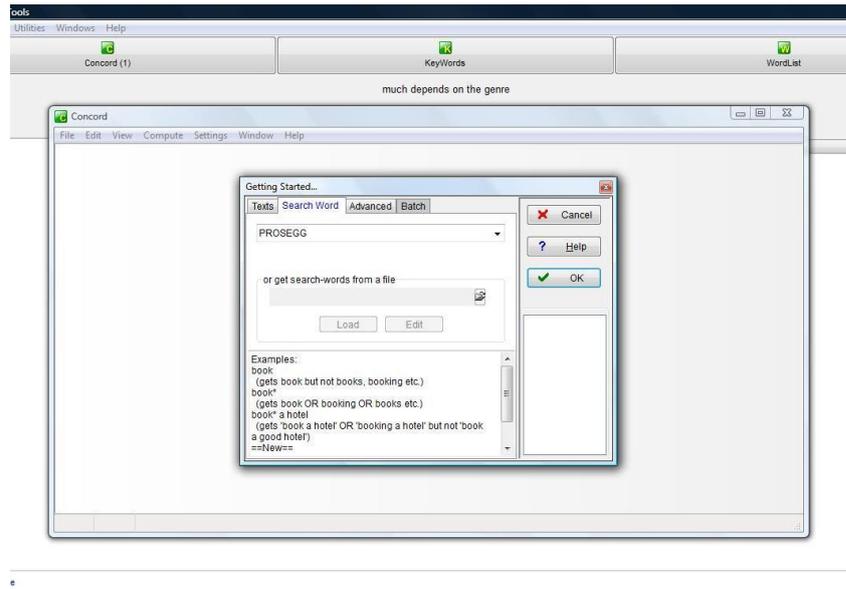


Figura 17: Interface do *WordSmith Tools 5.0* – *Concord-Search Word*

- 15) Depois de preencher o campo de busca com a etiqueta certa, é só clicar *ok* e aguardar a lista.

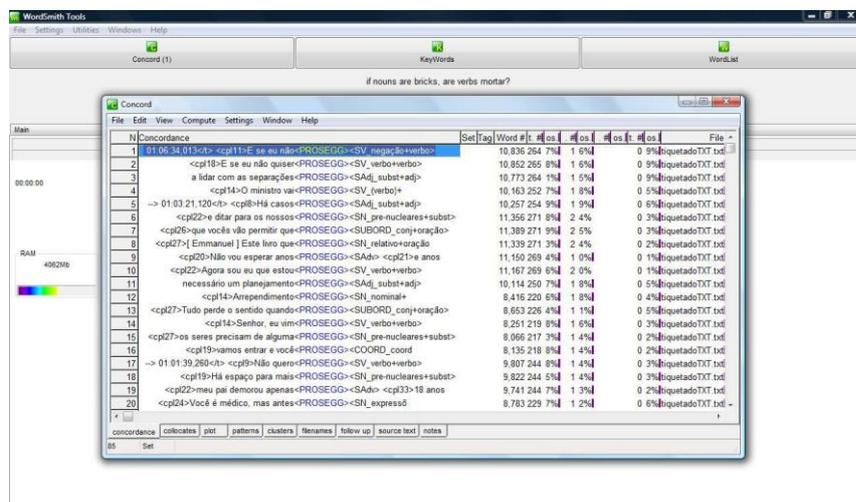


Figura 18: Interface do *WordSmith Tools 5.0* – Palavra de busca em concordância

- 16) Por último, salvar o arquivo gerado no formato padrão da *Concord*, *Concordance List*, e ele poderá ser visualizado pelo *WordSmith Tools* sempre que necessário.

4 ANÁLISE E DISCUSSÃO DOS DADOS

Este capítulo expõe a análise dos dados e uma discussão acerca dos resultados encontrados. A análise dos parâmetros técnicos e linguísticos da LSE do filme *Nosso Lar* foi feita sob uma ótica descritiva, de modo a explicar cada um dos parâmetros, em especial o da segmentação, enfatizando cada um dos problemas encontrados no *corpus*. Os resultados encontrados nos parâmetros técnicos, como número de caracteres por linha de legenda, velocidade de leitura etc. foram relacionados aos resultados encontrados sobre os parâmetros linguísticos com fins de mostrar como se deu a segmentação das legendas. O viés qualitativo, ou interpretativo, esteve presente em todos os momentos da análise, visto que dados eletrônicos fornecidos por computador não significam nada sem a interpretação do pesquisador.

4.1 Dados gerados pela *Word List*

Antes de detalhar os parâmetros técnicos e linguísticos analisados é importante apresentar os resultados estatísticos adquiridos através do gerador de lista de palavras *Word List*.

	N	1
text file	LSE	Nosso Lar.txt
tokens (running words) in text		16,313
tokens used for word list		6,122
types (distinct words)		1,587
type/token ratio (TTR)		25.92
numbers removed		10,191

Figura 20: Resultado estatístico gerado pela *Word List*

Nem todos os dados gerados pela *Word List* foram utilizados na análise dos dados. Os dados interessantes para esta pesquisa foram extraídos somente da lista estatística, ficando de fora as listas alfabética e de frequência. Na primeira linha, a letra *N* significa o número, ou a ordem dos textos utilizados na *Word List*, a segunda linha *text file* mostra o arquivo de texto analisado (*LSE Nosso Lar*), a terceira linha *tokens (running words) in text* apresenta um total de 16.313 itens linguísticos no texto analisado, que inclui os numerais contidos no arquivo srt (número da legenda e tempos inicial e final) como sendo itens linguísticos, mas, na verdade,

esses números não foram válidos para a análise. A quarta linha *tokens used for word list*, exclui os numerais e apresenta um total de 6.122 itens linguísticos, válidos para a análise. A quinta linha *types (distinct words)* apresentou um total de 1.587 palavras distintas no *corpus*, a sexta *type/token ratio* apresentou um percentual de 25,92% da razão entre as palavras distintas (*types*) e o total de palavras (*tokens*) no *corpus*. A última linha, *numbers removed*, apresenta um total de 10.191 números removidos do *corpus*, que corresponde exatamente à diferença entre o total de palavras com e sem números.

4.2 Parâmetros Técnicos da LSE do filme *Nosso Lar*

Para analisar os parâmetros técnicos foram criadas etiquetas personalizadas de codificação do *corpus*, algumas, pouco interessantes para os resultados, como as etiquetas indicativas de número da legenda e de tempos inicial e final que servem apenas de orientação para o pesquisador porque indicam a posição da legenda no texto, ou seja, facilitam a localização das legendas em contexto, portanto, ficam de fora da análise. Os demais parâmetros, linhas por legenda, número de caracteres por linha e velocidade da legenda, são importantes para a análise, por isso foram explanados.

4.2.1 Número de Linhas por Legenda

Etiquetas criadas: <1L>, <2L>

No *corpus* foram encontradas legendas de uma e duas linhas, das quais, 574, ou 50,7% foram de uma linha e 558, ou 49,3% foram de duas linhas. Todas as ocorrências de problema de segmentação aconteceram dentro de legendas de duas linhas e nenhum problema de segmentação aconteceu entre legendas de apenas uma linha.

570	28c/2,1s></sub412> <sub413><1L>413 <t>00:41:23,981 -->	5,754 147 0%	1 6%	0 7%	tiquetadoTXT.txt	36%
571	16c/2,6s></sub401> <sub412><1L>402 <t>00:40:13,077 -->	5,604 144 6%	1 5%	0 6%	tiquetadoTXT.txt	35%
572	12c/1,3s></sub400> <sub401><1L>401 <t>00:40:10,408 -->	5,592 144 6%	1 5%	0 6%	tiquetadoTXT.txt	35%
573	11c/1,9s></sub403> <sub414><1L>404 <t>00:40:52,883 -->	5,626 145 5%	1 5%	0 6%	tiquetadoTXT.txt	36%
574	19c/3,4s></sub402> <sub403><1L>403 <t>00:40:40,304 -->	5,616 144 6%	1 5%	0 6%	tiquetadoTXT.txt	36%

concordance	collocates	plot	patterns	clusters	filenames	follow up	source text	notes
574	Set	médico espiritual <velocidade da legenda_baixa 62c/4,7s></sub433> <sub434><1L>434 <t>00:42:39,023 --> 00:42:40,923</t> O que eu posso fazer? <velocidade						

Figura 21: Tela da *Concord* mostrando resultado de 574 legendas de 1 linha

554	31c/2,5s></sub363> <sub364><2L>364 <t>00:37:25,977 -->	5,063	137	0%	1	8%	0	2%	tiquetadoTXT.txt	32%
555	35c/2,7s></sub354> <sub355><2L>355 <t>00:36:28,052 -->	4,935	131	4%	1	7%	0	2%	tiquetadoTXT.txt	31%
556	15c/1,8s></sub353> <sub354><2L>354 <t>00:36:23,881 -->	4,920	131	3%	1	7%	0	1%	tiquetadoTXT.txt	31%
557	18c/2,5s></sub357> <sub358><2L>358 <t>00:36:40,464 -->	4,979	134	2%	1	7%	0	2%	tiquetadoTXT.txt	32%
558	34c/2,2s></sub355> <sub356><2L>356 <t>00:36:30,354 -->	4,951	131	7%	1	7%	0	2%	tiquetadoTXT.txt	31%

concordance collocates plot patterns clusters filenames follow up source text notes

558 Set

Figura 22: Tela da *Concord* mostrando resultado de 558 legendas de 2 linhas

Como dito no capítulo 2, na literatura encontram-se opiniões divergentes sobre o processamento de legendas de uma ou duas linhas, enquanto alguns pesquisadores defendem que legendas de uma linha são melhores de processar, outros defendem as de duas linhas como ideais no processamento.

Neste *corpus*, ambas apareceram praticamente com a mesma frequência, porém, todos os casos de problemas de segmentação ocorreram nas legendas de duas linhas. Para esta pesquisa a preferência entre legendas de uma ou duas linhas não foi verificada, pois seriam necessários estudos direcionados ao processamento de legendas e esse não foi o foco deste trabalho. Uma hipótese que levanto sobre essa questão é a de que ambas as legendas de uma ou duas linhas podem ser bem compreendidas se forem segmentadas no mais alto nível sintático possível. Quanto mais completa a informação contida na legenda e quanto menos ocorrências de quebra dentro das estruturas linguísticas (sintagmas e orações), menor será o esforço para o processamento e maior será a compreensão das legendas. Em outras palavras, a segmentação, ou distribuição do texto, poderia exercer maior influência sobre a compreensão das legendas do que a quantidade de texto processado.

4.2.2 Número de Caracteres por Linha

Etiqueta criada <cpl>

Como o número máximo de caracteres por linha, tido como referência para esta pesquisa foi 35, foram analisadas eletronicamente apenas as ocorrências de linhas com mais 35 caracteres. O programa mostrou que há apenas 19 ocorrências com valores a partir de 36 caracteres, das quais, 4 são de 36c, 9 são de 37c, 4 são de 38c e 2 são de 39c. Um dos fatores que influenciou esse aumento de caracteres por linha foi a quantidade de espaços desnecessários. Todas as identificações de falante e de efeitos sonoros entre colchetes contém espaços desnecessários, como na legenda 713 mostrada abaixo.

```

<sub713><2L>713
<t>01:01:51,207 --> 01:01:55,109</t>
<cpl36>[ Iolanda ] Feita com muito carinho.
<cpl34>[ Lísias ] E acompanhado por água.
<velocidade da legenda_alta 70c/3,9s></sub713>

```

Quadro 3: Legenda 713 etiquetada - Exemplo de identificação de falante e de efeito sonoro

Outro fator que influenciou o aumento de caracteres nas legendas foi a falta de condensação do texto, como ilustrado na legenda 914. Por exemplo, se na construção: <cpl38>*Foi então que comprovei, mais uma vez*, a expressão *mais uma vez* fosse substituída por *novamente* a construção estaria com exatos 35 caracteres.

```

<sub914><2L>914
<t>01:21:21,076 --> 01:21:25,137</t>
<cpl38>Foi então que comprovei, mais uma vez,
<cpl30>o que o amor é capaz de fazer.
<velocidade da legenda_alta 68c/4s></sub914>

```

Quadro 4: Legenda 914 etiquetada - Exemplo de identificação de falante e de efeito sonoro

O que também pode influenciar no aumento de caracteres é a decisão do tradutor, pois ele tem poder de decisão e pode optar, conscientemente, por aumentar os caracteres de uma linha, em prol da boa segmentação, ou de algum outro parâmetro.

A questão dos caracteres por linha é divergente na literatura, mas no que concerne esse *corpus* foram encontradas linhas de legendas com máximo de 39 caracteres. Apesar de ter como referência um valor máximo de 35 caracteres, não é interessante sacrificar a boa segmentação da legendagem em prol dessa restrição de caracteres. Manipular alguns parâmetros da LSE, como condensação e número de caracteres, em prol da segmentação parece ser uma boa saída para lidar com as restrições da legendagem. São muitas as restrições, entretanto é necessário correlacioná-las com vistas a sempre melhorar a qualidade das legendas.

4.2.3 Velocidade da Legenda

Etiquetas criadas: legendas de até 13 cps <velocidade da legenda_baixa>, legendas com 14 e 15 cps <velocidade da legenda_ideal>, legendas acima de 16 cps <velocidade da legenda_alta>.

Os resultados da análise eletrônica mostraram que 620 legendas, ou 54,7% estão com baixa velocidade, 257 legendas, ou 22,7% estão com velocidade ideal e 255 legendas, ou 22,5%, estão com velocidade alta. A maioria das legendas com velocidade baixa são identificações de efeitos sonoros. Isso leva a pensar que pode haver uma tendência, por parte dos legendistas, de aumentar o tempo de exibição das legendas quando se trata de efeitos sonoros. A velocidade média das legendas com velocidade baixa foi de 11,1cps, valor obtido através da divisão do total de caracteres das 620 legendas pelo tempo total das legendas em segundos. O número de caracteres das legendas bem como o tempo em segundos foram visualizados na tela da *Concord* e obtidos a partir da soma dos totais encontrados em cada legenda, como mostra a figura.

N	Concordance	Set
1	--> 00:01:21,112</t> [Música suave] <velocidade da legenda_baixa 16c/7,4s></sub1>	
2	--> 00:02:24,206</t> [Som de trovão] <velocidade da legenda_baixa 17c/3,3s></sub2>	
3	<t>00:02:46,766 --> 00:02:49,530</t> [Gritos] <velocidade da legenda_baixa 10c/2,7s></sub3>	
4	[Pessoas gritando e gemendo] <velocidade da legenda_baixa 30c/2,2s></sub8>	
5] <cpl31>[Gemidos de dor e sofrimento] <velocidade da legenda_baixa 54c/4,6s></sub12>	
6	--> 00:04:10,711</t> [Mulher] André Luiz! <velocidade da legenda_baixa 22c/1,7s></sub13>	
7	filho! <cpl32>Já cumprimentou o papai, a vovó? <velocidade da legenda_baixa 52c/4,7s></sub14>	
8	--> 00:04:17,420</t> [Música suave] <velocidade da legenda_baixa 16c/1,3s></sub15>	
9	<cpl12>Benção, pai. <cpl9>Oi, vovó. <velocidade da legenda_baixa 21c/3,1s></sub16>	
10	--> 00:04:30,101</t> Menino levado! <velocidade da legenda_baixa 14c/1,4s></sub18>	
11	<t>00:04:33,006 --> 00:04:34,200</t> André! <velocidade da legenda_baixa 6c/1,2s></sub19>	
12	Trovoadas] <cpl23>[Gritos desesperados] <velocidade da legenda_baixa 36c/3,8s></sub20>	
13	dores no corpo <cpl16>e dores na alma. <velocidade da legenda_baixa 37c/3,1s></sub22>	
14	O sonho se transformou em pesadelo. <velocidade da legenda_baixa 35c/2,7s></sub25>	
15	<cpl24>[Risadas e burburinho] <cpl9>[Piano] <velocidade da legenda_baixa 36c/3,6s></sub26>	
16	<cpl11>-Um brinde! <cpl11>-Um brinde. <velocidade da legenda_baixa 22c/1,9s></sub29>	
17	00:05:37,701</t> <cpl10>[Mulher] <cpl6>André! <velocidade da legenda_baixa 16c/1,4s></sub30>	
18	<t>00:05:52,985 --> 00:05:55,783</t> [Silêncio] <velocidade da legenda_baixa 12c/2,8s></sub31>	
19	--> 00:06:23,111</t> [Música de suspense] <velocidade da legenda_baixa 22c/4,7s></sub32>	
20	[Trovões e sons de aves de rapina] <velocidade da legenda_baixa 36c/4,7s></sub33>	
21	--> 00:07:08,082</t> Sim, papai. <velocidade da legenda_baixa 11c/1,5s></sub35>	
22	--> 00:07:10,890</t> Pode ir brincar. <velocidade da legenda_baixa 16c/1,7s></sub36>	
23	--> 00:07:22,705</t> [Música suave] <velocidade da legenda_baixa 16c/2,8s></sub37>	
24	<t>00:07:33,219 --> 00:07:34,186</t> Por vezes, <velocidade da legenda_baixa 10c/1s></sub40>	
25	<cpl35>numa falsa imagem de tranquilidade. <velocidade da legenda_baixa 66c/5s></sub41>	
26	--> 00:07:52,430</t> E nunca é tarde demais. <velocidade da legenda_baixa 23c/2,1s></sub43>	
27	00:07:56,333</t> [Som de gritos desesperados] <velocidade da legenda_baixa 30c/2,6s></sub44>	
28	--> 00:07:59,207</t> [Música de suspense] <velocidade da legenda_baixa 22c/2,8s></sub45>	
29] Sai daqui. <cpl19>Suicida! Sai daqui! <velocidade da legenda_baixa 44c/4,4s></sub46>	
30	--> 00:08:43,451</t> A culpa é sua! <velocidade da legenda_baixa 14c/1,3s></sub47>	
31	--> 00:08:46,617</t> Sai, sai, sai! <velocidade da legenda_baixa 14c/3s></sub48>	
32	--> 00:08:48,827</t> Vai morrer de sede! <velocidade da legenda_baixa 19 c/1,9s></sub49>	
33	Mulher] <cpl26>Sai que essa água é minha! <velocidade da legenda_baixa 36c/2,7s></sub50>	

Figura 23: Tela da *Concord* exibindo o número de caracteres e o tempo de cada legenda

Na figura, a linha destacada na horizontal mostra a etiqueta <velocidade da legenda_baixa 10c/2,7s> que significa que essa legenda foi classificada como sendo de velocidade baixa por apresentar 10c em 2,7s. A coluna destacada na vertical mostra uma sequência de caracteres e de segundos. As somas do total de caracteres e do total de segundos das legendas foram feitas manualmente, uma por uma. Desta forma, apresentam-se para as três velocidades os seguintes resultados:

	Número de legendas no <i>corpus</i>	Número de caracteres	Tempo de exibição (em segundos)	Velocidade média (em cps)
Legendas com velocidade baixa (até 13cps)	620	19438c	1748,2s	11,1cps
Legendas com velocidade ideal (entre 14 e 15cps)	257	8935c	604s	14,7cps
Legendas com velocidade alta (acima de 16cps)	255	15324c	830,7s	18,4cps
Média Global da velocidade das legendas do <i>corpus</i>	1132	43697c	3182,9s	13,7cps

Tabela 4: Velocidade das legendas do filme *Nosso Lar*

A tabela de velocidade das legendas mostra os valores encontrados para obter as médias das velocidades baixa, ideal e alta, além da média global de todas as legendas do *corpus*. A velocidade média das legendas do filme *Nosso Lar* foi de 13,7cps e o mesmo resultado em palavras por minuto foi de 115,5ppm.

Diante dos resultados apresentados pode-se afirmar que as legendas do filme *Nosso Lar* são, em sua maioria, lentas e que comparando a velocidade média dessas legendas, 115,5ppm à velocidade definida pelos surdos cearenses vê-se que a quantidade de palavras por minuto das legendas do filme *Nosso Lar* está abaixo daquela preferida pelos surdos. Por um lado, legendas menores e mais lentas podem garantir a harmonização entre legenda e imagem, por outro lado, legendas menores e mais lentas podem provocar a releitura das legendas. A releitura é um fator negativo, pois demanda mais esforço no processamento da mesma porção de texto podendo dar sensação de cansaço ou fadiga. Karamitroglou (1998) defende que legendas de duas linhas completas devem permanecer na tela por no máximo 6

segundos, pois duração maior pode causar releitura automática da legenda, principalmente quando se trata de um leitor rápido.

Relacionando o parâmetro de velocidade de leitura às legendas com problema de segmentação pode-se apresentar que das 88 legendas com problema de segmentação, 22, ou 25% são de velocidade baixa, 24, ou 27,3% são de velocidade ideal e 42 ou 47,7% são de velocidade alta. De antemão já é possível afirmar que quase metade dos problemas de segmentação ocorre em legendas de velocidade alta. Abaixo, a relação entre velocidade de legenda e problema de segmentação está representada num gráfico.

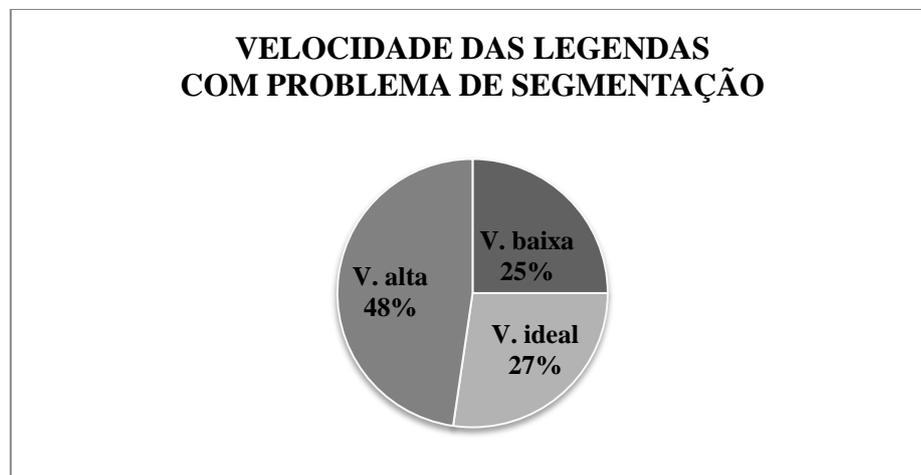


Gráfico 1: Velocidades das legendas com problemas de segmentação

O gráfico pode ser lido da seguinte forma: 48% das legendas com problema de segmentação são de velocidade alta, enquanto 27% são de velocidade ideal e 25% de velocidade baixa.

4.3 Parâmetros Linguísticos da LSE do filme *Nosso Lar*

Os problemas de segmentação linguística e retórica²⁴ encontrados nesse *corpus* são analisados de acordo com a análise linguística proposta na gramática de Perini (2010). Apesar de me valer de uma análise sintática das legendas para descrever a segmentação, essa pesquisa não teve como foco a análise sintática das legendas e sim a análise das legendas como um todo e especialmente da segmentação, que é um parâmetro tipicamente linguístico.

²⁴ As 87 legendas com 88 problemas de segmentação podem ser visualizadas no *Apêndice B* da seção de Apêndices.

Karamitroglou (1998) mostrou que a segmentação gramatical da legenda deve ser feita com base nos mais altos níveis sintáticos. Se não é possível distribuir toda a oração em uma única linha de legenda, a quebra da linha deve ser feita de modo que ambas as linhas contenham um pensamento completo. Por exemplo, se uma oração é composta por SN (substantivos, adjetivos) + SV (verbos auxiliares, principais) + SN + SAdv (advérbio) a divisão dessa oração deve ser feita entre os constituintes imediatos das orações, ou seja, entre os sintagmas e não dentro dos sintagmas.

Sobre a estrutura do português brasileiro, na visão de Perini (2010), pode-se dizer que uma construção linguística se define através de uma sequência de sintagmas, como, SN+V+SN, que representa um tipo de estrutura denominada oração. Uma oração, por sua vez, é uma estrutura que tipicamente contém um verbo e muitas vezes um ou mais complementos (sujeito, objetos e/ou SPs, SAdjs ou SAVs), podendo comportar um ou mais adjuntos e outros elementos. Em resumo, a oração é basicamente composta de um verbo e seus satélites: complementos, adjuntos e alguns outros elementos como a negação verbal e o verbo auxiliar (PERINI, 2010, p. 95).

Para este trabalho, considero cada um desses elementos que compõe a estrutura do português brasileiro como uma unidade linguística máxima que não deveria sofrer quebra na sua estrutura interna. Cada unidade máxima, ou constituinte, encontrada no *corpus* foi analisada e anotada conforme suas respectivas categorias linguísticas, por meio de etiquetas discursivas.

Como explicado anteriormente, no capítulo de metodologia, foram encontradas no *corpus*, constituintes linguísticas que sofreram quebra em sua estrutura interna e foram classificadas através das etiquetas, também apresentadas no capítulo 3. Os dados obtidos através da análise eletrônica serão mostrados neste capítulo de análise e discussão dos dados. Vale lembrar que cada etiqueta de análise dos parâmetros linguísticos indica a estrutura linguística na qual houve problema de segmentação e logo em seguida mostra a estrutura interna quebrada, ou seja, o problema de segmentação. Por exemplo, a etiqueta <SN_pre-nucleares+subst> indica que o problema de segmentação ocorreu no nível do sintagma nominal porque teve a estrutura interna *pre-nucleares + substantivo* quebrada.

Dependendo dos objetivos de cada pesquisa, nem sempre é necessário apresentar todos os dados no capítulo de análise, pois a apresentação de parte dos dados pode ser suficiente para representar o todo e demonstrar os resultados. Porém, para esta pesquisa, que tem objetivo descritivo, julguei interessante apresentar os 88 problemas de segmentação encontrados em 87 legendas porque eles representam um percentual de apenas 7,7% do *corpus*. Os problemas de segmentação encontrados são distintos e para tanto, precisam ser descritos com vistas a encontrar as regularidades e especificidades que os caracterizam. Na sequência, foram detalhados todos os problemas de segmentação retórica e linguística.

4.3.1 Problema de Segmentação Retórica (PROSEGR)

Os problemas de segmentação retórica acontecem nos casos em que a legenda antecipa ou atrasa informações por não acompanhar o fluxo da fala, incluindo hesitações, pausas e características peculiares da linguagem oral. Para Diaz Cintas e Remael (2007, p. 179) a segmentação pode refletir na dinâmica dos diálogos e afirmam: “Boa segmentação retórica ajuda a transmitir surpresa, suspense, ironia, hesitação etc.”²⁵

A legendagem do filme *Nosso Lar* pode ser caracterizada como uma boa legendagem no que diz respeito à segmentação retórica. Em todo o *corpus* de 1132 legendas foram encontradas apenas três ocorrências de problema de segmentação retórica por antecipar informação <PROSEGR_antecipouinformação> e nenhuma ocorrência por atrasar informação <PROSEGR_atrasouinformação>. Ainda assim, os três casos em que a retórica não foi mantida não comprometeram a surpresa, o suspense, a ironia, nem a hesitação, pois a cena em questão não trazia nenhum desses elementos. O motivo pelo qual foram classificadas como legendas com problema de segmentação retórica se deu porque a legenda não manteve o fluxo da fala, adiantando a fala seguinte.

As legendas nas quais foram identificados os problema de segmentação retórica são as de números 98, 222 e 223, como mostra a figura abaixo.

²⁵ Minha tradução de: *Good rhetorical segmentation helps convey surprise, suspense, irony, hesitation, etc.*

N	Concordance	Set	Tag	Word	#	t	#	os	#	os	#	os	File	%
1	parte de um diagnóstico?<PROSEGR><antecipouinformação>			3,054	95	0%	1	5%	0	0%	0	0%	tiquetadoTXT.txt	19%
2	<cpl24>que raiva, ódio, inveja, <PROSEGR><antecipouinformação>			3,039	94	5%	1	4%	0	9%	0	9%	tiquetadoTXT.txt	19%
3	minha alma, mas fui ouvido.<PROSEGR_antecipouinformação>			1,329	36	2%	0	0%	0	8%	0	8%	tiquetadoTXT.txt	8%

Figura 24: Tela da *Concord* mostrando as 3 ocorrências de PROSEGR

O primeiro trecho analisado diz:

<sub97>Não sei quanto tempo/durou aquela súplica</sub97>

<sub98>que vinha do fundo/da minha alma, mas fui ouvido.</sub98>.

Pela retórica do personagem, o trecho da legenda 98: *mas fui ouvido* foi antecipado em 750 milésimos de segundo. A pausa de 750ms pode parecer curta e insignificante, porém, do ponto de vista da legendagem e dos estudos de processamento de legendas com rastreamento ocular, 750 milissegundos pode ser fator determinante na qualidade da legenda. Essa afirmação é corroborada por (D'YDEWALLE; DE BRUYCKER, 2003, p. 269) quando relatam que o tempo médio de fixação de uma palavra por indivíduos adultos, sob um contexto de leitura normal (145, 160 e 180ppm), é de 200 a 250ms. Ou seja, com relação ao problema de segmentação retórica, 750ms configuram um atraso ou um adiantamento de três palavras, que pode comprometer o entendimento do produto audiovisual. Dependendo do conteúdo adiantado ou atrasado o espectador pode perder informações importantes de um filme, como um segredo, o *timing* de uma piada, o suspense, dentre outras coisas.

O segundo trecho analisado diz:

<sub222>Quando você poderia imaginar/que raiva, ódio, inveja,</sub222>

<sub223>egoísmo, intolerância,/fariam parte de um diagnóstico?</sub223>.

Seguindo o fluxo da fala, ou retórica do personagem, as duas legendas deveriam estar segmentadas em 3 blocos (legendas), a saber:

<sub222>Quando você poderia imaginar</sub222>

<sub223>que raiva, ódio, inveja,/egoísmo, intolerância,</sub223>

<sub224>fariam parte de um diagnóstico?</sub224>

É possível afirmar que não houve problemas de segmentação retórica na legendagem do filme *Nosso Lar*, pois nenhum dos três casos comprometeu o sentido das legendas. Embora os problemas de segmentação retórica não tenham sido recorrentes neste *corpus*, eles existem e podem causar dificuldades na recepção das legendas. Essa dificuldade foi relatada pelos informantes da pesquisa exploratória realizada pelo LEAD que demonstrou que surdos de todas as regiões do Brasil compreenderam legendas bem segmentadas, independente da velocidade na qual foram exibidas (ARAÚJO; NASCIMENTO, 2011).

4.3.2 Problemas de Segmentação Gramatical (PROSEGG)

A partir deste ponto serão analisados os problemas de segmentação gramatical nos níveis do sintagma e da oração. O sintagma, ou oração simples, segundo Perini (2010, p. 94) consta de um verbo, que pode ser acompanhado de certo número de complementos e adjuntos que podem ter a forma de SNs, SPs, SAdjs ou SAdvs. Um quadro do filme *Nosso Lar* ilustra os problemas de segmentação no nível do Sintagma.



Figura 25: Problema de segmentação no SN pela quebra de pre-nucleares+substantivo
Quadro do filme brasileiro *Nosso Lar* (2010)

4.3.2.1 Sintagma Nominal (SN)

Sintagma nominal é o recurso que a língua tem para se referir a entidades (coisas, pessoas ou abstrações), pode ser sujeito ou objeto (tudo que não é sujeito) de uma oração (PERINI, 2010). A estrutura interna do SN pode variar amplamente: pode ser composta de um possessivo + um nominal (*minha irmã*), ou apenas um nominal (*sombrinha*), e pode mesmo conter uma oração subordinada (*esses biscoitos que você faz*) (PERINI, 2010, p. 251).

No nível do SN foram encontradas 23 legendas com problemas de segmentação linguística ou 26,1% dos casos de problemas de segmentação, como mostra a tela da *Concord* na figura abaixo.

N	Concordance	Set	Tag	Word #	t. #	os. #	os. #	os. #	os. #	File	%
1] Este livro que<PROSEGG><SN_relativo+oração incompleta>			11,336	271	3%	2	4%	0	2%tiquetadoTXT.txt	72%
2	ditar para os nossos<PROSEGG><SN_pre-nucleares+subst>			11,353	271	8%	2	4%	0	3%tiquetadoTXT.txt	72%
3	espaço para mais<PROSEGG><SN_pre-nucleares+subst> <cpl24>um			9,819	244	5%	1	4%	0	3%tiquetadoTXT.txt	63%
4	<cpl14>Arrependimento<PROSEGG><SN_nominal+modif/modif+nominal>			8,413	220	6%	1	8%	0	4%tiquetadoTXT.txt	54%
5	é médico, mas antes<PROSEGG><SN_expressões/idiom/conv> <cpl26>de			8,780	229	7%	1	2%	0	6%tiquetadoTXT.txt	56%
6	<cpl12>já que é uma<PROSEGG><SN_pre-nucleares+subst> <cpl23>das			11,404	271	4%	2	6%	0	3%tiquetadoTXT.txt	73%
7	<cpl12>Ouvi a mesma<PROSEGG><SN_colocações/idiom/conv>			13,561	292	1%	2	4%	0	7%tiquetadoTXT.txt	87%
8	<cpl14>É por isso que<PROSEGG><SN_relativo+oração incompleta>			14,265	301	5%	2	9%	0	1%tiquetadoTXT.txt	91%
9	um dia todas<PROSEGG><SN_pre-nucleares+subst> <cpl22>as			12,320	275	4%	2	6%	0	9%tiquetadoTXT.txt	79%
10	que estão nos serviços<PROSEGG><SN_nominal+modif/modif+nominal>			11,816	273	9%	2	5%	0	6%tiquetadoTXT.txt	76%
11] Todas as cidades<PROSEGG><SN_nominal+modif/modif+nominal>			12,003	273	2%	2	9%	0	7%tiquetadoTXT.txt	77%
12	precisam de alguma<PROSEGG><SN_pre-nucleares+subst>			8,063	217	3%	1	4%	0	2%tiquetadoTXT.txt	51%
13	um histórico de todas<PROSEGG><SN_pre-nucleares+subst> <cpl38>as			2,920	91	9%	1	3%	0	9%tiquetadoTXT.txt	18%
14	<cpl12>Seu aparelho<PROSEGG><SN_nominal+modif/modif+nominal>			3,109	95	2%	1	5%	0	0%tiquetadoTXT.txt	20%
15	<cps10>Muito mais<PROSEGG><SN_superlativo+adj> <cps25>bem			1,934	59	4%	0	3%	0	2%tiquetadoTXT.txt	12%
16	vinha do fundo<PROSEGG><SN_colocações/idiom/conv> <cpl30>da			1,323	35	3%	0	0%	0	8%tiquetadoTXT.txt	8%
17	menos tira aquela<PROSEGG><SN_pre-nucleares+subst>			1,894	59	6%	0	1%	0	2%tiquetadoTXT.txt	12%
18	que fazemos é o nosso<PROSEGG><SN_pre-nucleares+subst>			3,459	99	6%	1	9%	0	2%tiquetadoTXT.txt	22%
19	um oceano de matéria<PROSEGG><SN_nominal+modif/modif+nominal>			6,131	157	9%	1	1%	0	9%tiquetadoTXT.txt	39%
20	muitos ainda mantinham<PROSEGG><SN_expressões/idiom/conv> <cpl37>as			6,337	163	8%	1	3%	0	1%tiquetadoTXT.txt	40%
21	<cpl14>São as grandes<PROSEGG><SN_nominal+modif/modif+nominal>			5,968	154	1%	1	9%	0	8%tiquetadoTXT.txt	38%
22	vida na Terra é que <PROSEGG><SN_relativo+oração incompleta>			4,006	113	5%	1	6%	0	6%tiquetadoTXT.txt	25%
23	pode procurar o irmão<PROSEGG><SN_titulo+nome proprio>			5,445	143	8%	1	3%	0	5%tiquetadoTXT.txt	34%

Figura 26: Tela da *Concord* mostrando as 23 ocorrências de problema de segmentação no nível do SN.

Os problemas no nível do SN são encontrados nas legendas onde há quebra de *pré-nucleares + substantivo*; *nominal + modificador*, ou na ordem inversa, *modificador + nominal*; *superlativo+adjetivo*; *relativo+oração incompleta*; *nome próprio*; *título+nome próprio*; e *colocações nominais, expressões idiomáticas e convencionais*.

Os 23 problemas em nível de sintagma nominal, encontrados no *corpus*, estão distribuídos nas 7 categorias:

- 8 ocorrências de <SN_pre-nucleares+subst>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SN_pre-nucleares+subst>
sub142	2	não	alta	Pelo menos tira <u>aquela/sensação</u> de fome.
sub215	2	38c	alta	guarda um histórico <u>de todas/as ações</u> praticadas no mundo material.
sub252	2	não	baixa	O bem que fazemos é <u>o nosso/advogado</u> pela eternidade.
sub580	2	não	alta	os seres precisam <u>de alguma/espiritualidade</u> , seja ela qual for.

sub709	2	não	alta	Há espaço <u>para mais/um</u> em nossa casa, André.
sub820	2	não	ideal	e ditar para <u>os nossos/irmãos</u> da Terra
sub823	2	não	alta	já que é <u>uma/das</u> minhas encarnações.
sub892	2	não	ideal	mas um dia <u>todas/as respostas</u> chegam...

Quadro 5: Problemas de segmentação gramatical - SN_pre-nucleares+subst

- 6 ocorrências de <SN_nominal+modif/modif+nominal>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SN_nominal+modif/modif+nominal>
sub227	2	não	ideal	<u>Seu aparelho/gastrointestinal</u> estava
sub428	2	não	baixa	São <u>as grandes/chagas</u> da humanidade...
sub439	2	37c	alta	[André] Existe um oceano <u>de matéria/invisível</u> ao redor da Terra.
sub606	2	não	baixa	<u>Arrependimento/inconsciente</u> , meu irmão.
sub853	2	não	baixa	aos que estão <u>nos serviços/edificantes</u> de nossa colônia.
sub868	2	não	ideal	[Governador] <u>Todas as cidades/espirituais</u> ao redor do planeta

Quadro 6: Problemas de segmentação gramatical - SN_nominal+modif/modif+nominal

- 1 ocorrência de <SN_superlativo+adj>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SN_superlativo+adj>
sub145	2	não	ideal	<u>Muito mais/bem disposto</u> , pelo visto.

Quadro 7: Problemas de segmentação gramatical - SN_superlativo+adj

- 3 ocorrências de <SN_relativo+oração incompleta>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SN_relativo+oração incompleta>
sub288	2	não	alta	A vida na Terra <u>é que/é uma cópia daqui</u> , André.
sub819	2	não	alta	[Emmanuel] Este livro <u>que/acabamos de completar</u>
sub1031	2	não	alta	É por isso <u>que/decidi ser médico</u> .

Quadro 8: Problemas de segmentação gramatical - SN_ relativo+oração incompleta

- Nenhuma ocorrência de <SN_nome próprio>
- 1 ocorrência de <SN_título+nome próprio>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SN_título+nome próprio>
sub389	2	não	ideal	Você pode procurar <u>o irmão/Genésio</u> no Ministério do Auxílio,

Quadro 9: Problemas de segmentação gramatical - SN_título+nome próprio

- 4 ocorrências de <SN_colocações/idiom/conv>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SN_colocações/idiom/conv>
sub98	2	não	baixa	que vinha <u>do fundo/da minha alma</u> , <u>mas fui ouvido</u> .
sub454	2	37c	alta	Aqui, muitos <u>ainda mantinham/as aparências</u> de suas vidas na Terra.
sub633	2	não	baixa	Você é médico, <u>mas antes/de tudo</u> é um homem de bem.
sub979	2	não	baixa	Ouvi <u>a mesma/coisa</u> anos atrás...

Quadro 10: Problemas de segmentação gramatical - SN_colocações/idiom/conv

Das 23 legendas com problemas de segmentação no nível do SN, 7 são de velocidade baixa, 6 de velocidade ideal e 10 de velocidade alta. Além disso, 3 legendas apresentaram linhas com número de caracteres superior a 35 caracteres. Diante desses resultados é possível afirmar que os problemas de segmentação no nível do sintagma nominal aconteceram em

legendas de duas linhas, com velocidade predominantemente alta e com maior frequência pela quebra de pre-nucleares + substantivos.

4.3.2.2 Sintagma Preposicionado (SP)

Os complementos nem sempre são SN, alguns são SP (preposição + SN), como em: [9] O vizinho apanhou da Elvira, [10] O vizinho apanhou por causa da Elvira, [11] O vizinho fugiu da Elvira (PERINI, 2010, p.89). A estrutura interna do SP é formada de preposição + SN e se relaciona com o verbo.

No nível do SP foram encontradas duas legendas com problemas de segmentação linguística ou 2,3% dos casos de problemas de segmentação, como mostra a tela da *Concord* na figura abaixo.

N	Concordance	Set	Tag	Word #	t	#	os	%	File	%	
1	que eu ando pelo-<PROSEGG><SP_prep+subst> <cpl26>vale da sombra			12,045	273	0%	2	0%	0 7%	tiquetadoTXT.txt	77%
2	que sabe sobre-<PROSEGG><SP_prep+subst> <cpl22>a medicina			7,271	200	1%	1	4%	0 6%	tiquetadoTXT.txt	46%

Figura 27: Tela da *Concord* mostrando as 2 ocorrências de problema de segmentação no nível do SP.

Os problemas no nível do SP são encontrados nas legendas quando há quebra de *preposição + substantivo*. Foram encontradas duas ocorrências em nível de SP:

➤ 2 ocorrências de <SP_prep+subst>

N° da legenda	N° de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SP_prep+subst>
sub523	2	não	alta	O que sabe <u>sobre/a medicina espiritual?</u>
sub871	2	não	baixa	Mesmo que eu ande <u>pelo/vale da sombra e da morte,</u>

Quadro 11: Problemas de segmentação gramatical - SP_prep+subst

Os problemas no SP foram os de menor ocorrência no *corpus*. Uma das legendas tem velocidade alta enquanto a outra, baixa e todos os problemas aconteceram em legendas de 2 linhas. Diante desses resultados é possível afirmar que os problemas de segmentação no nível do SP foram pouco frequentes nesse *corpus*, contudo, eles existem e podem ser evitados se a segmentação ocorrer entre os constituintes imediatos e não dentro deles.

4.3.2.3 Sintagma Verbal (SV)

Embora Perini (2010), em momento algum utilize o termo *sintagma verbal* em sua gramática descritiva do português brasileiro para se referir a construções verbais com mais de um verbo, tomei a iniciativa de adotar a nomenclatura *sintagma verbal* para as construções verbais constituídas de verbo principal, verbo auxiliar, modais, advérbios, colocações verbais, negações verbais e oblíquos.

No nível do SV foram encontradas 36 legendas com problemas de segmentação linguística ou 41% dos casos de problemas de segmentação, como mostra a tela da *Concord* na figura abaixo.

N	Concordance	Set	Tag	Word #	t	#	os	#	os	#	os	File	%
5	tempo atrás, eu inia<PROSEGG><SV_verbo+verbo> <cpl32>pensar que			9,419	237	6%	1	0%	0	0%	0%	tiquetadoTXT.txt	60%
6	<cpl14>Isso não podia<PROSEGG><SV_verbo+verbo> <cpl22>ter acontecido			7,539	206	9%	1	7%	0	8%	0%	tiquetadoTXT.txt	48%
7	Luiz] O que é que está<PROSEGG><SV_verbo+verbo> <cpl19>acontecendo			7,362	201	8%	1	5%	0	7%	0%	tiquetadoTXT.txt	47%
8	<cpl14>Senhor, eu vim<PROSEGG><SV_verbo+verbo> <cpl25>implorar a sua			8,248	219	8%	1	6%	0	3%	0%	tiquetadoTXT.txt	53%
9	estava lá nos<PROSEGG><SV_verbo>+oblíquo+verbo>			7,892	216	9%	1	2%	0	0%	0%	tiquetadoTXT.txt	50%
10] Mamãe não quer<PROSEGG><SV_verbo+verbo> <cpl15>falar do papai.			14,132	299	4%	2	6%	0	0%	0%	tiquetadoTXT.txt	90%
11	que pensei<PROSEGG><SV_verbo+adv> <cpl21>muito no papai.			13,963	295	0%	2	3%	0	9%	0%	tiquetadoTXT.txt	89%
12	<cpl12>ele deve ser<PROSEGG><SV_verbo+verbo> <cpl18>recebido por			15,401	323	6%	2	5%	0	8%	0%	tiquetadoTXT.txt	99%
13	do que vai adiantar<PROSEGG><SV_verbo+verbo> <cpl22>lembrar dele,			14,148	299	1%	2	7%	0	0%	0%	tiquetadoTXT.txt	90%
14	acho que eu não<PROSEGG><SV_negação+verbo> <cpl27>aguentaria			13,534	292	7%	2	3%	0	7%	0%	tiquetadoTXT.txt	86%
15	<cpl16>os clarins foram<PROSEGG><SV_verbo+verbo> <cpl17>ouvidos ao			11,638	273	8%	2	1%	0	4%	0%	tiquetadoTXT.txt	74%
16	sou eu que estou<PROSEGG><SV_verbo+verbo> <cpl31>lendo seus			11,164	269	6%	2	0%	0	1%	0%	tiquetadoTXT.txt	71%
17	<cpl15>É preciso viver<PROSEGG><SV_verbo+adv> <cpl21>de novo para			12,698	282	3%	2	4%	0	1%	0%	tiquetadoTXT.txt	81%
18	por que voltar<PROSEGG><SV_verbo+verbo> <cpl17>a viver na			12,668	281	7%	2	4%	0	0%	0%	tiquetadoTXT.txt	81%
19	Eu lhe ensino<PROSEGG><SV_verbo+verbo> <cpl27>a materializar			3,564	103	5%	1	1%	0	3%	0%	tiquetadoTXT.txt	23%
20	amigo parece ter<PROSEGG><SV_verbo+verbo> <cpl31>compreendido			2,666	84	7%	1	0%	0	7%	0%	tiquetadoTXT.txt	17%
21	atenção porque vou<PROSEGG><SV_verbo+verbo> <cpl33>dizer tudo			4,498	128	3%	1	2%	0	9%	0%	tiquetadoTXT.txt	28%
22	estavam sofrendo<PROSEGG><SV_verbo+adv> <cpl25>muito com a			3,789	105	3%	1	3%	0	4%	0%	tiquetadoTXT.txt	24%
23	eul Logo eu que nem<PROSEGG><SV_negação+verbo> <cpl32>acreditava			2,392	73	5%	0	0%	0	5%	0%	tiquetadoTXT.txt	15%
24	minha história estava <PROSEGG><SV_verbo+adv> <cpl19>apenas			152	0	2%	0	6%	0	1%	0%	tiquetadoTXT.txt	1%
25	que não pertencia<PROSEGG><SV_verbo+adv> <cpl24>mais ao mundo			63	0	4%	0	2%	0	0%	0%	tiquetadoTXT.txt	0%
26	<cpl14>Alguém precisa<PROSEGG><SV_verbo>+oblíquo+verbo> <cpl22>me			2,187	64	5%	0	2%	0	4%	0%	tiquetadoTXT.txt	14%
27	enquanto, precisa <PROSEGG><SV_verbo+verbo> <cpl28>receber			1,981	59	5%	0	5%	0	3%	0%	tiquetadoTXT.txt	12%
28	que não poderia<PROSEGG><SV_verbo+verbo> <cpl23>ouvir definição			6,764	171	6%	1	8%	0	0%	0%	tiquetadoTXT.txt	43%
29	<cpl13>Tio, me manda<PROSEGG><SV_verbo+adv> <cpl20>de volta, por			6,630	168	5%	1	7%	0	2%	0%	tiquetadoTXT.txt	42%
30	estou tentando<PROSEGG><SV_verbo+verbo> <cpl14>ajudar			7,085	188	2%	1	2%	0	5%	0%	tiquetadoTXT.txt	45%
31	<cpl13>Pode levantar<PROSEGG><SV_verbo+adv> <cpl19>um pouco por			6,818	174	3%	1	9%	0	4%	0%	tiquetadoTXT.txt	43%
32	<cpl14>Meu noivo deve<PROSEGG><SV_verbo+verbo> <cpl19>estar me			6,564	166	2%	1	6%	0	2%	0%	tiquetadoTXT.txt	42%
33	André] Só que eu ainda<PROSEGG><SV_verbo+adv> <cpl18>era o mesmo			4,989	134	1%	1	7%	0	2%	0%	tiquetadoTXT.txt	32%
34	<cpl11>Já me perdi<PROSEGG><SV_verbo+adv> <cpl24>muito por essas			4,926	131	5%	1	7%	0	1%	0%	tiquetadoTXT.txt	31%
35	as mudanças tinham<PROSEGG><SV_verbo+verbo> <cpl23>que ser mais			6,356	163	0%	1	4%	0	1%	0%	tiquetadoTXT.txt	40%
36	posso voltar<PROSEGG><SV_verbo+verbo> <cpl20>a atuar como			6,008	154	3%	1	9%	0	8%	0%	tiquetadoTXT.txt	38%

Figura 28: Tela da *Concord* mostrando as 36 ocorrências de problema de segmentação no nível do SV.

Os problemas no nível do SV são encontrados nas legendas quando há quebra de *verbo + verbo* (sejam principais, auxiliares ou modais); *verbo + advérbio*; *colocações verbais*; *negação* (não, nem etc.) + *verbo*; e *oblíquo* (antecedido ou não de verbo) + *verbo principal*. Os 36 problemas estão distribuídos nas 5 categorias, a saber:

- 21 ocorrências de <SV_verbo+verbo>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SV_verbo+verbo>
sub148	2	não	baixa	Por enquanto, <u>precisa/receber</u> apenas o tratamento.
sub197	2	não	alta	O amigo <u>parece ter/compreendido</u> o sentido da água,
sub260	2	não	baixa	Vem. <u>Eu lhe ensino/a materializar</u> outra roupa.
sub324	2	não	ideal	Mas preste atenção <u>porque vou/dizer tudo</u> uma vez só, combinado?
sub431	2	não	baixa	então <u>posso voltar/a atuar</u> como médico?

sub455	2	não	alta	Mas as mudanças <u>tinham/que ser</u> mais profundas.
sub470	2	não	alta	Meu noivo <u>deve/estar</u> me esperando.
sub484	2	não	ideal	Sabe que <u>não poderia/ouvir</u> definição melhor.
sub509	2	não	ideal	Não, <u>estou tentando/ajudar</u> também.
sub529	2	não	alta	[André Luiz] O que é que <u>está/acontecendo</u> comigo?
sub543	2	não	alta	Isso <u>não podia/ter acontecido</u> comigo!
sub593	2	não	baixa	Senhor, <u>eu vim/implorar</u> a sua permissão.
sub682	2	não	alta	Há algum tempo atrás, <u>eu iria/pensar</u> que você está reclamando,
sub708	2	não	ideal	<u>Não quero/incomodá-la</u> com isto.
sub782	2	não	alta	E se eu não <u>quiser/ficar ouvindo</u> estas lições?
sub806	2	não	alta	Agora sou eu <u>que estou/lendo</u> seus pensamentos, doutor.
sub839	2	não	ideal	os clarins <u>foram/ouvidos</u> ao longe.
sub916	2	não	alta	então por que <u>voltar/a viver</u> na Terra?
sub1022	2	não	alta	[Clarice] Mamãe <u>não quer/falar</u> do papai.
sub1023	2	não	alta	Mas do que <u>vai adiantar/lembrar</u> dele, Clarice?
sub1117	2	não	ideal	ele <u>deve ser/recebido</u> por todos

Quadro 12: Problemas de segmentação gramatical - SV_verbo+verbo

➤ 9 ocorrências de <SV_verbo+adv>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SV_verbo+adv>
sub5	2	não	alta	E logo descobri que <u>não pertencia/mais</u> ao mundo dos vivos.

sub11	2	não	alta	E minha história <u>estava/apenas recomeçando</u> .
sub274	2	não	baixa	Eles <u>estavam sofrendo/muito</u> com a minha doença.
sub354	2	não	baixa	<u>Já me perdi/muito</u> por essas trilhas.
sub358	2	não	alta	[André] Só que eu <u>ainda/era</u> o mesmo homem,
sub474	2	não	ideal	Tio, <u>me manda/de volta</u> , por favor!
sub488	2	não	baixa	<u>Pode levantar/um pouco</u> por favor?
sub918	2	não	ideal	É preciso <u>viver/de novo</u> para evoluir,
sub1011	2	não	ideal	Sabe que <u>pensei/muito</u> no papai, hoje?

Quadro 13: Problemas de segmentação gramatical - SV_verbo+adv

- Nenhuma ocorrência de <SV_colocações>
- 3 ocorrências de <SV_negação+verbo>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SV_negação+verbo>
sub179	2	não	alta	Logo eu! Logo eu <u>que nem/acreditava</u> em vida após a morte!
sub781	2	não	ideal	E se eu <u>não/quiser entender</u> , vó?
sub977	2	não	alta	Eu acho que eu <u>não/aguentaria</u> uma segunda vez.

Quadro 14: Problemas de segmentação gramatical - SV_negação+verbo

- 3 ocorrências de <SV_(verbo)+oblíquo+verbo>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SV_(verbo)+oblíquo+verbo>
sub163	2	não	baixa	Alguém <u>precisa/me dizer</u> alguma coisa!
sub568	2	não	ideal	Você estava lá <u>nos/acompanhando</u> . Sabe disso.

sub731	2	não	ideal	O ministro <u>vai/nos receber</u> em breve.
--------	---	-----	-------	---

Quadro 15: Problemas de segmentação gramatical - SV_(verbo)+oblíquo+verbo

Os problemas no nível do SV foram os de maior ocorrência no *corpus*, representando 41% dos problemas de segmentação. Dessas 36 legendas, 8 são de velocidade baixa, 12 de velocidade ideal e 16 de velocidade alta. Nenhuma legenda apresentou linhas com 36 caracteres ou mais. Diante dos resultados, é possível afirmar que esse tipo de problema de segmentação no SV acontece em legendas de 2 linhas, com velocidade alta e com maior frequência pela quebra de *verbo + verbo*.

Como o tipo de problema mais frequente no SV e no *corpus* em geral foi a quebra de *verbo + verbo*, com 21 ocorrências, maior atenção pode ser atribuída à segmentação do sintagma verbal, com vistas a diminuir os problemas enfrentados no processo de tradução. A partir dessas informações, as estratégias de tradução, no que diz respeito à segmentação, podem ser aprimoradas para que treinamentos oferecidos aos tradutores em formação, ou novatos, possam ser melhores e mais conscientes.

4.3.2.4 Sintagma Adverbial (SAdv)

Além de SN e SP, as orações também podem incluir constituintes de outras formas, como SAdv e SAdj sublinhadas nos exemplos encontrados em Perini (2010, p. 91): [17] Eu acho vocês geniais, [18] A garota telefonou agora mesmo. Segundo Perini (2010), esses constituintes não são sintagmas nominais porque, primeiro, não se referem a objetos do mundo real ou imaginário, depois, nunca poderiam, por exemplo, ser sujeitos da oração. Também não são SP porque não são precedidos de preposição. No entanto, eles também têm papel temático, ou seja, são semanticamente relacionados ao verbo de alguma maneira.

O Sintagma Adverbial (SAdv) é uma constituinte invariável em gênero e número que não concorda com SN nem com verbo, mas se relaciona semanticamente com o verbo. Diferente da noção de advérbio enquanto palavra terminada em *mente* o sintagma adverbial é muito mais que uma palavra, constitui um valor adverbial que pode ser entendido como um sintagma com papéis temáticos como lugar, tempo etc. Perini (2010) explica que não se conhece a lista completa dos advérbios

No nível do SAdv foram encontradas 6 legendas com problemas de segmentação, ou 5,8%, como mostra a tela da *Concord* na figura abaixo.

N	Concordance	Set	Tag	Word #	t. #	os. %	# os.	os. %	File	%
1	pai demorou apenas<PROSEGG><SAdv> <cpl33>18 anos para conseguir			9,738	244	7%	1	3%	0 2%tiquetadoTXT.txt	62%
2	vou esperar anos<PROSEGG><SAdv> <cpl21>e anos para retornar.			11,147	269	4%	1	0%	0 1%tiquetadoTXT.txt	71%
3	André morreu ao longo<PROSEGG><SAdv> <cpl32>de todos aqueles anos			13,358	291	5%	2	9%	0 5%tiquetadoTXT.txt	85%
4	ato realizado durante<PROSEGG><SAdv> <cpl21>longos e longos anos,			3,151	95	8%	1	6%	0 0%tiquetadoTXT.txt	20%
5	principal abriga em cada<PROSEGG><SAdv> <cpl30>uma das seis pontas da			4,750	130	5%	1	5%	0 0%tiquetadoTXT.txt	30%
6	Lar está no topo<PROSEGG><SAdv> <cpl39>de uma cadeia de			6,149	157	0%	1	1%	0 0%tiquetadoTXT.txt	39%

Figura 29: Tela da *Concord* mostrando as 6 ocorrências de problema de segmentação no nível do SAdv.

Os problemas no nível do SAdv são encontrados nas legendas quando há quebra da estrutura interna de um advérbio, ou valor adverbial. Neste *corpus* foram encontrados 6 problemas em nível de SAdv:

➤ 6 ocorrências de <SAdv>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SAdv>
sub230	2	não	baixa	Um ato realizado <u>durante/longos e longos anos</u> ,
sub341	2	não	alta	A praça principal abriga <u>em cada/uma das seis pontas da estrela</u>
sub440	2	39c	alta	Nosso Lar está <u>no topo/de uma cadeia</u> de montanhas espirituais.

sub704	2	não	ideal	meu pai demorou <u>apenas/18 anos</u> para conseguir esta casa.
sub805	2	não	alta	Não vou esperar <u>anos/e anos</u> para retornar.
sub965	2	não	ideal	um André morreu <u>ao longo/de todos aqueles anos</u> na cidade.

Quadro 16: Problemas de segmentação gramatical - SAdv

Os problemas no nível do SAdv representam 5,8% dos problemas de segmentação. Dentre as 6 legendas com problemas, uma é de velocidade baixa, 2 de velocidade ideal e 3 de velocidade alta, e apenas uma legenda (sub440) apresentou uma linha com 39 caracteres. Diante desses resultados, é possível afirmar que esse tipo de problema de segmentação no SAdv acontece em legendas de 2 linhas, com velocidade alta, quando há quebra da constituinte de valor adverbial.

4.3.2.5 Sintagma Adjetivo (SAdj)

O SAdj, para Perini (2010), concorda e se relaciona com algum SN da oração, além de ter valor adjetivo, serem variáveis em gênero e número e concordarem com algum SN da oração.

No nível do SAdj foram encontradas 9 legendas com problemas de segmentação ou 10,2%, como mostra a tela da *Concord* na figura abaixo.

N	Concordance	Set	Tag	Word #	t. #	os. #	.. #	os. #	.. #	os. #	File	%
1	lidar com as separações<PROSEGG><SAdj_subst+adj> <cpl20>temporárias da			10,770	264	1%	1	5%	0	9%	tiquetadoTXT.txt	69%
2	<cpl8>Há casos<PROSEGG><SAdj_subst+adj> <cpl24>mais difíceis			10,254	254	9%	1	9%	0	6%	tiquetadoTXT.txt	65%
3	lições que só a vida<PROSEGG><SAdj_subst+adj> <cpl22>na Terra pode			12,716	282	6%	2	5%	0	1%	tiquetadoTXT.txt	81%
4	de renovação da vida<PROSEGG><SAdj_subst+adj> <cpl34>humana chega			11,785	273	4%	2	4%	0	5%	tiquetadoTXT.txt	75%
5	um planejamento<PROSEGG><SAdj_subst+adj> <cpl17>minucioso			10,111	250	7%	1	8%	0	5%	tiquetadoTXT.txt	64%
6	com a forma<PROSEGG><SAdj_subst+adj> <cpl22>semelhante ao			3,353	96	7%	1	8%	0	1%	tiquetadoTXT.txt	21%
7	certeza, o deixou<PROSEGG><SAdj_subst+adj> <cpl18>mais bem			1,909	59	3%	0	2%	0	2%	tiquetadoTXT.txt	12%
8	a separação<PROSEGG><SAdj_subst+adj> <cpl20>temporária da			5,733	147	0%	1	6%	0	7%	tiquetadoTXT.txt	36%
9	Eu sei, foram momentos<PROSEGG><SAdj_subst+adj> <cpl21>difíceis no			3,630	104	3%	1	1%	0	3%	tiquetadoTXT.txt	23%

Figura 30: Tela da *Concord* mostrando as 9 ocorrências de problema de segmentação no nível do SAdj.

Os problemas no nível do SAdj são encontrados nas legendas onde há quebra de *substantivo* + *adjetivo* ou na ordem inversa *adjetivo* + *substantivo*. Foram 9, os problemas encontrados em nível de SAdj:

- 9 ocorrências de <SAdj_subst+adj>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SAdj_subst+adj>
sub143	2	não	alta	Com certeza, <u>o deixou/mais bem disposto</u> .
sub245	2	não	ideal	que se surpreendem <u>com a forma/semelhante</u> ao planeta.
sub264	2	não	alta	[Lísias] Eu sei, foram <u>momentos/difíceis</u> no hospital.
sub411	2	não	baixa	com <u>a separação/temporária</u> da morte,
sub727	2	não	ideal	É necessário <u>um planejamento/minucioso</u> para...
sub738	2	não	alta	Há <u>casos/mais difíceis</u> que o seu,

sub777	2	não	alta	aprenderá a lidar <u>com as separações/temporárias</u> da vida.
sub851	2	não	baixa	O processo de renovação <u>da vida/humana</u> chega ao ponto de ebulição.
sub919	2	não	alta	há lições que só <u>a vida/na Terra</u> pode nos dar,

Quadro 17: Problemas de segmentação gramatical – SAdj_subst+adj

Os problemas no nível do SAdj representam 10,2% dos problemas de segmentação. Dentre as 9 legendas com problemas, 2 são de velocidade baixa, 2 de velocidade ideal e 5 de velocidade alta. Nenhuma legenda apresentou linhas com mais de 35 caracteres. Com esses resultados, é possível afirmar que esse tipo de problema de segmentação no SAdj acontece em legendas de 2 linhas, com velocidade, predominantemente alta, e quando há quebra da constituinte *substantivo + adjetivo*.

Até agora foram analisados problemas de segmentação em orações simples, ou no nível do sintagma, mas a partir desse tópico serão analisados em construções mais complexas da língua, como *coordenação e subordinação*, que, segundo Perini (2010, p. 157) são recursos que a língua tem para juntar diversas orações em uma estrutura sintaticamente coesa. Um quadro do filme *Nosso Lar* ilustra os problemas de segmentação no nível da Oração.

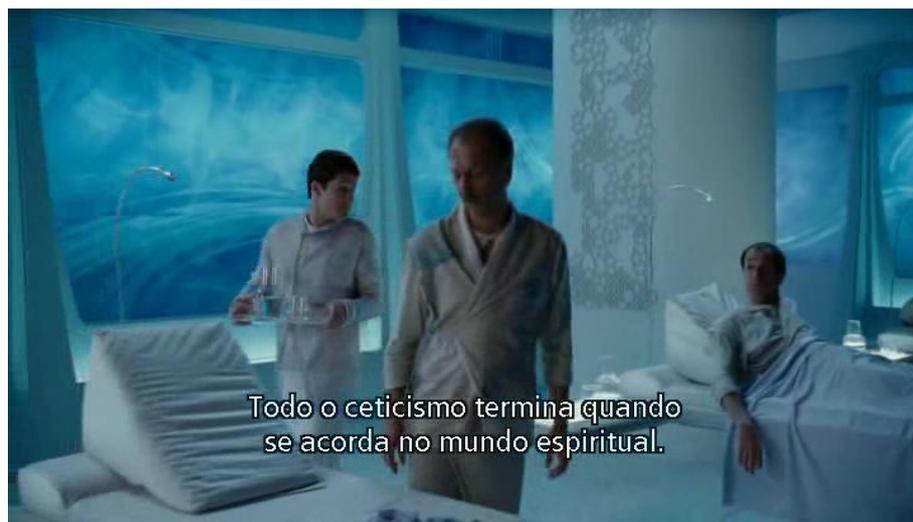


Figura 31: Problema de segmentação na Oração Subordinada pela quebra de conjunção+oração
Quadro do filme brasileiro *Nosso Lar* (2010)

4.3.2.6 Oração Coordenada (COORD)

A oração coordenada exige uma ou mais orações uma ao lado da outra, sem que uma faça parte da outra. As orações são chamadas de *coordenadas*, e o item linguístico *e* se chama *coordenador*, ou como na gramática tradicional, *conjunção coordenativa*. Alguns coordenadores são: *e*, *mas*, *logo* etc.

No tocante às orações coordenadas foram encontradas apenas 2 legendas com problemas de segmentação ou 2,3%, como mostra a tela da *Concord* na figura abaixo

N	Concordance	Set	Tag	Word #	t.	# os	%	# os	%	# os	t.	# os	%	File	%
1	como estou e você<PROSEGG><COORD_coordenador+oração>			14,018	298	2%	2.4%	0	0%	0	0%	0	0%	tiquetadoTXT.txt	90%
2	entrar e você<PROSEGG><COORD_coordenador+oração>			8,134	218	8%	1.4%	0	0.2%	0	0.2%	0	0.2%	tiquetadoTXT.txt	52%

Figura 32: Tela da *Concord* mostrando as 2 ocorrências de problema de segmentação no nível da oração coordenada.

Os problemas no nível da oração coordenada são encontrados nas legendas quando há quebra entre *coordenador e oração*, e entre partícula *negativa* (antecedida de coordenador) + *oração*. Os 2 problemas em nível de oração coordenada se encontram em uma das 2 categorias, a saber:

- 2 ocorrências de <COORD_coordenador+oração>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <COORD_conj+oração>
sub585	2	não	ideal	vamos entrar e <u>você/fala com o governador.</u>
sub1015	2	não	alta	Aflita como estou e <u>você/me vêm com essa melancolia.</u>

Quadro 18: Problemas de segmentação gramatical – COORD_coordenador+oração

- Nenhuma ocorrência de <COORD_negativa>

Os problemas no nível das coordenadas representam 2,3% dos problemas de segmentação. Dentre as 2 legendas com problemas, uma é de velocidade ideal e a outra de velocidade alta. Nenhuma legenda apresentou linhas com mais de 35 caracteres. Com esses resultados, é possível afirmar que esse tipo de problema de segmentação acontece em legendas de 2 linhas, com velocidade entre ideal e alta.

4.3.2.7 Oração Subordinada (SUBORD)

Em Perini (2010, p. 158) explica-se que a subordinada também é utilizada na língua para juntar orações, no entanto, diferentemente das coordenadas, as subordinadas não ficam uma ao lado da outra, mas uma *dentro* da outra. Ex: [8] [A tia Rosa disse que {o Rafael é médico}]. A oração maior – que se chama oração principal – está entre colchetes, []; e a oração menor, chamada oração subordinada, está entre chaves, {} (PERINI, 2010, p. 159). Algumas conjunções são: *que, porque, quando* etc.

No tocante às orações subordinadas foram encontradas 7 legendas com problemas de segmentação ou 8%, como mostra a tela da *Concord* na figura abaixo

N	Concordance	S	Tag	Word #	t	#	os	#	os	#	t	#	os	File	%
1	<cpl31>Todo o ceticismo termina quando<PROSEGG><SUBORD_conj+oração> <cpl30>se acorda no mundo espiritual	2,652	84	0%	0	0%	0	7%	tiquetadoTXT.txt	17%					
2	<cpl33>Lísias, eu não vou parar enquanto<PROSEGG><SUBORD_conj+oração> <cpl34>não souber o que está	2,724	84	7%	1	1%	0	7%	tiquetadoTXT.txt	17%					
3	<cpl28>Apenas uma entre as milhares<PROSEGG><SUBORD_conj+oração> <cpl38>de cidades que pairam sobre o	6,170	157	1%	1	1%	0	9%	tiquetadoTXT.txt	39%					
4	<cpl27>Tudo perde o sentido quando<PROSEGG><SUBORD_conj+oração> <cpl32>a gente acorda depois de morrer.	8,652	226	4%	1	1%	0	5%	tiquetadoTXT.txt	55%					
5	<cpl26>que vocês vão permitir que<PROSEGG><SUBORD_conj+oração> <cpl33>eu trate dele na terceira pessoa.	11,388	271	9%	2	5%	0	3%	tiquetadoTXT.txt	73%					
6	-> 01.21.52.867<lt> <cpl13>É por ela que<PROSEGG><SUBORD_conj+oração> <cpl19>estou fazendo isso. <velocidade	12,781	283	0%	2	6%	0	2%	tiquetadoTXT.txt	82%					
7	-> 01.33.40.245<lt> <cpl14>Aprendi quando<PROSEGG><SUBORD_conj+oração> <cpl25>tinha a sua idade, sabia?	14,431	307	0%	2	3%	0	2%	tiquetadoTXT.txt	92%					

Figura 33: Tela da *Concord* mostrando as 7 ocorrências de problema de segmentação no nível da oração subordinada.

Os problemas no nível da oração subordinada são encontrados nas legendas quando há quebra entre *conjunção e oração*; e entre *condicional “se” e oração*. Os 7 problemas em nível de oração subordinada foram encontrados apenas em uma das duas categorias, a saber:

- 7 ocorrências de <SUBORD_conj+oração>

Nº da legenda	Nº de linhas	A partir de 36cpl	Velocidade da legenda	LEGENDA com problema <SUBORD_conj+oração>
sub196	2	não	alta	Todo o ceticismo termina <u>quando/se acorda no mundo espiritual</u> .
sub201	2	não	alta	Lísias, eu não vou parar <u>enquanto/não souber o que está acontecendo</u> .
sub441	2	38c	alta	<u>Apenas uma entre as milhares/de cidades</u> que pairam sobre o planeta.
sub624	2	não	alta	Tudo perde o sentido <u>quando/a gente acorda depois de morrer</u> .
sub822	2	não	alta	que vocês vão permitir <u>que/eu trate dele na terceira pessoa</u> ,
sub923	2	não	ideal	É por ela <u>que/estou fazendo isso</u> .
sub1044	2	não	alta	Aprendi <u>quando/tinha a sua idade</u> , sabia?

Quadro 19: Problemas de segmentação gramatical – SUBORD_conj+oração

- Nenhuma ocorrência de <SUBORD_se>

Os problemas no nível da oração subordinada representaram 8% dos problemas. Dentre as 7 legendas com problemas, uma é de velocidade ideal e as demais são de velocidade alta. Apenas uma legenda apresentou uma linha com 38 caracteres. Desta forma é possível afirmar que esse tipo de problema de segmentação acontece em legendas de 2 linhas, com velocidade predominantemente alta.

4.4 Discussão dos Dados

Tendo em vista a análise dos 88 problemas de segmentação é possível concluir que os problemas de segmentação da LSE do filme *Nosso Lar* são encontrados em maior frequência no nível do sintagma, representando 86% do total, e em geral se caracterizam por acontecerem em legendas de duas linhas, que raramente ultrapassam os 35 caracteres, e em legendas com velocidade alta (acima de 16cps). O resultado geral está abaixo representado em forma de gráfico.

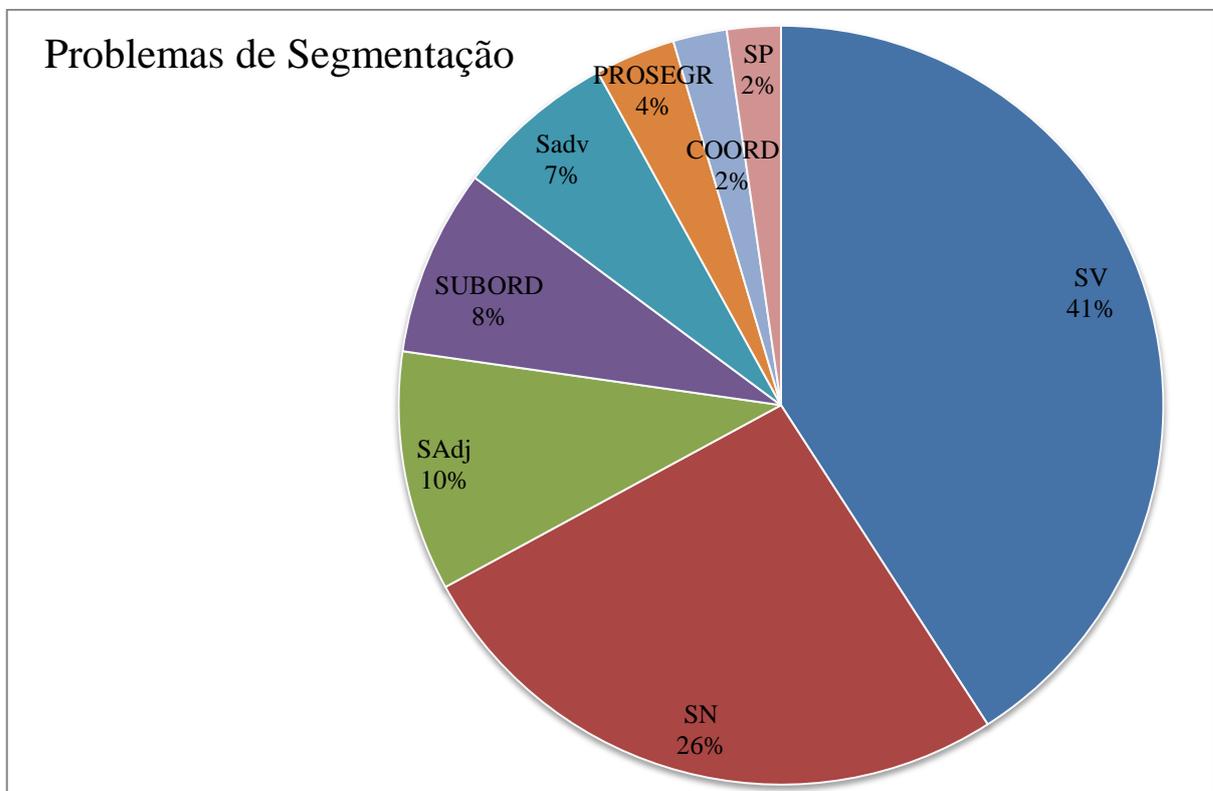


Gráfico 2: Resultado geral da Segmentação na legendagem do filme *Nosso Lar*

Por ordem de frequência, temos em primeiro lugar, os problemas no nível do sintagma verbal, que representam 41% do gráfico, em segundo lugar, os do sintagma nominal, que representam 26% do gráfico, em terceiro lugar, os do sintagma adjetivo, que correspondem a 10% do gráfico, em quarto lugar, os problemas no nível da oração subordinada, representando 8% do gráfico, em quinto lugar, os do sintagma adverbial, que representam 7% do gráfico, em sexto lugar os problemas de segmentação retórica, representando 4% do gráfico e por último os problemas no nível do sintagma preposicionado e da oração coordenada, que representam cada um 2% do gráfico.

A maior ocorrência de problemas no âmbito dos sintagmas do que no âmbito das orações e da retórica pode ser atribuída à dificuldade por parte dos tradutores de identificar os sintagmas, dado o dinamismo dos constituintes que podem exercer funções sintáticas ou papéis temáticos semelhantes, dependendo do contexto nos quais estão inseridos. Já as orações coordenadas e subordinadas podem ser identificadas com mais facilidade pelos tradutores, porque compreendem uma estrutura mais intuitiva formada de orações ligadas por coordenadores ou conjunções subordinadas. Essa estrutura mais estática das orações facilita o trabalho do tradutor de identificar o início e o fim das orações evitando a segmentação indevida do texto.

Os resultados dessa pesquisa, em especial as categorias definidas e os procedimentos de análise, servem de recursos para profissionais, estudantes e estudiosos da área de tradução que buscam desenvolver práticas de tradução mais conscientes e melhor fundamentadas. Os resultados também sugerem que maior atenção deve ser atribuída à análise de constituintes no nível do sintagma, não excluindo, é claro, os constituintes no nível da oração nem os problemas de retórica. Com tantos recursos oferecidos nessa pesquisa, um melhor treinamento pode ser dado a tradutores em formação visando a questão da quebra dos sintagmas.

A partir da análise dos problemas de segmentação na LSE do filme *Nosso Lar* e da discussão gerada, é possível concluir, com propriedade, que as orações são intuitivamente mais fáceis de serem segmentadas do que os sintagmas.

Essa dificuldade de caracterizar um sintagma ou mesmo uma oração pode ser superada por meio de um curso de legendagem que desenvolva estratégias de segmentação. O suporte

teórico metodológico desse curso sugerido compreenderia as categorias de problemas de segmentação definidas no *corpus* desta pesquisa. Vale informar que essas categorias gramatical e retórica não se aplicam apenas a análise do filme *Nosso Lar*, pelo contrário, a proposta desse trabalho é que essas categorias sejam passíveis de aplicação em qualquer outro *corpus* de legendas que se encontrem em língua portuguesa (variante brasileira).

É importante salientar que o foco desta pesquisa não foi avaliar o grau de dificuldade da análise de sintagmas, de orações e de retórica, nem a eficácia da boa e da má segmentação no processamento de LSE, pois, para isso seriam necessárias pesquisas experimentais com rastreador ocular, direcionadas à análise do desempenho dos espectadores quando expostos a legendas com e sem problemas de segmentação. Esta é uma sugestão para futuros trabalhos no âmbito da legendagem.

Esse desempenho dos espectadores quando expostos a duas condições: legendas com e sem problemas de segmentação, foi observado por Perego, et al. (2010) que não encontrou diferenças significativas em ambas as condições. Perego, et al. (2010) deduziram que um dos motivos para que não houvesse diferença no desempenho de sujeitos se deu pela escolha do vídeo, cuja compreensão dependia mais das legendas do que das imagem para acontecer e, partindo de uma observação pessoal, as categorias de problemas de segmentação não foram devidamente testadas, pois a pesquisadora verificou o processamento de legendas com problemas apenas nos níveis do Sintagma Nominal e Preposicionado, sem verificar legendas com problemas nos níveis do Sintagma Verbal, nos níveis das Coordenadas e Subordinadas, entre outros. A meu ver, se os sujeitos fossem expostos a legendas com problemas de segmentação de toda natureza e se os sujeitos fossem heterogêneos os resultados possivelmente seriam divergentes.

Os resultados desta pesquisa são perfeitamente comparáveis aos da pesquisa anterior (CHAVES, 2009), que teve como objetivo verificar a viabilidade do uso de *corpora* na análise e descrição de LSE e para isso, realizou uma análise de alguns parâmetros da legendagem, que foram contemplados também por esta pesquisa (Densidade Lexical (Velocidade da Legenda) e Segmentação) em seis trechos de variados gêneros da programação da *Rede Globo* (CHAVES, 2009). A Densidade Lexical corresponde à quantidade de palavras dividida pela duração das legendas em minuto, ou seja, corresponde ao

parâmetro de Velocidade da Legenda na presente pesquisa. O termo Densidade Lexical foi substituído por Velocidade da Legenda porque gerava ambiguidade semântica com o mesmo termo bastante utilizado na literatura de Linguística de *Corpus* e em outras áreas, mas que não possuíam o mesmo significado. Densidade Lexical, na Linguística de *Corpus* significa a riqueza do léxico, ou seja, o cálculo procura saber se aquele *corpus* compreende mais palavras (*tokens*) repetidas ou se compreende mais palavras distintas (*types*).

No que diz respeito à velocidade das legendas pode-se caracterizar que a legendagem do filme *Nosso Lar* foi de velocidade baixa, apresentando uma média de 115ppm, percentual bem abaixo daquele sugerido pelos surdos cearenses (ARAÚJO, 2008). Já os programas da *Globo* se caracterizam por terem legendas muito rápidas, apresentando uma média de 189ppm, enquanto as legendas dos mesmos programas confeccionadas pelo LEAD apresentaram uma média de 147ppm (CHAVES, 2009), podendo, com isso, inferir que as legendas do LEAD seguem o padrão sugerido pelos surdos cearenses. O fato de as legendas do filme *Nosso Lar* apresentarem velocidade de 115ppm não a classifica como boa ou má legenda, longe desse intuito, esses achados sugerem que pesquisas sejam feitas para verificar se a velocidade compromete ou não a compreensão do filme. Um fato possível de se afirmar diante desses resultados é que houve uma tendência por parte dos legendistas de aumentar o tempo de exibição das legendas de efeitos sonoros, pois os resultados demonstraram que as 620 legendas de velocidade baixa são, em sua maioria, identificação de efeitos sonoros. Essa tendência não pode ser generalizada como recorrente em todas as legendas de filmes, para isso mais pesquisas envolvendo a investigação de efeitos sonoros seriam necessárias.

A Segmentação na atual pesquisa foi analisada qualitativamente, de modo descritivo, enquanto que na pesquisa anterior foi observada apenas no aspecto quantitativo, considerando a segmentação como quantidade de legendas (CHAVES, 2009). Os seis trechos analisados da programação da *Globo* tiveram duração de 32 minutos, mas apenas 24min correspondia ao tempo de exibição das legendas. Nesses 24 minutos de exibição de legendas a *Globo* apresentou 491 legendas, enquanto o LEAD apresentou 589. Como as legendas da *Globo* foram caracterizadas como densas e rápidas, pôde-se inferir que se os trechos fossem mais segmentados, ou seja, fossem divididos em maior número de legendas, a leitura seria mais fluida (CHAVES, 2009), mas essa foi apenas uma hipótese que ainda não foi comprovada ou refutada. Já os resultados da análise qualitativa da legendagem do filme *Nosso Lar* foram bem

mais abrangentes e revelaram muitas características com relação à segmentação, como por exemplo, que os problemas de segmentação na legendagem acontecem em vários níveis da língua, em sua maioria no nível do sintagma. Com relação à esfera quantitativa, o filme apresentou 1132 legendas em 53 minutos de exibição de legendas, ou seja, 21 legendas por minuto, enquanto a Globo exibiu 20 e o LEAD 24.

Uma das considerações feitas em Chaves (2009) sobre o porquê de não ter realizado análise qualitativa, dizia: “*Por questões de tempo e espaço, a segmentação só será analisada quantitativamente. A análise qualitativa ficará para outro momento, pois talvez fosse preciso uma pesquisa inteira em torno da segmentação dessas legendas.*” (CHAVES, 2009, p. 33) Ressalto esse trecho acima citado, porque esse é um dado importante para a academia. Uma pesquisa de graduação gerou perspectivas tão positivas quanto à metodologia baseada em *corpus*, que incitou a ampliação da investigação, em larga escala, e que agora, com os resultados da atual pesquisa, gera mais expectativa e mais discussões no âmbito acadêmico.

Essa comparação dos resultados da pesquisa anterior (CHAVES, 2009) com a pesquisa atual, comprova mais uma vez que os estudos baseados em *corpora* estão gerando frutos no campo da Tradução Audiovisual e da Linguística de *Corpus*, além de surpreender com a quantidade de dados gerada nas análises eletrônicas. Essa gama de possibilidades encontrada na metodologia baseada em *corpus* abre espaço para investigações inéditas e relevantes nos campos da linguística, em especial no campo da tradução, como é o caso da próxima pesquisa do LEAD, que pretende compilar um *corpus* com legendas de programas de TV com vistas a analisar os problemas de segmentação a partir das etiquetas propostas na atual pesquisa, e propor a ressegmentação das legendas conforme categorias também encontradas aqui.

E para concluir, acredito que mais importante do que ser um *expert* em análise sintática do português brasileiro e saber distinguir todas as construções da língua, é saber apenas identificar os constituintes nos mais altos níveis sintáticos possíveis. Claro que isso não exime o legendista de ter uma noção mínima da gramática do português, apenas para fins de identificar as construções da nossa língua, mas o que pretendo com essa discussão é mostrar que o legendista não precisa realizar uma análise sintática pormenorizada para saber segmentar gramaticalmente uma legenda e ser um bom legendista. Esse trabalho

pormenorizado pode ficar a cargo do pesquisador, assim como foi nesta pesquisa. Por fim, se o legendista consegue identificar um sintagma, mesmo sem saber explicá-lo, ele deve ter sempre em mente que a segmentação não deve acontecer dentro da estrutura daquele sintagma, mas sim, entre sintagmas imediatos.

5 CONSIDERAÇÕES FINAIS

Esta pesquisa se dedicou a investigar os problemas de segmentação gramatical e de retórica, além de alguns parâmetros técnicos (velocidade da legenda, n° de linhas e de caracteres por linha) na LSE do filme *Nosso Lar*. Recapitulando, os objetivos desta pesquisa foram: 1) Estabelecer parâmetros para análise da segmentação na LSE, utilizando a Linguística de *Corpus*; 2) Descrever por meio de *corpus* quais são e como se caracterizam os problemas de segmentação na LSE, em língua portuguesa, de filmes. Para isso, outras metas precisaram ser cumpridas: Para estabelecer esses parâmetros, tive que propor um conjunto de etiquetas discursivas, além de realizar uma análise eletrônica da LSE com vistas a descrever os problemas de segmentação. Para propor o conjunto de etiquetas recorri aos parâmetros técnicos e linguísticos da legendagem e me aprofundi nos estudos linguísticos do português brasileiro. E para realizar a análise eletrônica da LSE recorri à Linguística de *Corpus*, mais especificamente ao programa *WordSmith Tools 5.0*.

Respondendo aos objetivos pretendidos, os resultados mostraram que os parâmetros para análise da segmentação na legendagem consistem no conjunto de etiquetas definido nesta pesquisa, que ajuda a identificar no *corpus* as ocorrências de problemas de segmentação gramatical, retórica e visual, além das etiquetas de análise de parâmetros técnicos, como velocidade da legenda e número de caracteres por linha, que ajudam a entender sob quais restrições esses problemas acontecem. Além disso, o passo a passo dessa pesquisa, bem como os procedimentos e as ferramentas também servem de parâmetros para a realização de uma análise e descrição das legendas e dos problemas de segmentação.

Os resultados da análise mostraram que os problemas de segmentação no nível do sintagma são os de maior ocorrência no *corpus*, em especial os do sintagma verbal, representando 41% dos 88 problemas encontrados. Esses problemas acontecem devido a quebra das estruturas *verbo + verbo* (sejam principais, auxiliares ou modais); *verbo + advérbio*; *colocações verbais*; *negação* (não, nem etc.) + *verbo*; e *oblíquo* (antecedido ou não de verbo) + *verbo principal*. De todos esses problemas encontrados no sintagma verbal, a quebra da estrutura *verbo + verbo* foi a de maior ocorrência, representado 18,5% dos problemas de segmentação.

O resultado geral da análise demonstrou que a legendagem do filme *Nosso Lar* apresenta poucos problemas de segmentação, representado apenas 7,7% das 1132 legendas analisadas. Além disso, a velocidade das legendas se caracteriza como lenta, tendo em vista que quase a metade das legendas, ou 84%, apresentaram velocidade média de 115ppm. Portanto, foi possível categorizar a legendagem do filme *Nosso Lar* como sendo de baixa velocidade, que não excede a média de 35 caracteres por linha de legenda, que não apresenta legendas com mais de duas linhas e que não tem grandes problemas de segmentação.

Esses resultados, tanto os que responderam ao objetivo de estabelecer parâmetros pra análise de LSE quanto os que responderam ao objetivo de descrever quais são e como se caracterizam os problemas de segmentação na LSE do filme *Nosso Lar*, estão bem estabelecidos neste trabalho, portanto servirão como base para muitos trabalhos da área e para profissionais que trabalham em prol da acessibilidade e da inclusão de surdos aos meios audiovisuais. Também para aqueles que lidam com a legendagem para ouvintes e aqueles que pretendem aplicar esses resultados em cursos e treinamentos de recursos humanos na área de tradução audiovisual. A pesquisa baseada em *corpus*, intimamente ligada às pesquisas descritivas, mais uma vez provou ser eficiente na análise e descrição de legendas.

Acredito que essa pesquisa poderá contribuir com o modo de praticar a segmentação na LSE, uma vez que a decisão do tradutor seja guiada, não por regras, mas pela reflexão de que a segmentação na LSE varia conforme suas características linguísticas e retóricas. Segmentar uma legenda através de regras é ficar preso à ideia de certo ou errado, além de correr o risco das regras não se aplicarem a todos os casos. Entretanto, se a segmentação for feita de modo descritivo, observando as características linguísticas e retóricas, os prejuízos podem ser diminuídos, tornando a recepção às legendas mais confortável e eficiente. Os problemas na legendagem acontecem e precisam ser descritos conforme suas peculiaridades, de modo a incitar reflexão teórica sobre as práticas. A meu ver, praticar uma segmentação linguística consciente é ter capacidade de reconhecer uma construção nos mais altos níveis sintáticos possíveis. E praticar uma segmentação retórica consciente é ter conhecimento dos prejuízos de se adiantar ou atrasar qualquer informação chave ao entendimento, um segredo, uma surpresa etc. Nesta pesquisa, optei por investigar a origem dos problemas de segmentação na LSE porque acredito que compreender a raiz do problema ajuda a desenvolver formas de melhorar o produto final.

Sobre os problemas que enfrentei ao longo dessa pesquisa de mestrado, gostaria de relatar que, como legendista, me encontrava em constantes conflitos na hora de segmentar uma legenda, pois sabia identificar os casos de má segmentação, mas não sabia explicá-los. Nesses momentos de dúvida contava muito com a intuição e com a questão da escolha, mas, como pesquisadora, tinha consciência de que a tarefa do tradutor exigia reflexão sobre as próprias práticas, e que essa reflexão me faria entender e explicar como e por que esses problemas aconteciam. Para prosseguir com a investigação sobre os problemas de segmentação precisava de um aporte teórico linguístico que se encaixasse na proposta dessa pesquisa, mas sem ter que me filiar a nenhuma corrente linguística. Após muita pesquisa recebi indicação da gramática descritiva do português brasileiro, que tinha como foco a análise sintática do português falado (PERINI, 2010). Essa gramática foi um achado para essa pesquisa, visto que há poucas, ou quase nenhuma gramática descritiva baseada em *corpus*, focada na sintaxe e semântica das construções linguísticas do português brasileiro, para tanto, tive que me aprofundar nesses estudos para entender o funcionamento da nossa língua e só a partir de então, analisar as legendas dos filmes quanto à segmentação linguística.

Outro problema que enfrentei ao longo dessa caminhada diz respeito ao procedimento de etiquetagem do *corpus*, ou análise das legendas, que demorou quase seis meses, muito mais do que imaginava e havia programado, pois tive que definir todas as categorias de problemas de segmentação à medida que eles iam aparecendo no *corpus* e, quando necessário, tinha que redefini-las.

As categorias de problemas de segmentação linguística, sugeridas nessa pesquisa não são representativas de todos os casos de má segmentação possíveis de acontecer em legendas. Isso porque as construções linguísticas se apresentam de formas variadas na língua, sendo inviável mapear todas as formas e variações dessas construções dentro de uma língua em constante processo de modificação. Portanto, o intuito foi observar as características de cada problema de segmentação no *corpus*, bem como descrever os padrões que os regem.

O tema tratado nesta pesquisa está longe de ser esgotado e os resultados gerados corroboram essa afirmação. Uma questão de pesquisa que pode ser explorada futuramente gira em torno da hipótese de que os sintagmas são mais difíceis de analisar que as orações

coordenadas e subordinadas. Outra proposta para futuras pesquisas pode estar na análise dos problemas de segmentação visual, que infelizmente, não foi possível dar conta nessa pesquisa, pois não haveria tempo suficiente e fugiria da proposta, tendo em vista um estudo minucioso das imagens do filme.

Outro tema bastante interessante a ser investigado na LSE é a questão dos efeitos sonoros, que demonstraram, nesta pesquisa, serem apresentados a uma velocidade diferente das demais legendas. Eles permanecem mais tempo na tela e geralmente não respeitam a relação de caracteres por segundo. São legendas sempre pequenas, apresentando uma média de 15 caracteres por legenda, mas costumam durar mais que 1 segundo, ou o tempo do som representado.

As discussões e reflexões levantadas nessa pesquisa são bastante promissoras para os estudos de Tradução Audiovisual e Linguística de *Corpus*, além de relevantes, do ponto de vista teórico, para a formação de tradutores, para o aperfeiçoamento das práticas profissionais e para subsidiar as pesquisas do LEAD. Concluo este trabalho expondo que os frutos gerados por esta pesquisa são de extrema importância para a formação do legendista e para o desenvolvimento de pesquisas visando a acessibilidade e a inclusão de surdos e ensurdecidos aos meios audiovisuais.

REFERÊNCIAS

ADORÁVEL Rosa. Direção: Aurora Miranda Leão. Ceará, Brasil, 2008, 1 DVD (19min), color., legendas (para surdos em português), audiodescrição e audionavegação.

ÁGUAS de Romanza. Direção: Gláucia Soares e Patrícia Baía. Ceará, Brasil, 2002, 1 DVD (15min), color., legendas (para surdos em português), audiodescrição e audionavegação.

ALONSO, F. Algo más que suprimir barreras: conceptos y argumentos para una accesibilidad universal. In: **Trans**, Revista de Traductología, Universidad de Málaga, n. 11, p. 15-30, 2007.

ARAÚJO, V. L. S.; NASCIMENTO, A. K. P. Investigando parâmetros de legendas para Surdos e Ensurdecidos no Brasil. In: FROTA, M. P.; MARTINS, M. A. P. (orgs.). **Tradução em Revista**, v. 2, p. 1-18, 2011. Disponível em: <<http://www.maxwell.lambda.ele.puc-rio.br/18862/18862.>>. Acesso em: 01 de fevereiro de 2012.

ARAÚJO, V. L. S. Por um modelo de legendagem para Surdos no Brasil. In VERAS, V. (org.). **Tradução e Comunicação**, Revista Brasileira de Tradutores, São Paulo: UNBERO, n. 17, p. 59-76, 2008.

ARAÚJO, V.L.S. Subtitling for the deaf and hard-of-hearing in Brazil In: ORERO, P.; REMAEL, A. (orgs.). **Media for All: Subtitling for the Deaf, Audio Description and Sign Language**. Kenilworth: Nova Jersey, EUA: Rodopi, v. 30, p. 99-107, 2007.

ARAÚJO, V.L.S. A legendagem para surdos no Brasil. In: LIMA, P. L.C.; ARAÚJO, A. D. (orgs.). **Questões de Linguística Aplicada: Miscelânea**. Fortaleza: EdUECE, p. 163-188, 2005.

ARAÚJO, V. L. S. Closed subtitling in Brazil In: ORERO, P. (org.). **Topics in Audiovisual translation**. Amsterdã: John Benjamins Publishing Company, v. 1, p. 199-212, 2004.

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, p. 233-250, 1993.

BAKER, M. Corpora in translation studies: an overview and some suggestions for future research. In: **Target**, v. 7, n. 3, p. 223-243, 1995.

BAKER, M. Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, H. (ed.). **Terminology, LSP and translation**. Amsterdã, Filadélfia: John Benjamins, p. 175-187, 1996.

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

BEZERRA de Menezes: O Diário de um Espírito. Direção: Glauber Filho e Joe Pimentel. Brasil: Fox, 2008, 1 DVD (75min), região 4, NTSC, color., legendas e audiodescrição.

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

BRASIL. Ministério Público Federal. Inquérito Civil Tutela Coletiva n° 1.34.001.005949/2010-32, da Vara Cível da Seção Judiciária do Estado de São Paulo, São Paulo, SP, 13 de fevereiro de 2012. **Ação Civil Publica com pedido de Tutela Antecipada** da Procuradoria Regional dos Direitos do Cidadão, São Paulo, 2012. Disponível em <http://www.prsp.mpf.gov.br/sala-de-imprensa/pdfs-das-noticias/PRDC_ACP_Legenda_0002444-7.2012.4.03.6100.PDF/at_download/file>. Acesso em: 08 de março de 2012.

CAPISTRANO no Quilo. Direção: Firmino Holanda. Ceará, Brasil, 2006, 1 DVD (21min), color., legendas (para surdos em português), audiodescrição e audionavegação.
REISADO Miudim. Direção: Petrus Cariry. Ceará, Brasil, 2008, 1 DVD (13min), color., legendas (para surdos em português), audiodescrição e audionavegação.

CHAVES, É. G. **Legenda para Surdos no Brasil**: uma análise baseada em corpus. 52f. Monografia (Bacharelado em Letras Inglês). Universidade Estadual do Ceará, Fortaleza-CE, 2009.

CHICO Xavier. Direção: Daniel Filho. Brasil: Sony Pictures, 2010, 1 DVD (125min), região 4, NTSC, color., legendas (para surdos em português) e audiodescrição.

CORISCO e Dadá. Direção: Rosemberg Cariry. Brasil: Riofilme, 1996, 1 DVD (110min), região 4, NTSC, color., legendas (para surdos em português), janela de LIBRAS, audiodescrição, audionavegação.

DÍAZ CINTAS, J. **Teoría y práctica de la subtítulos**. Barcelona: Ariel Cine, 2003.

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

DIAZ CINTAS, J.; ANDERMAN, G. **New Trends in Audiovisual Translation**. Bristol, UK: Multilingual Matters, 2009.

D'YDEWALLE, G.; Muylle, P.; VAN RENSBERGEN, J. Attention shifts in partially redundant information situations. In: GRONER, R.; MCCONKIE, G. W.; MENZ, C. (eds.). **Eye Movements and Human Information Processing**. Amsterdam: North-Holland, p. 375-84, 1985.

D'YDEWALLE, G.; PRAET, C.; VERFAILLIE, K.; VAN RENSBERGEN, J. 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. (eds.). **Eye Movements: From Physiology to Cognition**. Amsterdam, New York: Elsevier Science Publishers, p. 313-321, 1987.

D'YDEWALLE, G.; PRAET, C.; VERFAILLIE, K.; DE GRAEF, P.; VAN RENSBERGEN, J. A One Line Text is not Half a Two Line Text', In: GRONER, R.; D'YDEWALLE, G.; PARHAM, R. (eds.). **From Eye to Mind: Information Acquisition in Perception, Search, and Reading**. Amsterdam: Elsevier Science Publishers, p. 205-213, 1990.

D'YDEWALLE, G.; DE BRUYCKER, W. Reading Native and Foreign Language Television Subtitles in Children and Adults. In: HYÖNÄ, J.; RADACH, R.; DEUBEL, H. (eds.). **The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research**. Amsterdam: North-Holland, p. 671-684, 2003. ISBN: 0-444-51020-6.

ENSAIO sobre a Cegueira. Direção: Fernando Meirelles. Brasil: Fox do Brasil, 2008, 1 DVD (121min), região 4, NTSC, color, legendas, dublagem e audiodescrição. Tradução de: Blindness.

EUGENI, C. Respeaking political debate for the deaf: the Italian case. In: BALDRY, A.; MONTAGNA, E. (eds.). **Interdisciplinary perspectives on multimodality: theory and practice**. Proceedings of the third international conference on multimodality. Campobasso: Palladino, p. 191-205, 2008.

FEITOSA, M. P. **Legendagem comercial e legendagem pirata: um estudo comparado**. 162f. Tese (Doutorado em Linguística), Universidade Federal de Minas Gerais, Brasil, 2009.

FRANCO, E.; ARAUJO, 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.

FRIES, P. H. Themes, methods of developments and texts. In: FRIES, P. H.; HASAN, R. (eds.). **On Subject and Theme**. Amsterdã: John Benjamins, p. 317-359, 1995.

FRIES, P. H. The flow of information in a written English text. In: **Relations and Functions within and around a Language**. Londres: Continuum, p. 117-155, 2002.

FRONTEIRA. Direção: Rafael Conde. Brasil: Filmegraph e Camisa Listrada, 2008, 1 DVD (85min), região 4, NTSC, color., legendas (para surdos em português) e audiodescrição.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. **An introduction to Functional Grammar**. 3 ed. Londres: Edward Arnold, 2004.

HURTADO, C J. Una gramática local del guión audiodescrito. Desde la semántica a la pragmática de un nuevo tipo de traducción. In: Catalina Jiménez Hurtado (ed.). **Traducción y accesibilidad**. Frankfurt: Peter Lang Internationaler Verlag der Wissenschaften, p. 55-80, 2007.

HURTADO, C. J.; RODRÍGUEZ, A.; SEIBEL, C. (Eds.). **Por um corpus de cine: teoria y práctica de la audiodescripción**. Granada: Ediciones Tragacanto, 2010.

IBGE. **Censo Demográfico 2010**, Pessoas com deficiência. Instituto Brasileiro de Pesquisa e Estatística: Rio de Janeiro, 2010. Disponível em:
<ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_Gerais_da_Amostra/tab1.pdf> Acesso em 27 de maio de 2012.

IRMÃOS de Fé. Direção: Moacyr Góes. Brasil: Sony Pictures, 2004, 1 DVD (105min), 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, Dimitra. **Subtitling for the Deaf and Hard of Hearing: A corpus-based methodology for the analysis of subtitles with a focus on segmentation and deletion**. 366f. 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: 07 de julho de 2011.

LEECH, G. Adding Linguistic Annotation. In: WYNNE, M. **Developing Linguistic Corpora: a Guide to Good Practice**. Oxford: Oxbrow Books, p. 17-29, 2005. Disponível em: <<http://ahds.ac.uk/linguistic-corpora/>>. Acesso em: 07 de julho de 2011.

MCENERY, T.; XIAO, R, Y. **Corpus-based language studies: an advanced resource book**. London: Routledge, 2006.

NEVES, J. A world of change in a changing world. In: DIAZ CINTAS, J.; ORERO, P.; REMAEL, A. (org.). **Media for All: Subtitling for the Deaf, Audio Description and Sign Language**. Kenilworth, Nova Jersey: Rodopi, p. 89-98, 2007.

NEVES, J. (2005) **Audiovisual Translation: Subtitling for the Deaf and the Hard-of-Hearing**. 357f. Tese (Doutorado): Roehampton University, London, 2005.

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.

O'DONNELL, M. B. The Use of Annotated Corpora for New Testament Discourse Analysis: A Survey of Current Practice and Future Prospects. In: PORTER, S. E.; REED, J. T. (eds.). **Discourse Analysis and the New Testament: Results and Applications**. Sheffield: Sheffield Academic Press, p. 71-117, 1999.

O GRÃO. Direção: Petrus Cariry. Brasil: Iluminura Filmes, 2008, 1 DVD (88min), região 4, NTSC, color., legendas (para surdos em português), janela de LIBRAS, audiodescrição e audionavegação.

O SIGNO da Cidade. Direção: Carlos Alberto Riccelli. Brasil: Europa Filmes, 2008, 1 DVD (95min), região 4, NTSC, color., legendas (para surdos em português) e audiodescrição.

PEREGO, E. Evidence of explicitation in subtitling: towards a characterization. In: **Across Languages and Cultures**, A Multidisciplinary Journal for Translation and Interpreting Studies. Budapest: Adadémiai Kiadó, v. 4, n. 1, p. 63-88, 2003.

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.

PEREGO, E. The codification of non-verbal information in subtitled texts. In: DIAZ CINTAS, J. (ed.). **New trends in audiovisual translation**. Bristol, UK: Multilingual Matters, p. 58-69, 2009.

PEREGO, E.; DEL MISSIER, F.; PORTA, M.; MOSCONI, M. The Cognitive Effectiveness of Subtitle Processing. In: **Media Psychology**. Philadelphia, PA: Routledge, p. 243-272, 2010. ISSN: 1521-3269 print/1532-785X online. Disponível em: <<http://www2.units.it/delmisfa/papers/SubtitlesProcessing2010.pdf>>. Acesso em: 07 de julho de 2011.

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

PORTAL Acessibilidade Total: definição de legenda descritiva. Disponível em: <<http://www.acessibilidadetotal.com.br/legenda-ou-janela-de-interpretres/>>. Acesso em: 13 de março de 2012.

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.

SALWAY, A. A corpus-based analysis of audio description. In: DIAZ CINTAS, J.; ORERO, P.; REMAEL, A. (eds.). **Media for All** - Subtitling for the Deaf, Audio Description, and Sign Language. Amsterdam, New York: Rodopi, p. 151-174, 2007.

SELVATICI, C. **Closed Caption: conquistas e questões**. 80f. Dissertação (Mestrado) Pontífca Universidade Católica do Rio, 2010.

SUBTITLE Workshop. Sítio do programa. Disponível em: <www.urusoft.net>. Acesso em: 10 de março de 2012.

TAGNIN, S. E. O. Um corpus multilíngue para ensino e tradução – o CoMET: da construção à exploração. In: TAGNIN, S. E. O. (Org.). **TradTerm**. São Paulo: Humanitas/FFLCH/USP, v. 10, p. 117-142, 2004.

TAGNIN, S. E. O. Glossário de Linguística de Corpus. In: VIANA, V.; TAGNIN, S. E. O. (orgs.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, p. 357-361, 2010. ISBN: 978-85-63623-66-9.

TAGNIN, S. E. O. Corpora on-line. In: VIANA, V.; TAGNIN, S. E. O. (orgs.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, p. 363-370, 2010. ISBN: 978-85-63623-66-9.

THOMPSON, G. **Introducing Functional Grammar**. 2 ed. Londres: Hodder Arnold, 2004.

THOMPSON, G. Unfolding Theme: the development of clausal and textual perspectives on Theme. In: HASAN, R; MATTHIESSEN, C; WEBSTER, J. **Continuing Discourse on Language**. A functional perspective, v. 2. Londres: Equinox, p. 671-696, 2007.

TORRENTE 3, El Protector. Direção: Santiago Segura. Espanha: Santiago Segura, 2005. 1 DVD (91min), região 2, PAL, color., legendas (para surdos em espanhol, e inglês) e audiodescrição.

TURMA da Mônica em: Cine Gibi 5. Direção: Maurício de Sousa. Brasil: Maurício de Sousa, 2010, 1 DVD (71min), região 4, NTSC, color., audiodescrição.

WINCAPS: sítio do programa disponível em: <www.sysmedia.com>. Acesso em: 10 de fevereiro de 2012.

WORDSMITH Tools: sítio do programa disponível em: <<http://www.lexically.net/wordsmith/index.html>>. Acesso em: 10 de março de 2012.

ZANETTIN, F; BERNARDINI, S.; STEWART, D. **Corpora in Translator Education**. Manchester: St. Jerome Publishing, 2003.

APÊNDICES

APÊNDICE A – Conjunto de etiquetas

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>	
ETIQUETAS INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO RETÓRICA	
<PROSEGR_antecipouinformação>	
<PROSEGR_atrasouinformação>	
ETIQUETAS 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>	

APÊNDICE B - Legendas etiquetadas com problemas de segmentação

<sub5><2L>5
 <t>00:02:59,012 --> 00:03:02,607</t>
 <cpl33>E logo descobri que não pertencia<PROSEGG><SV_verbo+adv>
 <cpl24>mais ao mundo dos vivos.
 <velocidade da legenda_alta 57c/3,6s></sub5>

<sub11><2L>11
 <t>00:03:36,349 --> 00:03:38,977</t>
 <cpl23>E minha história estava <PROSEGG><SV_verbo+adv>
 <cpl19>apenas recomeçando.
 <velocidade da legenda_alta 42c/2,6s></sub11>

<sub98><2L>98
 <t>00:15:44,109 --> 00:15:48,375</t>
 <cpl18>que vinha do fundo<PROSEGG><SN_colocações/idiom/conv>
 <cpl30>da minha alma, mas fui ouvido.<PROSEGR_antecipou>
 <velocidade da legenda_baixa 48c/4,2s ></sub98>

<sub142><2L>142
 <t>00:23:30,275 --> 00:23:32,743</t>
 <cpl22>Pelo menos tira aquela<PROSEGG><SN_pre-nucleares+subst>
 <cpl17>sensação de fome.
 <velocidade da legenda_alta 39c/2,4s></sub142>

<sub143><2L>143
 <t>00:23:32,811 --> 00:23:35,177</t>
 <cps21>Com certeza, o deixou<PROSEGG><SAdj_subst+adj>
 <cps18>mais bem disposto.
 <velocidade da legenda_alta 39c/2,3s></sub143>

<sub145><2L>145
 <t>00:23:37,783 --> 00:23:40,308</t>
 <cps10>Muito mais<PROSEGG><SN_superlativo+adj>
 <cps25>bem disposto, pelo visto.
 <velocidade da legenda_ideal 35c/2,5s></sub145>

<sub148><2L>148
 <t>00:23:46,525 --> 00:23:50,120</T>
 <cpl21>Por enquanto, precisa <PROSEGG><SV_verbo+verbo>
 <cpl28>receber apenas o tratamento.
 <velocidade da legenda_baixa 49c/3,6s></sub148>

<sub163><2L>163
 <t>00:24:28,667 --> 00:24:31,033</t>
 <cpl14>Alguém precisa<PROSEGG><SV_(verbo)+oblíquo+verbo>
 <cpl22>me dizer alguma coisa!

<velocidade da legenda_baixa 36c/2,6s></sub163>

<sub179><2L>179

<t>00:26:01,560 --> 00:26:05,052</t>

<cpl24>Logo eu! Logo eu que nem<PROSEGG><SV_negação+verbo>

<cpl32>acreditava em vida após a morte!

<velocidade da legenda_alta 56c/3,5></sub179>

<sub196><2L>196

<t>00:26:49,441 --> 00:26:53,309</t>

<cpl31>Todo o ceticismo termina quando<PROSEGG><SUBORD_conj+oração>

<cpl30>se acorda no mundo espiritual.

<velocidade da legenda_alta 61c/3,8></sub196>

<sub197><2L>197

<t>00:26:54,646 --> 00:26:57,740</t>

<cpl18>O amigo parece ter<PROSEGG><SV_verbo+verbo>

<cpl31>compreendido o sentido da água,

<velocidade da legenda_alta 49c/3,1s></sub197>

<sub201><2L>201

<t>00:27:05,190 --> 00:27:09,126</t>

<cpl33>Lísias, eu não vou parar enquanto<PROSEGG><SUBORD_conj+oração>

<cpl34>não souber o que está acontecendo.

<velocidade da legenda_alta 67c/4s></sub201>

<sub215><2L>215

<t>00:27:42,193 --> 00:27:46,220</t>

<cpl28>guarda um histórico de todas<PROSEGG><SN_pre-nucleares+subst>

<cpl38>as ações praticadas no mundo material.

<velocidade da legenda_alta 66c/4s></sub215>

<sub222><2L>222

<t>00:28:03,248 --> 00:28:07,685</t>

<cpl28>Quando você poderia imaginar

<cpl24>que raiva, ódio, inveja,<PROSEGR_antecipouinformação>

<velocidade da legenda_baixa 52c/4,4s></sub222>

<sub223><2L>223

<t>00:28:07,752 --> 00:28:11,688</t>

<cpl22>egoísmo, intolerância,

<cpl31>fariam parte de um diagnóstico?<PROSEGR><antecipouinformação>

<velocidade da legenda_baixa 53c/3,9s></sub223>

<sub227><2L>227

<t>00:28:22,100 --> 00:28:24,534</t>

<cpl12>Seu aparelho<PROSEGG><SN_nominal+modif/modif+nominal>

<cpl23>gastrointestinal estava

<velocidade da legenda_ideal 35c/2,4s></sub227>

<sub230><2L>230

<t>00:28:33,244 --> 00:28:37,237</t>

<cpl24>Um ato realizado durante<PROSEGG><SAdv>

<cpl21>longos e longos anos,

<velocidade da legenda_baixa 45c/4s></sub230>

<sub245><2L>245

<t>00:30:10,708 --> 00:30:14,371</t>

<cpl30>que se surpreendem com a forma<PROSEGG><SAdj_subst+adj>

<cpl22>semelhante ao planeta.

<velocidade da legenda_ideal 52c/3,6s></sub245>

<sub252><2L>252

<t>00:30:35,800 --> 00:30:40,863</t>

<cpl27>O bem que fazemos é o nosso<PROSEGG><SN_pre-nucleares+subst>

<cpl25>advogado pela eternidade.

<velocidade da legenda_baixa 52c/5s></sub252>

<sub260><2L>260

<t>00:31:12,971 --> 00:31:17,874</t>

<cpl18>Vem. Eu lhe ensino<PROSEGG><SV_verbo+verbo>

<cpl27>a materializar outra roupa.

<velocidade da legenda_baixa 45c/4,9s></sub260>

<sub264><2L>264

<t>00:31:30,688 --> 00:31:33,851</t>

<cpl33>[Lísias] Eu sei, foram momentos<PROSEGG><SAdj_subst+adj>

<cpl21>difíceis no hospital.

<velocidade da legenda_alta 54c/3,1></sub264>

<sub274><2L>274

<t>00:32:03,087 --> 00:32:06,716</t>

<cpl21>Eles estavam sofrendo<PROSEGG><SV_verbo+adv>

<cpl25>muito com a minha doença.

<velocidade da legenda_baixa 46c/3,6s></sub274>

<sub288><2L>288

<t>00:32:39,590 --> 00:32:42,286</t>

<cpl21>A vida na Terra é que <PROSEGG><SN_relativo+oração incompleta>

<cpl25>é uma cópia daqui, André.

<velocidade da legenda_alta 46c/2,7s></sub288>

<sub324><2L>324

<t>00:34:26,297 --> 00:34:30,256</t>

<cpl29>Mas preste atenção porque vou<PROSEGG><SV_verbo+verbo>

<cpl33>dizer tudo uma vez só, combinado?

<velocidade da legenda_ideal 62c/3,9s></sub324>

<sub341><2L>341
 <t>00:35:28,559 --> 00:35:32,154</t>
 <cpl32>A praça principal abriga em cada<PROSEGG><SAdv>
 <cpl30>uma das seis pontas da estrela
 <velocidade da legenda_alta 62c/3,6s></sub341>

<sub354><2L>354
 <t>00:36:23,881 --> 00:36:26,645</t>
 <cpl11>Já me perdi<PROSEGG><SV_verbo+adv>
 <cpl24>muito por essas trilhas.
 <velocidade da legenda_baixa 35c/2,7s></sub354>

<sub358><2L>358
 <t>00:36:40,464 --> 00:36:42,955</t>
 <cpl25>[André] Só que eu ainda<PROSEGG><SV_verbo+adv>
 <cpl18>era o mesmo homem,
 <velocidade da legenda_alta 43c/2,5s></sub358>

<sub389><2L>389
 <t>00:38:54,632 --> 00:38:58,568</t>
 <cpl26>Você pode procurar o irmão<PROSEGG><SN_título+nome próprio>
 <cpl33>Genésio no Ministério do Auxílio,
 <velocidade da legenda_ideal 59/3,9s></sub389>

<sub411><2L>411
 <t>00:41:17,107 --> 00:41:19,871</t>
 <cpl15>com a separação<PROSEGG><SAdj_subst+adj>
 <cpl20>temporária da morte,
 <velocidade da legenda_baixa 35c/2,7s></sub411>

<sub428><2L>428
 <t>00:42:13,631 --> 00:42:16,464</t>
 <cpl14>São as grandes<PROSEGG><SN_nominal+modif/modif+nominal>
 <cpl23>chagas da humanidade...
 <velocidade da legenda_baixa 37c/2,8s></sub428>

<sub431><2L>431
 <t>00:42:28,312 --> 00:42:31,839</t>
 <cpl18>então posso voltar<PROSEGG><SV_verbo+verbo>
 <cpl20>a atuar como médico?
 <velocidade da legenda_baixa 38c/3,5s></sub431>

<sub439><2L>439
 <t>00:42:59,143 --> 00:43:03,170</t>
 <cpl37>[André] Existe um oceano de
matéria<PROSEGG><SN_nominal+modif/modif+nominal>
 <cpl28>invisível ao redor da Terra.
 <velocidade da legenda_alta 65c/4s></sub439>

<sub440><2L>440

<t>00:43:03,514 --> 00:43:07,075</t>

<cpl22>Nosso Lar está no topo<PROSEGG><SAdv>

<cpl39>de uma cadeia de montanhas espirituais.

<velocidade da legenda_alta 61c/3,5s></sub440>

<sub441><2L>441

<t>00:43:07,551 --> 00:43:11,544</t>

<cpl28>Apenas uma entre as milhares<PROSEGG><SUBORD_conj+oração>

<cpl38>de cidades que pairam sobre o planeta.

<velocidade da legenda_alta 66c/4s></sub441>

<sub454><2L>454

<t>00:44:05,042 --> 00:44:09,103</t>

<cpl28>Aqui, muitos ainda mantinham<PROSEGG><SN_colocações/idiom/conv>

<cpl37>as aparências de suas vidas na Terra.

<velocidade da legenda_alta 65c/4s></sub454>

<sub455><2L>455

<t>00:44:09,680 --> 00:44:12,513</t>

<cpl22>Mas as mudanças tinham<PROSEGG><SV_verbo+verbo>

<cpl23>que ser mais profundas.

<velocidade da legenda_alta 45c/2,8s></sub455>

<sub470><2L>470

<t>00:44:52,489 --> 00:44:54,480</t>

<cpl14>Meu noivo deve<PROSEGG><SV_verbo+verbo>

<cpl19>estar me esperando.

<velocidade da legenda_alta 33c/2s></sub470>

<sub474><2L>474

<t>00:45:05,169 --> 00:45:07,330</t>

<cpl13>Tio, me manda<PROSEGG><SV_verbo+adv>

<cpl20>de volta, por favor!

<velocidade da legenda_ideal 33c/2,1s></sub474>

<sub484><2L>484

<t>00:45:43,207 --> 00:45:46,176</t>

<cpl20>Sabe que não poderia<PROSEGG><SV_verbo+verbo>

<cpl23>ouvir definição melhor.

<velocidade da legenda_ideal 43c/2,9s></sub484>

<sub488><2L>488

<t>00:45:57,187 --> 00:45:59,985</t>

<cpl13>Pode levantar<PROSEGG><SV_verbo+adv>

<cpl19>um pouco por favor?

<velocidade da legenda_baixa 32c/2,8s></sub488>

<sub509><2L>509

<t>00:46:49,306 --> 00:46:51,433</t>
 <cpl19>Não, estou tentando<PROSEGG><SV_verbo+verbo>
 <cpl14>ajudar também.
 <velocidade da legenda_ideal 33c/2,1s></sub509>

<sub523><2L>523
 <t>00:47:19,870 --> 00:47:22,065</t>
 <cpl16>O que sabe sobre<PROSEGG><SP_prep+subst>
 <cpl22>a medicina espiritual?
 <velocidade da legenda_alta 38c/2,2s></sub523>

<sub529><2L>529
 <t>00:47:34,818 --> 00:47:37,685</t>
 <cpl31>[André Luiz] O que é que está<PROSEGG><SV_verbo+verbo>
 <cpl19>acontecendo comigo?
 <velocidade da legenda_alta 50c/2,8s></sub529>

<sub543><2L>543
 <t>00:48:30,641 --> 00:48:32,871</t>
 <cpl14>Isso não podia<PROSEGG><SV_verbo+verbo>
 <cpl22>ter acontecido comigo!
 <velocidade da legenda_alta 36c/2,2s></sub543>

<sub568><2L>568
 <t>00:50:06,636 --> 00:50:09,469</t>
 <cpl18>Você estava lá nos<PROSEGG><SV_(verbo)+oblíquo+verbo>
 <cpl25>acompanhando. Sabe disso.
 <velocidade da legenda_ideal 43c/2,8s></sub568>

<sub580><2L>580
 <t>00:52:18,802 --> 00:52:22,465</t>
 <cpl27>os seres precisam de alguma<PROSEGG><SN_pre-nucleares+subst>
 <cpl35>espiritualidade, seja ela qual for.
 <velocidade da legenda_alta 62c/3,6s></sub580>

<sub585><2L>585
 <t>00:52:33,483 --> 00:52:36,384</T>
 <cpl19>vamos entrar e você<PROSEGG><COORD_coordenador+oração>
 <cpl22>fala com o governador.
 <velocidade da legenda_ideal 41c/2,9s></sub585>

<sub593><2L>593
 <t>00:53:36,246 --> 00:53:40,273</t>
 <cpl14>Senhor, eu vim<PROSEGG><SV_verbo+verbo>
 <cpl25>implorar a sua permissão.
 <velocidade da legenda_baixa 39c/4s></sub593>

<sub606><2L>606
 <t>00:54:14,784 --> 00:54:18,220</t>

<cpl14>Arrependimento<PROSEGG><SN_nominal+modif/modif+nominal>
 <cpl24>inconsciente, meu irmão.
 <velocidade da legenda_baixa 38c/3,4s></sub606>

<sub624><2L>624
 <t>00:55:53,450 --> 00:55:56,851</t>
 <cpl27>Tudo perde o sentido quando<PROSEGG><SUBORD_conj+oração>
 <cpl32>a gente acorda depois de morrer.
 <velocidade da legenda_alta 59c/3,4s></sub624>

<sub633><2L>633
 <t>00:56:16,673 --> 00:56:21,542</t>
 <cpl24>Você é médico, mas antes<PROSEGG><SN_colocações/idiom/conv>
 <cpl26>de tudo é um homem de bem.
 <velocidade da legenda_baixa 50c/4,8s></sub633>

<sub682><2L>682
 <t>01:00:17,880 --> 01:00:21,407</t>
 <cpl29>Há algum tempo atrás, eu iria<PROSEGG><SV_verbo+verbo>
 <cpl32>pensar que você está reclamando,
 <velocidade da legenda_alta 61c/3,5s></sub682>

<sub704><2L>704
 <t>01:01:24,147 --> 01:01:27,981</t>
 <cpl22>meu pai demorou apenas<PROSEGG><SAdv>
 <cpl33>18 anos para conseguir esta casa.
 <velocidade da legenda_ideal 55c/3,8s></sub704>

<sub708><2L>708
 <t>01:01:37,326 --> 01:01:39,260</t>
 <cpl9>Não quero<PROSEGG><SV_verbo+verbo>
 <cpl21>incomodá-la com isto.
 <velocidade da legenda_ideal 30c/1,9s></sub708>

<sub709><2L>709
 <t>01:01:39,328 --> 01:01:41,853</t>
 <cpl19>Há espaço para mais<PROSEGG><SN_pre-nucleares+subst>
 <cpl24>um em nossa casa, André.
 <velocidade da legenda_alta 43c/2,5s></sub709>

<sub727><2L>727
 <t>01:02:48,431 --> 01:02:51,525</t>
 <cpl28>É necessário um planejamento<PROSEGG><SAdj_subst+adj>
 <cpl17>minucioso para...
 <velocidade da legenda_ideal 45c/3,1s></sub727>

<sub731><2L>731
 <t>01:03:00,977 --> 01:03:03,275</t>
 <cpl14>O ministro vai<PROSEGG><SV_(verbo)+oblíquo+verbo>

<cpl21>nos receber em breve.
<velocidade da legenda_ideal 35c/2,3s></sub731>

<sub738><2L>738
<t>01:03:19,061 --> 01:03:21,120</t>
<cpl8>Há casos<PROSEGG><SAdj_subst+adj>
<cpl24>mais difíceis que o seu,
<velocidade da legenda_alta 32c/2s></sub738>

<sub777><2L>777
<t>01:06:17,039 --> 01:06:20,497</t>
<cpl35>aprenderá a lidar com as separações<PROSEGG><SAdj_subst+adj>
<cpl20>temporárias da vida.
<velocidade da legenda_alta 55c/3,4s></sub777>

<sub781><2L>781
<t>01:06:31,954 --> 01:06:34,013</t>
<cpl11>E se eu não<PROSEGG><SV_negação+verbo>
<cpl20>quiser entender, vó?
<velocidade da legenda_ideal 31c/2s></sub781>

<sub782><2L>782
<t>01:06:34,390 --> 01:06:37,120</t>
<cpl18>E se eu não quiser<PROSEGG><SV_verbo+verbo>
<cpl27>ficar ouvindo estas lições?
<velocidade da legenda_alta 45c/2,7s></sub782>

<sub805><2L>805
<t>01:09:49,285 --> 01:09:51,845</t>
<cpl20>Não vou esperar anos<PROSEGG><SAdv>
<cpl21>e anos para retornar.
<velocidade da legenda_alta 41c/2,5s></sub805>

<sub806><2L>806
<t>01:09:55,291 --> 01:09:58,488</t>
<cpl22>Agora sou eu que estou<PROSEGG><SV_verbo+verbo>
<cpl31>lendo seus pensamentos, doutor.
<velocidade da legenda_alta 53c/3,2s></sub806>

<sub819><2L>819
<t>01:11:27,650 --> 01:11:30,619</t>
<cpl27>[Emmanuel] Este livro que<PROSEGG><SN_relativo+oração incompleta>
<cpl21>acabamos de completar
<velocidade da legenda_alta 48c/2,9s></sub819>

<sub820><2L>820
<t>01:11:30,686 --> 01:11:33,314</t>
<cpl22>e ditar para os nossos<PROSEGG><SN_pre-nucleares+subst>
<cpl15>irmãos da Terra

<velocidade da legenda_ideal 37c/2,6s></sub820>

<sub822><2L>822

<t>01:11:38,560 --> 01:11:42,223</t>

<cpl26>que vocês vão permitir que<PROSEGG><SUBORD_conj+oração>

<cpl33>eu trate dele na terceira pessoa,

<velocidade da legenda_alta 59c/3,6s></sub822>

<sub823><2L>823

<t>01:11:42,531 --> 01:11:44,658</t>

<cpl12>já que é uma<PROSEGG><SN_pre-nucleares+subst>

<cpl23>das minhas encarnações.

<velocidade da legenda_alta 35c/2,1s></sub823>

<sub839><2L>839

<t>01:12:47,529 --> 01:12:49,656</t>

<cpl16>os clarins foram<PROSEGG><SV_verbo+verbo>

<cpl17>ouvidos ao longe.

<velocidade da legenda_ideal 33c/2,1s></sub839>

<sub851><2L>851

<t>01:13:26,435 --> 01:13:34,171</t>

<cpl31>O processo de renovação da vida<PROSEGG><SAdj_subst+adj>

<cpl34>humana chega ao ponto de ebulição.

<velocidade da legenda_baixa 65c/7,7s></sub851>

<sub853><2L>853

<t>01:13:40,115 --> 01:13:45,576</t>

<cpl26>aos que estão nos serviços<PROSEGG><SN_nominal+modif/modif+nominal>

<cpl29>edificantes de nossa colônia.

<velocidade da legenda_baixa 55c/5,4s></sub853>

<sub868><L>868

<t>01:14:58,026 --> 01:15:02,395</t>

<cpl31>[Governador] Todas as cidades<PROSEGG><SN_nominal+modif/modif+nominal>

<cpl31>espirituais ao redor do planeta

<velocidade da legenda_ideal 62c/4,3s></sub868>

<sub871><2L>871

<t>01:15:17,379 --> 01:15:22,840</t>

<cpl22>Mesmo que eu ande pelo<PROSEGG><SP_prep+subst>

<cpl26>vale da sombra e da morte,

<velocidade da legenda_baixa 48c/5,4s></sub871>

<sub892><2L>892

<t>01:19:51,486 --> 01:19:53,886</t>

<cpl16>mas um dia todas<PROSEGG><SN_pre-nucleares+subst>

<cpl22>as respostas chegam...

<velocidade da legenda_ideal 38c/2,4s></sub892>

<sub916><2L>916
 <t>01:21:31,720 --> 01:21:34,018</t>
 <cpl20>então por que voltar<PROSEGG><SV_verbo+verbo>
 <cpl17>a viver na Terra?
 <velocidade da legenda_alta 37c/2,3s></sub916>

<sub918><2L>918
 <t>01:21:36,758 --> 01:21:39,226</t>
 <cpl15>É preciso viver<PROSEGG><SV_verbo+adv>
 <cpl21>de novo para evoluir,
 <velocidade da legenda_ideal 36c/2,4s></sub918>

<sub919><2L>919
 <t>01:21:39,294 --> 01:21:42,024</t>
 <cpl23>há lições que só a vida<PROSEGG><SAAdj_subst+adj>
 <cpl22>na Terra pode nos dar,
 <velocidade da legenda_alta 45c/2,7s></sub919>

<sub923><2L>923
 <t>01:21:50,672 --> 01:21:52,867</t>
 <cpl13>É por ela que<PROSEGG><SUBORD_conj+oração>
 <cpl19>estou fazendo isso.
 <velocidade da legenda_ideal 32c/2,2s></sub923>

<sub965><2L>965
 <t>01:25:45,673 --> 01:25:49,439</t>
 <cpl24>um André morreu ao longo<PROSEGG><SAAdv>
 <cpl32>de todos aqueles anos na cidade.
 <velocidade da legenda_ideal 56c/3,7s></sub965>

<sub977><2L>977
 <t>01:27:16,264 --> 01:27:18,926</t>
 <cpl18>Eu acho que eu não<PROSEGG><SV_negação+verbo>
 <cpl27>aguentaria uma segunda vez.
 <velocidade da legenda_alta 45c/2,6s></sub977>

<sub979><2L>979
 <t>01:27:20,768 --> 01:27:23,134</t>
 <cpl12>Ouvi a mesma<PROSEGG><SN_colocações/idiom/conv>
 <cpl19>coisa anos atrás...
 <velocidade da legenda_baixa 31c/2,3s></sub979>

<sub1011><2L>1011
 <t>01:30:53,748 --> 01:30:56,308</t>
 <cpl15>Sabe que pensei<PROSEGG><SV_verbo+adv>
 <cpl21>muito no papai, hoje?
 <velocidade da legenda_ideal 36c/2,5s></sub1011>

<sub1015><2L>1015
 <t>01:31:01,989 --> 01:31:05,049</t>
 <cpl24>Aflita como estou e você<PROSEGG><COORD_coordenador+oração>
 <cpl27>me vêm com essa melancolia.
 <velocidade da legenda_alta 51c/3s></sub1015>

<sub1022><2L>1022
 <t>01:31:25,680 --> 01:31:28,114</t>
 <cpl26>[Clarice] Mamãe não quer<PROSEGG><SV_verbo+verbo>
 <cpl15>falar do papai.
 <velocidade da legenda_alta 41c/2,4s></sub1022>

<sub1023><2L>1023
 <t>01:31:28,182 --> 01:31:30,776</t>
 <cpl23>Mas do que vai adiantar<PROSEGG><SV_verbo+verbo>
 <cpl22>lembrar dele, Clarice?
 <velocidade da legenda_alta 45c/2,6s></sub1023>

<sub1031><2L>1031
 <t>01:31:53,641 --> 01:31:55,632</t>
 <cpl14>É por isso que<PROSEGG><SN_relativo+oração incompleta>
 <cpl18>decidi ser médico.
 <velocidade da legenda_alta 32c/2s></sub1031>

<sub1044><2L>1044
 <t>01:33:37,845 --> 01:33:40,245</t>
 <cpl14>Aprendi quando<PROSEGG><SUBORD_conj+oração>
 <cpl25>tinha a sua idade, sabia?
 <velocidade da legenda_alta 39c/2,4s></sub1044>

<sub1117><2L>1117
 <t>01:41:47,534 --> 01:41:49,502</t>
 <cpl12>ele deve ser<PROSEGG><SV_verbo+verbo>
 <cpl18>recebido por todos
 <velocidade da legenda_ideal 30c/1,9s></sub1117>