



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

MARIA HELENA CLARINDO GABRIEL

**PROBLEMAS DE SEGMENTAÇÃO LINGUÍSTICA NA LEGENDAGEM PARA
SURDOS E ENSURDECIDOS (LSE) DE “CHEIAS DE CHARME”:
UMA ANÁLISE BASEADA EM *CORPUS***

FORTALEZA – CEARÁ

2015

MARIA HELENA CLARINDO GABRIEL

PROBLEMAS DE SEGMENTAÇÃO LINGUÍSTICA NA LEGENDAGEM PARA
SURDOS E ENSURDECIDOS (LSE) DE “CHEIAS DE CHARME”:
UMA ANÁLISE BASEADA EM *CORPUS*

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

Orientadora: Profa. Dra. Vera Lúcia Santiago Araújo.

FORTALEZA – CEARÁ

2015

Dados Internacionais de Catalogação na Publicação

Universidade Estadual do Ceará

Sistema de Bibliotecas

Gabriel, Maria Helena Clarindo.

Problemas de segmentação linguística na Legendagem para Surdos e Ensurdidos (LSE) de "Cheias de charme": uma análise baseada em corpus [recurso eletrônico] / Maria Helena Clarindo Gabriel. - 2015.

1 CD-ROM: 4 ¼ pol.

CD-ROM contendo o arquivo no formato PDF do trabalho acadêmico com 109 folhas, acondicionado em caixa de DVD Slim (19 x 14 cm x 7 mm).

Dissertação (mestrado acadêmico) - Universidade Estadual do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Linguística Aplicada, Fortaleza, 2015.

Área de concentração: Linguagem e interação.

Orientação: Prof.^a Dra. Vera Lúcia Santiago Araújo.

1. Tradução Audiovisual Acessível. 2. Legendagem para surdos e ensurdidos. 3. Linguística de Corpus. 4. Problemas de Segmentação. 5. Telenovela. I. Título.

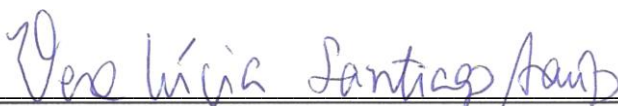
MARIA HELENA CLARINDO GABRIEL

PROBLEMAS DE SEGMENTAÇÃO LINGUÍSTICA NA LEGENDAGEM PARA
SURDOS E ENSURDECIDOS (LSE) DE “CHEIAS DE CHARME”:
UMA ANÁLISE BASEADA EM *CORPUS*

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


Aprovada em: 02 /07 /2015.

BANCA EXAMINADORA



Profa. Dra. Vera Lúcia Santiago Araújo (Orientadora)

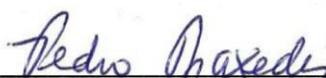
Universidade Estadual do Ceará – UECE



Profa. Dra. Renata Oliveira Mascarenhas

Fundação Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico – FUNCAP

Universidade Estadual do Ceará – UECE



Prof. Dr. Pedro Henrique Lima Praxedes Filho

Universidade Estadual do Ceará – UECE

As minhas filhas, Beatriz Helena Clarindo Elias e Bruna Helena Clarindo Elias, por serem a motivação primeira deste trabalho.

AGRADECIMENTOS

A Deus, fonte de luz, força e coragem que me sustentam e me sustentaram durante as atribuições vividas durante esta fase da minha vida.

A minha mãe, exemplo de mulher guerreira, pelos sacrifícios que fez para priorizar a minha educação e a dos meus irmãos. Ainda por me ajudar bastante na educação das minhas filhas.

A minha família, meus pais, meus irmãos, minhas filhas, meus tios (irmãos de criação), por permanecerem ao meu lado em todos os momentos, mostrando o real significado de uma família.

A minha orientadora, professora Dra. Vera Lúcia Santiago Araújo, por ter me apresentado a Tradução Audiovisual e Legendagem para Surdos e Ensurdidos. Acima de tudo, por ser fonte de conhecimento, inspiração e perseverança durante toda a minha caminhada no curso de mestrado, em especial na realização desta dissertação, que chamo de ‘sonho realizado’.

Ao Programa de Mestrado em Linguística Aplicada (POSLA), pela oportunidade de participar dessa experiência única.

A todos os professores do Programa de Mestrado em Linguística Aplicada (POSLA) e a todos os colegas do curso, pelos momentos de aprendizagem e prazer compartilhados.

A todos os membros do Grupo LEAD, em especial a Élide Gama Chaves, Ana Katarinna do Nascimento, Daniel de Albuquerque e Arraes e Sofia Nicolau Amoreira, por terem contribuído com minha aprendizagem tanto no âmbito teórico da Legendagem para Surdos e Ensurdidos, quanto no âmbito das tecnologias e habilidades com os programas computacionais, fundamentais para a realização deste trabalho.

Disse-lhe, então, o anjo: Maria, não temas,
porque achaste graça diante de Deus.

(Lucas, 1:30)

RESUMO

Este trabalho dedica-se a uma das modalidades da Tradução Audiovisual acessível (TAVa) que promove uma forma de acessibilidade sociocultural às pessoas com deficiência auditiva: a Legendagem para Surdos e Ensurdidos (LSE). Seu objetivo principal é identificar os problemas de segmentação linguística (PROSEGL) na LSE da telenovela brasileira ‘Cheias de Charme’, via Linguística de *Corpus* (LC) e tendo como base teórico-metodológica os estudos da tradução, mais especificamente a tradução audiovisual (TAV). Os problemas de segmentação linguística são divisões inapropriadas dentro de uma legenda e podem ocorrer entre linhas e, ou entre legendas. Elas também podem ocorrer em nível de sintagmas: sintagma verbal, sintagma nominal, sintagma adjetival, sintagma adverbial, sintagma preposicional; e orações coordenadas e subordinadas. Para alcançar o objetivo desta pesquisa, analisou-se 1725 LSEs, pertencentes a cinco capítulos de ‘Cheias de Charme’, exibida pela emissora brasileira Rede Globo de Televisão. Esse *corpus* é um *subcorpus* do projeto intitulado *A segmentação na legendagem para surdos e ensurdidos (LSE): um estudo baseado em corpus* (Projeto CORSEL), desenvolvido pelo grupo Legendagem e Audiodescrição (LEAD) da Universidade Estadual do Ceará (UECE) coordenados pela Profa. Dra. Vera Lúcia Santiago Araújo. O projeto CORSEL se dedica a segmentação em LSE colaborando com a construção de um modelo de legendagem para surdos e ensurdidos. A metodologia envolve uma dimensão descritiva através de análises quanti-qualitativas baseadas em *corpus*, viabilizada através do programa *WordSmith Tools 5.0* (SCOTT,2012), precisamente, as ferramentas *Wordlist* e *Concord*. As etiquetas desenvolvidas por Chaves (2012) e Assis (2013) permitiram as análises eletrônicas do *corpus*, identificando as ocorrências de problemas de segmentação em valores absolutos. Ainda, por se tratar de diferentes *corpora*, não podemos usar os dados absolutos para comparar os problemas de segmentação linguística. Para esse propósito, usa-se o Índice de Frequência Simples (IFS), que é um recurso estatístico adotado para o controle de variáveis, proposto por Praxedes Filho e Magalhães (2013) e Praxedes Filho e Silva (2014). Assim, com o seu uso, as médias de PROSEGL em valores de IFS passam a ser aproximadas, possibilitando a comparação entre as mesmas. Os resultados apresentam 242 ocorrências de PROSEGL, representando um Índice de Frequência Simples (IFS) de 19,4%. Os problemas ocorreram com maior frequência nos tipos de sintagma verbal e sintagma nominal em legendas de três linhas e com alta velocidade. Esses resultados indicam que há um número considerável de PROSEGL nas

LSEs, (19,4), *closed caption-pop-on* da telenovela Cheias de Charme. Vale ressaltar que essas evidências são reforçadas ao comparar esses resultados com os obtidos nas demais pesquisas desenvolvidas no âmbito do projeto CORSEL envolvendo o gênero das telenovelas, pois os resultados desta pesquisa foram semelhantes aos resultados obtidos concluindo que a maior ocorrência de PROSEGL no gênero das telenovelas também ocorreu nos tipos de sintagma verbal e sintagma nominal.

Palavras-chave: Tradução Audiovisual Acessível. Legendagem para surdos e ensurdecidos. Linguística de *Corpus*. Problemas de Segmentação. Telenovela.

ABSTRACT

This work deals with one of the branches of Accessible Audiovisual Translation, which promotes a sociocultural form of accessibility to people with hearing impairment: Subtitling for the deaf and the hard-of-hearing (SDH). Its main goal was to identify the linguistic segmentation problems in the SDH of the Brazilian Soap Opera ‘Cheias de Charme’ via Corpus Linguistics, and having as theoretical foundation the Translation Studies, more specifically, Audiovisual Translation (AVT). The linguistic segmentation problems are inappropriate divisions in one subtitle and they can occur between lines and, or between subtitles. They also can occur in levels of phrases: noun phrases, verb phrases, adjectival phrases, adverbial phrases, prepositional phrases; coordinate and subordinate clauses. To reach its goal, 1725 SDH were analyzed. They belong to five chapters of ‘Cheias de Charme’, broadcasted by the Brazilia TV channel Rede Globo de Televisão. This *corpus* is a *subcorpus* of the project entitled The segmentation in the Subtitling for the deaf and the hard-of-hearing (SDH): a *corpus* based study (CORSEL Project), developed by the Subtitling and Audiodescription group (LEAD) in the Ceara State University and coordinated by the Professor Dra. Vera Lúcia Santiago Araújo. The CORSEL Project is dedicated to the segmentation in SDH collaborating with the building of a subtitling model to the deaf and hard-of-hearings. The methodology involved a descriptive dimension through *corpus*-based quanti-qualitative analysis with the help of the software *WordSmith Tools 5.0*, specifically, the *Wordlist* and *Concord* tools. The tags developed by Chaves (2012) and Assis (2013), allowed the electronic analysis of the *corpus*, identifying the occurrences of linguistic segmentation problems in relative values. Therefore, for dealing with different corpora, we cannot use the absolute values to compare the linguistic segmentation problems. To this purpose we use the simple frequency index (IFS), which is a statistical feature to control variables, Proposed by Praxedes Filho e Magalhães (2013) e Praxedes Filho e Silva (2014). Therefore, with the use of it, the average of the occurrences of linguistic segmentation problems turn to be approximated, enabling the comparison among them. The results presented 242 linguistic segmentation problems, representing a simple frequency index (IFS) of 19, 4. The problems occurred most frequently at the noun phrase and the verb phrase; in 3-lined subtitles at high speed. These results indicate a relevant number of segmentation problems at the SDHs, (19, 4), closed caption pop-on, of ‘Cheias de Charme’. It is noteworthy that these evidences are strengthened when one compares the results with those of the remaining results for the genre Soap Operas investigated by the CORSEL project, concluding

that the highest occurrence of linguistics segmentation problems in the soap opera genre happens also at noun and verb phrases.

Keywords: Accessible Audiovisual Translation. Subtitling for the deaf and the hard-of-hearing. *Corpus* Linguistics. Segmentation Problems. Soap opera.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Árvore sintática usada por Karamitroglou (1998) para explicar a segmentação linguística.....	31
Figura 2 –	Imagens da vinheta de abertura da telenovela ‘Cheia de Charme’ e de suas protagonistas.....	50
Figura 3 –	CCExtractor.....	50
Figura 4 –	<i>Subtitle Workshop</i>	51
Figura 5 –	Arquivo de legenda em formato srt convertido para txt - <i>corpus</i> anotado.....	52
Figura 6 –	Tela do <i>Wordlist</i>	56
Figura 7 –	Tela com resultados da ferramenta <i>Concord</i>	56
Figura 8 –	Imagem com exemplo de PROSEGL do SV_composto.....	60
Figura 9 –	Imagem com exemplo de PROSEGL do SV_verbo+SAdv.....	62
Figura 10 –	Imagem com exemplo de PROSEGL do SN_especif+SN.....	66
Figura 11 –	Imagem com exemplo de PROSEGL do SN_subst+SP.....	67
Figura 12 –	Imagem com exemplo de PROSEGL do SN_subst+SAdj.....	69

LISTA DE TABELAS

Tabela 1 –	Relação de caracteres por segundos para velocidade de 145 ppm.....	26
Tabela 2 –	Relação de caracteres por segundos para velocidade de 160 ppm.....	26
Tabela 3 –	Relação de caracteres por segundos para velocidade de 180 ppm.....	27
Tabela 4 –	IFS e VA dos PROSEGL encontrados no gênero das telenovelas.....	47
Tabela 5 –	VA e IFS dos PROSEGL em ‘Cheias de Charme’	58
Tabela 6 –	VA dos tipos de PROSEGL em ‘Cheias de Charme’	59
Tabela 7 –	Relação entre PROSEGL e parâmetros técnicos em ‘Cheia de Charme em VA.....	75
Tabela 8 –	PROSEGL do SV (verbo)+oblíquo+SV.....	76
Tabela 9 –	Relação entre PROSEGL e parâmetros técnicos em ‘Cheia de Charme em VA.....	77
Tabela 10 –	Relação entre PROSEGL e parâmetros técnicos do gênero telenovela.....	78

LISTA DE QUADROS

Quadro 1 –	Formato das legendas na tela.....	32
Quadro 2 –	Legendas com problemas de segmentação linguística.....	33
Quadro 3 –	Legendas com problemas de segmentação linguística.....	43
Quadro 4 –	Quadro sinótico de etiquetas do projeto CORSEL relativas aos parâmetros técnicos da legendagem.....	53
Quadro 5 –	Quadro sinótico de etiquetas do projeto CORSEL relativa aos problemas de segmentação linguística.....	54
Quadro 6 –	PROSEGL do SV_composto.....	60
Quadro 7 –	PROSEGL do SV_verbo+SAdv.....	62
Quadro 8 –	PROSEGL do SV (verbo)+oblíquo+SV.....	63
Quadro 9 –	PROSEGL do SV_verbo+SP.....	64
Quadro 10 –	PROSEGL do SN_especif+SN.....	66
Quadro 11 –	PROSEGL do SN_subst+SP.....	68
Quadro 12 –	PROSEGL do SN_subst+SAdj.....	69
Quadro 13 –	PROSEGL do SN_subst+núcleo.....	70
Quadro 14 –	PROSEGL do SAdj.....	71
Quadro 15 –	PROSEGL do SAdj.....	72
Quadro 16 –	PROSEGL do SP.....	72
Quadro 17 –	PROSEGL de orações coordenadas.....	73
Quadro 18 –	PROSEGL de orações subordinadas.....	74

LISTA DE ABREVIATURAS E SIGLAS

CORSEL	<i>Corpus</i> , Segmentação e Legendagem
ADV	Advérbio
BBC	British Broadcast Television
CAS	Centro de Atendimento ao Surdo
Cpl	Caracteres por linha
Cps	Caracteres por segundo
COORD	Coordenada
ETBC	Estudos de Tradução Baseados em <i>Corpus</i>
IFS	Índice de Frequência Simples
L	Linha
LATAV	Laboratório de Tradução Audiovisual
LC	Linguística de <i>Corpus</i>
LEAD	Legendagem e Audiodescrição
LSE	Legendagem para Surdos e Ensurdecidos
MOLES	Modelo de legendagem Surdos e Ensurdecidos
NCC	Novela Cheia de Charme
PosLA	Programa de Mestrado em Linguística Aplicada
PB	Português Brasileiro
PPM	Palavras por minuto
PROSEGG	Problema de Segmentação Gramatical
PROSEGL	Problema de Segmentação Linguística
PREP	Preposição
SAdj	Sintagma Adjetival
SAdv	Sintagma Adverbial
SGML	Standard Mark-up Language
SL	Segmentação Linguística
SN	Sintagma Nominal
SP	Sintagma Preposicionado
SV	Sintagma Verbal
Sub	Legenda (<i>subtitle</i>)
SUBORD	Subordinada

T	Tempo
TAV	Tradução audiovisual
TAVa	Tradução audiovisual acessível
UECE	Universidade Estadual do Ceará

SUMÁRIO

1	INTRODUÇÃO.....	18
2	APORTE TEÓRICO.....	24
2.1	TRADUÇÃO AUDIOVISUAL: A LEGENDAGEM E SEUS PARÂMETROS.....	24
2.2	LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE).....	29
2.3	SEGMENTAÇÃO LINGUÍSTICA.....	30
2.4	A DESCRIÇÃO DO PORTUGUÊS SEGUNDO CASTILHO.....	33
2.4.1	A Oração.....	34
2.4.1.1	Oração coordenada.....	34
2.4.1.2	Oração subordinada.....	35
2.4.2	Os sintagmas.....	36
2.4.2.1	Sintagma verbal.....	37
2.4.2.2	Sintagma nominal.....	38
2.4.2.3	Sintagma adjetival.....	40
2.4.2.4	Sintagma adverbial.....	41
2.4.2.5	Sintagma preposicional.....	42
2.5	LEGENDAGEM E LINGUÍSTICA DE <i>CORPUS</i>	43
3	METODOLOGIA.....	49
3.1	TIPO DA PESQUISA.....	49
3.2	CONSTITUIÇÃO DO <i>CORPUS</i>	49
3.3	PROCEDIMENTOS.....	50
3.3.1	Extração de legendas.....	50
3.3.2	Etiquetagem.....	52
3.4	ANÁLISE DOS DADOS.....	55
4	RESULTADOS E DISCUSSÃO DOS DADOS.....	58
4.1	RESULTADOS QUANTITATIVOS DE ‘CHEIAS DE CHARME’.....	58
4.2	TIPOS DE PROBLEMA DE SEGMENTAÇÃO LINGUÍSTICA (PROSEGL).....	59
4.2.1	PROSEGL no sintagma verbal.....	59
4.2.2	PROSEGL no sintagma nomina.....	66
4.2.3	PROSEGL no sintagma adjetival.....	71

4.2.4	PROSEGL no sintagma adverbial.....	71
4.2.5	PROSEGL no sintagma preposicional.....	72
4.2.6	Orações coordenadas.....	73
4.2.7	Orações subordinadas.....	74
4.3	RELAÇÃO ENTRE PROSEGL E PARÂMETROS TÉCNICOS.....	74
4.4	A SEGMENTAÇÃO LINGUÍSTICA NO GÊNERO TELENOVELA.....	75
4.4.1	Quantidade de problemas de segmentação.....	76
4.4.2	Tipos de problemas de segmentação.....	77
4.4.3	Relação entre problemas de segmentação e parâmetros técnicos.....	78
5	CONSIDERAÇÕES FINAIS.....	80
	REFERÊNCIAS.....	83
	ANEXOS.....	87
	ANEXO A – SINTÁGMA VERBAL COMPOSTO.....	88
	ANEXO B – SINTÁGMA VERBAL (VERBO + SINTÁGMA ADVERBIAL).....	91
	ANEXO C – SINTÁGMA VERBAL (VERBO + PRONOME OBLÍQUO + SINTÁGMA ADVERBIAL).....	93
	ANEXO D – SINTÁGMA VERBAL (VERBO + SINTÁGMA PREPOSICIONAL).....	94
	ANEXO E – SINTAGMA NOMINAL (ESPECIFICADOR + SINTÁGMA NOMINAL).....	95
	ANEXO F – SINTAGMA NOMINAL (SUBSTANTIVO + SINTÁGMA PREPOSICIONAL).....	97
	ANEXO G – SINTAGMA NOMINAL (SUBSTANTIVO + SINTÁGMA ADJETIVAL).....	99
	ANEXO H – SINTAGMA NOMINAL (NÚCLEO).....	101
	ANEXO I – SINTAGMA ADJETIVAL.....	102
	ANEXO J – SINTAGMA ADVERBIAL.....	103
	ANEXO K – SINTAGMA PREPOSICIONAL.....	104
	ANEXO L – ORAÇÕES COORDENADAS.....	106
	ANEXO M – ORAÇÕES SUBORDINADAS.....	107

1 INTRODUÇÃO

No Brasil, as pesquisas em tradução audiovisual (TAV), mais especificamente, legendagem para surdos e ensurdecidos (LSE) e audiodescrição (AD), apesar de recentes, vêm ganhando força e se destacando cada vez mais no cenário acadêmico. Essa realidade é uma tentativa de contribuir com a urgente necessidade de políticas públicas inclusivas, voltadas à promoção da acessibilidade de pessoas surdas e ensurdecidas e de pessoas com deficiência visual, respectivamente, a produtos (áudio) visuais, com o intuito de fazer valer o seu direito de apreender seu entorno *in totum*, que é o direito de todo cidadão brasileiro.

Para atender a essa necessidade, é relevante reconhecer que a LSE é um recurso de acessibilidade que difere de outra modalidade de TAV muito próxima, qual seja, a legenda para ouvintes. Portanto, a LSE precisa de parâmetros que atendam às necessidades do público para o qual ela é feita.

Atualmente, os canais de televisão brasileiros utilizam o sistema de legendagem norte-americano, *closed caption*, que é um tipo de legendagem fechada. Essas legendas, em sua maioria, constituem-se em transcrições diretas da fala e não se adequam às diretrizes próprias da legendagem, como número de linhas, tempo máximo de permanência da legenda na tela, velocidade da legenda etc. A ausência de adequação a esses parâmetros pode dificultar a compreensão das legendas pelos surdos (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004).

Ao me defrontar com estas questões, percebi uma relevância mais abrangente acerca da acessibilidade, promovida por meio da lei 10.098, Lei da Acessibilidade¹, que estabelece normas gerais e critérios básicos para assegurar o direito de acesso à informação, à comunicação, ao trabalho, à educação, ao transporte, à cultura, ao esporte e ao lazer às pessoas com deficiência. A partir dessa lei, esses direitos vão sendo evidenciados à medida que aumentam os meios de se promover esse acesso às pessoas surdas e ensurdecidas por intermédio da interpretação em LIBRAS e da LSE.

Assim, foi crescente, numa espectadora não-leiga como eu, o interesse por LSE, à medida em que nela se percebia nuances que pareciam dificultar a compreensão de quem as lia, pois, uma segmentação aparentemente sem critérios me ocasionava uma certa incapacidade cognitiva para acompanhar o fluxo da informação. Isso despertou em mim o

¹ http://www.planalto.gov.br/civil_03/leis/110098.htm.

desejo de investigar na LSE como as falas de uma produção audiovisual são segmentadas em forma de legenda.

Várias pesquisas desenvolvidas pelo grupo de pesquisas Legendagem e Audiodescrição (LEAD) – do Laboratório de Tradução Audiovisual (LATAV) do Programa de Pós-Graduação em Linguística Aplicada (PosLA) do Centro de Humanidades (CH) da Universidade Estadual do Ceará (UECE), doravante LEAD-LATAV, coordenado pela Profa. Dra. Vera Lucia Santiago Araújo – testaram a recepção da LSE no Brasil. Dentre estas, destacamos o projeto POR UM MODELO DE LEGENDAGEM PARA SURDOS NO BRASIL (projeto MOLES), que teve como objetivo propor um modelo de LSE com parâmetros que atendessem à comunidade de surdos e ensurdecidos brasileiros (ARAÚJO; NASCIMENTO, 2011). Essa pesquisa, que coletou dados em todas as regiões do país, testou hipóteses relacionadas à velocidade da legenda em números de palavras por minuto (ppm) com base nos padrões indicativos de três tipos de velocidade que podem proporcionar uma boa recepção para espectadores de filmes legendados: 145 ppm, 160 ppm e de 180 ppm. (D'YDEWALLE *et al.*, 1987; DIAZ-CINTAS; REMAEL, 2007).

Apesar de a hipótese levantada inicialmente ser de que a velocidade de 160 ppm seria a preferida pelos surdos e ensurdecidos brasileiros, os resultados revelaram que a recepção das legendas pelos participantes foi eficiente independentemente da velocidade da legenda. Uma interpretação possível para esses resultados seria a boa segmentação linguística das legendas apresentadas (ARAÚJO; NASCIMENTO, 2011). Os resultados do projeto MOLES foram fundamentais para o surgimento de novas perspectivas sobre LSE no LEAD-LATAV, pois, a partir disso, a segmentação passou a receber maior atenção, suscitando novas pesquisas e projetos, como: Segmentação na legendagem para surdos e ensurdecidos – LSE: um estudo baseado em *corpus* (CORSEL).

O projeto CORSEL, no qual minha pesquisa está inserida, pretende disponibilizar um *corpus* que possa ser acessado por profissionais e pesquisadores da área, com o fim de promover uma análise mais detalhada do parâmetro da segmentação e, assim, colaborar com a construção de um modelo de legendagem para surdos e ensurdecidos que promova, de fato, a acessibilidade audiovisual. O *corpus* selecionado é composto de gravações de programas exibidos durante duas semanas na programação da Rede Globo, sendo constituído por gêneros como telenovelas, filmes e seriados, que disponibilizam a legenda do tipo *closed caption*.

Vale ressaltar que a segmentação é um parâmetro da legendagem relacionado à distribuição do texto escrito dentro de uma legenda, ou entre duas ou mais legendas e que também é responsável pelo fluxo das informações (linguísticas, retóricas e visuais) exibidas

na tela dos produtos audiovisuais. Perego (2008, p. 35) sugere que a segmentação linguística deva ser criteriosa, seguindo as principais regras de sintaxe e privilegiando a não separação dos constituintes estruturais dos sintagmas – verbais, nominais, adjetivais, adverbiais ou preposicionais – e das orações coordenadas e subordinadas. A mesma autora acredita que as respostas das pessoas às legendas podem ser afetadas positivamente, havendo um menor esforço quanto ao aspecto cognitivo durante a leitura de legendas segmentadas criteriosamente.

Dentro do que é sugerido por Perego (2008), desenvolvi como monografia de especialização, o estudo *A Segmentação na Legendagem para Surdos e Ensurdecidos (LSE): uma análise baseada em corpus* (GABRIEL, 2013). O *corpus* analisado constituiu-se de um único capítulo da telenovela ‘Cheias de Charme’ e os resultados indicaram várias inadequações de segmentação linguística, além de outras, como questões técnicas de velocidade, número de linhas e caracteres. Senti, então, a curiosidade de saber se as legendas de outros capítulos apresentavam as mesmas inadequações, sendo este o problema investigado nesta pesquisa. Para tal, o *corpus* passou a ser uma expansão daquele que analisei em Gabriel (2013).

A partir do problema proposto, esta pesquisa procurou responder às seguintes perguntas:

- 1) Qual a quantidade de problemas de segmentação linguística (PROSEGL) nas legendas para surdos e ensurdecidos (LSEs) da telenovela brasileira ‘Cheias de Charme’?
- 2) Quais os tipos de problemas de segmentação linguística nessas LSEs?
- 3) Qual a relação entre os PROSEGL e os parâmetros técnicos de legendagem?
- 4) Qual a relação entre a ocorrência dos PROSEGL encontrados na LSE de ‘Cheias de Charme’ e a ocorrência de PROSEGL encontrada nos subcorpora do CORSEL para o gênero telenovela?

A partir das questões mencionadas na problematização, pude definir o objetivo geral deste estudo: analisar a segmentação linguística da LSE da telenovela ‘Cheias de Charme’ utilizando o suporte metodológico da Linguística de *Corpus*. Pude, também, definir seus objetivos específicos:

- 1) Identificar a ocorrência quantitativa dos problemas de segmentação linguística (PROSEGL) na LSE da telenovela;
- 2) Analisar as características dos PROSEGL na LSE da telenovela;
- 3) Verificar a relação entre a ocorrência dos casos de PROSEGL encontrados na LSE da telenovela e os parâmetros técnicos de legendagem;
- 4) Comparar a ocorrência dos PROSEGL encontrados em ‘Cheias de Charme’ e a ocorrência de PROSEGL encontrada nos subcorpora do CORSEL para o gênero telenovela.

Atualmente, as pesquisas sobre a segmentação linguística em LSE despontam como um novo horizonte na construção de parâmetros para a LSE, a partir de importantes trabalhos como os de Perego (2003; 2008; 2009), Kalantzi (2008), Chaves (2009; 2012), Diniz (2012), Assis (2013), Gabriel (2013). Perego (2003) descreveu a explicitação na legendagem e propôs uma categorização para os casos de explicitação em legendagem intralinguística. Kalantzi (2008) compilou um *corpus* de legendas para surdos e ensurdecidos e analisou a segmentação e a redução das legendas. Ambas as autoras viabilizaram suas pesquisas através de análise manual dos *corpora*. Kalantzi (2008) utilizou um *corpus* composto de 360 legendas para surdos de três programas da BBC, dois documentários e um programa de atualidades; e analisou as características da segmentação, redução e os parâmetros técnicos dessas legendas. Assim, em minha pesquisa: (i) analisei os parâmetros técnicos da legendagem; (ii) verifiquei a relação entre os problemas de segmentação linguística identificados no *corpus* e os parâmetros técnicos de legendagem; e (iii) comparei os resultados obtidos com outros resultados obtidos no projeto CORSEL para o gênero de telenovelas.

Ademais, outras pesquisas em LSE têm sido inovadas com a metodologia da Linguística de *Corpus*, trazendo vantagens para a área e possibilitando a análise eletrônica de grande quantidade de dados. Algumas dessas pesquisas utilizaram o programa de análise linguística *WordSmith Tools*. Chaves (2009) teve como objetivo verificar a viabilidade de análise descritiva automática de *corpora*, tendo em vista o estabelecimento de parâmetros para a elaboração de LSE em trechos de programas de gêneros variados da Rede Globo de Televisão. Chaves (2012) analisou eletronicamente as LSEs do filme em DVD ‘Nosso Lar’ (2010), com o objetivo de estabelecer parâmetros para a análise da segmentação linguística em LSE, e propôs as primeiras etiquetas para a análise eletrônica da segmentação em legendagem. Ainda, os resultados da autora sugeriram que os problemas de segmentação

estavam concentrados em sua maioria nos níveis dos sintagmas, sobressaindo-se o sintagma verbal, cuja segmentação ocorreu principalmente entre os constituintes verbo + verbo.

Diniz (2012) comparou as categorias da segmentação linguísticas entre linhas nas legendagens para ouvintes e para surdos e ensurdecidos produzidas em português brasileiro, do filme ‘Irmãos de Fé’. Os resultados dessa pesquisa revelaram a preferência do legendista pelo tipo de legendagem feita, em sua maioria, por transcrições da fala. Além disso, assim como Chaves (2012), revelou que os problemas de segmentação linguística ocorreram em sua maioria entre os constituintes do sintagma verbal.

Assis (2013) analisou a segmentação das legendas da telenovela brasileira ‘Amor Eterno Amor’ e aperfeiçoou o quadro de etiquetas para a identificação dos problemas de segmentação linguística com subsídios do trabalho de Chaves (2012) e das considerações funcionalistas da gramática de Castilho (2012) sobre os sintagmas no português brasileiro falado. Os resultados de Assis (2013) revelaram uma grande quantidade de PROSEGL no *corpus*, sendo a maioria das ocorrências no sintagma verbal (50,6%) e sintagma nominal (32,1%) em legendas de 3 linhas e com velocidade alta.

Utilizando as etiquetas propostas por Chaves (2012) e Assis (2013), a presente autora também investigou problemas de segmentação recorrentes nas LSEs. Gabriel (2013) identificou e descreveu os problemas de segmentação verificando a ocorrência dos principais problemas de segmentação linguística e analisando a relação dessas ocorrências com a velocidade, o número de caracteres e o número de linhas das legendas. Os resultados mostraram que os problemas de segmentação linguística ocorreram com maior frequência nos níveis do sintagma verbal (45%) e do sintagma nominal (31,6%), majoritariamente em legendas de três linhas e com velocidade alta.

Diante desses resultados, esta pesquisa teve o intuito de contribuir com os estudos em LSE em vários aspectos. Em primeiro lugar, buscando resultados mais categóricos acerca dos tipos de PROSEGL mais recorrentes nas LSEs de ‘Cheias de Charme’, através da análise de um *corpus* cinco vezes maior que aquele usado em Gabriel (2013). Em segundo lugar, proporcionando uma descrição mais detalhada das quebras dos sintagmas e das orações coordenadas e subordinadas, em que foram especificados seus constituintes, explicando como ocorreu o parâmetro da segmentação. Em terceiro lugar, comparando os casos de PROSEGL com as características dos parâmetros técnicos da legendagem dessas LSEs, de modo a reconhecer essas características e saber as razões plausíveis para as quebras de linhas. Além disso, também contribuiu para os resultados e as conclusões do projeto CORSEL.

E, por fim, como já mencionado, essa pesquisa também se justifica por levantar e discutir questões sociais, como a inclusão social e acessibilidade aos meios audiovisuais.

Esta dissertação está organizada em cinco capítulos, incluindo este, que constitui o primeiro capítulo, seguido do Aporte Teórico, da Metodologia, da Análise dos dados e das Considerações finais. O segundo capítulo traz as principais considerações sobre tradução audiovisual, legendagem para surdos e ensurdecidos – LSE, particularmente sobre segmentação linguística. O terceiro capítulo descreve o *corpus*, os procedimentos metodológicos para a realização desta pesquisa, desde a extração e etiquetagem até a análise eletrônica que gerou os dados para as discussões e análises dos resultados finais. O quarto capítulo apresenta os resultados encontrados e traz as análises detalhadas das segmentações linguísticas encontradas no *corpus* em estudo. Finalmente, o último capítulo apresenta as considerações finais sobre esta pesquisa, enfatizando os resultados e as conclusões sobre a análise, além de sugerir expectativas futuras para o desenvolvimento de outras pesquisas que possam também contribuir com os estudos em LSE.

2 APORTE TEÓRICO

Este capítulo inicia destacando a localização da pesquisa no campo disciplinar da Tradução Audiovisual. Em seguida, descreve as peculiaridades da legendagem para ouvintes, nos seus parâmetros técnicos e linguísticos, bem como apresenta a legendagem para surdos e ensurdecidos (LSE). Depois, discute as considerações gramaticais de Castilho (2012) sobre os sintagmas e as orações. Finalmente apresenta a importância dos Estudos Descritivos da Tradução e o papel da Linguística de *Corpus*.

2.1 TRADUÇÃO AUDIOVISUAL: A LEGENDAGEM E SEUS PARÂMETROS

A inclusão da LSE nos estudos da tradução, mais especificamente nos estudos em tradução audiovisual (TAV), tornou-se possível a partir da classificação dos tipos de tradução proposta por Jakobson (1995): tradução intralinguística ou reformulação, tradução interlinguística ou tradução propriamente dita e tradução intersemiótica ou transmutação. A primeira, tradução intralinguística, é uma interpretação de signos verbais mediante outros signos verbais do mesmo idioma. A segunda, tradução interlinguística ou tradução propriamente dita, é uma interpretação de signos verbais mediante outro idioma. A terceira, tradução intersemiótica ou transmutação, é uma interpretação de signos verbais mediante signos de sistemas de signos não verbais.

A classificação proposta por Jakobson (1995) permitiu que os estudos da tradução pudessem se expandir e deu início à construção de novas teorias e definições sobre a TAV, como a que temos a seguir:

[...] TAV [tradução audiovisual] era usado para encapsular diferentes práticas tradutórias usadas nos meios audiovisuais [...] nas quais há transferência de uma língua fonte para uma [língua] alvo, que envolve algum tipo de interação com som e imagens. (DÍAZ-CINTAS; REMAEL, 2007, p. 12, tradução nossa)²

Uma das principais práticas tradutórias da TAV é a legendagem. Segundo Franco e Araújo (2011, p. 5) uma das primeiras definições propostas acerca da legendagem é que se trata apenas de uma tradução interlinguística (ALVARENGA, 1998). Contudo, essa definição não alcança toda a extensão dos estudos e das teorias citadas. A seguir, encontramos uma definição mais abrangente sobre legendagem:

² “[...] AVT was used to encapsulate different translation practices used in the audiovisual media [...] in which there is a transfer from a source to a target language, which involves some form of interaction with sound and images.”

[...] uma prática da tradução que consiste na apresentação de um texto escrito, geralmente na parte inferior da tela, que tenta narrar o diálogo original do falante, bem como os elementos discursivos que aparecem na imagem (letras, inserções, grafite, inscrições, placares, etc.) e as informações de efeitos sonoros (DIAZ-CINTAS; REMAEL, 2007, p. 8, tradução nossa)³.

Atualmente, as legendas para ouvintes são feitas com o uso de um *software* que permite a marcação de seu início e término. Além disso, essa tecnologia permite a tradução, revisão e pré-visualização. Nesse processo de legendagem, os profissionais envolvidos são o legendador e o legendista. O legendador é o técnico que grava as legendas no vídeo, já o legendista faz a tradução. O termo “legendista”, proposto por Alvarenga (1998), justifica-se porque é o tradutor o protagonista da tradução. Além disso, o processo de legendagem para ouvintes pode se diferenciar entre legendação, o trabalho de tradução, e legendagem, o processo completo até a gravação. Ainda, é possível fazer todo o processo de legendagem com o auxílio de *softwares* livres, sendo desnecessária essa diferenciação. No processo de tradução, as legendas para ouvintes são condensadas para que haja uma sincronização das legendas, falas e imagens (ARAÚJO; ASSIS, 2014; FRANCO; ARAÚJO, 2011; ARAÚJO; NASCIMENTO, 2011).

Segundo Diaz-Cintas e Remael (2007, p. 17), na legendagem é possível estabelecer diferentes parâmetros, dos quais se destacam os técnicos e os linguísticos.

Os parâmetros técnicos são classificados em duas dimensões, espacial e temporal. A dimensão espacial abrange as características relacionadas ao número de linhas, disposição e aparência das legendas na tela, tamanho da fonte e número de caracteres por linha; enquanto a dimensão temporal está relacionada ao tempo de marcação e duração das legendas, sincronização, intervalo, velocidade de leitura e tempo de exposição das legendas. As legendas para ouvintes apresentam-se preferencialmente em no máximo duas linhas, são em geral de cor branca ou amarela e podem ocorrer em três velocidades de 145, 160 e 180 palavras por minuto (ppm) (D'YDEWALLE *et al.*, 1987; IVARSSON; CARROL, 1998; DIAZ CINTAS; REMAEL, 2007). Diaz-Cintas e Remael (2007) propuseram três tabelas que descrevem o número de caracteres por segundo permitidos para a legendagem. Pelas Tabelas 1, 2 e 3, podemos conhecer esses números quando se tratar de uma velocidade 145 ppm, 160 ppm e 180 ppm:

³ “ [...] a translation practice that consists of presenting a written text, generally on the lower part of the screen, that endeavours to recount the original dialogue of the speakers, as well as the discursive elements that appear in the image (letters, inserts, graffiti, inscriptions, placards, and the like), and the information that is contained on the soundtrack (songs, voices off).”

Tabela 1 – Relação de caracteres por segundos para velocidade de 145 ppm

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

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

Tabela 2 – Relação de caracteres por segundos para velocidade de 160 ppm

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

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

Tabela 3 – Relação de caracteres por segundos para velocidade de 180 ppm

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

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

As tabelas expostas mostram os espaços permitidos para que se editem legendas de 145, 160 e 180 ppm. Os números antes de dois pontos correspondem aos segundos e os números posteriores, aos *frames* ou quadros⁴. Ao observarmos a tabela de 145 ppm, podemos compreender que em uma fala que dure um segundo e vinte *frames* (ou quadros), é possível inserir 25 caracteres de legenda. Já para a tabela de 160 ppm e de 180 ppm, a mesma fala de um segundo e vinte *frames* ou quadros é possível inserir 28 e 30 caracteres de legenda, respectivamente. Caso o legendista ultrapasse esses números, as legendas ficarão muito rápidas, e talvez não proporcionem uma boa recepção por parte do espectador. As tabelas expostas acima são um instrumental fundamental para as análises dos parâmetros técnicos das legendas do meu *corpus* de estudo e de outros trabalhos acerca da LSE (DIAZ CINTAS; REMAEL, 2007).

Abaixo temos um exemplo de legenda que apresentou 59 caracteres distribuídos em 3,7 segundos, resultando numa legenda de velocidade média. De acordo com a tabela acima proposta para 160 ppm, em 3,7s poderíamos inserir até 60 caracteres para 3:20 (3 segundos e 20 frames), indicando a velocidade de 160 ppm (GABRIEL, 2013, p. 46).

**Bem diferente daquele buraco negro
que eu vivia no Sarmento.**

⁴ É uma unidade de medida usada para determinar o número de imagens de uma produção audiovisual, processa ou exibe por unidade de tempo. Também são chamados de quadros. No cinema, cada segundo de exibição é composto por 24 imagens ou quadros enquanto na TV e no vídeo temos 25 quadros por segundo (CINTAZ E REMAEL, 2007, p. 70).

Sobre o tempo de exposição de uma legenda na tela, na Europa o tempo máximo de permanência é de seis segundos. No Brasil, esse tempo máximo é de quatro segundos (ARAÚJO, 2004). Por outro lado, a duração mínima de exposição de uma legenda para que o espectador tenha tempo de ler a legenda e olhar para a imagem antes de ambas desaparecerem é de um segundo. Quando a velocidade da fala ultrapassa qualquer uma das três velocidades apontadas por Diaz-Cintas e Remael (2007), o texto da legenda precisa ser editado. Essa edição é um processo de tradução, viabilizado pelos parâmetros linguísticos da legendagem. Esses parâmetros ocupam pertinência para este estudo, pois englobam dois tipos principais, a redução textual e a segmentação; sendo o segundo tipo o foco principal do meu trabalho. A redução textual pode ocorrer por condensação ou reformulação, omissões ou exclusões. A segmentação pode ocorrer por critérios linguísticos, retóricos e visuais⁵ (DIAZ-CINTAS; REMAEL, 2007; REID, 1990; GOTTLIEB, 1994; KARAMITROGLOU, 1998; IVARSSON; CARROL, 1998).

Durante a segmentação de uma legenda, um tradutor pode usar os parâmetros da redução para obter um bom resultado nos casos em que haja necessidade de quebras de linhas ou separações entre legendas. Assim, esse recurso, que pode ser atrelado a outros, pode ser usado pelos legendistas durante o processo de tradução, como no exemplo da legenda abaixo, que apresentou problema de segmentação linguística de quebra no sintagma preposicional.

**Entrou com um processo lá na
empresa em que ele trabalhava,
ganhou uma grana ferrada.**⁶

Para compreendermos como ocorreu a redução da legenda, apresento a mesma legenda condensada através da omissão do advérbio “lá”, que causava redundância. Ainda, com a omissão do pronome “ele”, que pode ser oculto sem que houvesse problemas para a compreensão da legenda, já que o mesmo é facilmente identificado pela desinência número pessoal do verbo “entrou”, por sua terminação e pela semântica (ARAÚJO; ASSIS, 2014, p.176):

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

⁵ O parâmetro da segmentação e seus critérios serão abordados mais adiante na sessão 2.3.

⁶ Retirado de Araújo e Assis (2014, p.176).

Aqui não iremos nos aprofundar em todas as formas de segmentação, mas na segmentação linguística, por ser o foco principal desta dissertação. Essa discussão será realizada na próxima seção.

2.2 LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE)

Além da legendagem para ouvintes que segue os parâmetros técnicos e linguísticos tratados anteriormente, temos a legendagem para surdos e ensurdecidos (LSE), que é a tradução do inglês SDH (*Subtitling for the Deaf and the Hard-of-Hearing*). Dentre os tipos de LSE, temos as legendas do tipo *closed caption*, que não seguem os mesmos parâmetros de legendagem para ouvinte. É um modelo norte-americano de legendagem, um sistema tecnológico que utiliza números ou códigos numéricos para transcrever as falas em legendas que podem ser visualizadas pelo telespectador. Para a edição do *closed caption*, a linha de legenda precisa ser transformada em códigos eletrônicos e inserida na linha 21 do intervalo vertical em branco da TV, que é uma barra horizontal posta entre as imagens. São produzidas por um estenotipista, profissional que opera um teclado especial, o estenótipo⁷. Ainda, essas legendas são visíveis por meio de um decodificador disponível no controle remoto da TV (ARAÚJO; FRANCO; ALVARENGA, 2002; ARAÚJO; NASCIMENTO, 2011, ARAÚJO; ASSIS, 2014).

O sistema de LSE do tipo *closed caption* se apresenta de duas formas distintas: *roll-up e pop-on*. A legenda *roll-up* é um tipo rotativo, cujas linhas sobem (ou rolam) continuamente da parte inferior da tela da TV, em um máximo de quatro linhas por vez. As palavras que compõem essas legendas são exibidas da esquerda para a direita. Geralmente, é o tipo usado para legendagem ao vivo. A legenda *pop-on* é aquela cuja informação surge como um todo, exibida em um único bloco, ficando temporariamente na tela, na maioria das vezes em sincronismo com o áudio, desaparecendo em seguida. É o tipo de legenda usado em programas pré-gravados, como as telenovelas, os filmes e as minisséries exibidas nos canais de TV brasileiros (ARAÚJO; FRANCO; ALVARENGA, 2002).

Ademais, temos o tipo de LSE proposto pelas pesquisas do grupo LEAD, essas legendas diferem totalmente das mencionadas anteriormente, assemelhando-se mais às legendas para ouvintes, mas com o acréscimo de algumas informações: identificação dos

⁷ O estenótipo é um equipamento normalmente usado para transcrever pautas de reuniões e de assembleias do Congresso Nacional, de empresas e de sessões de tribunais. Atualmente, também, está sendo usado agora para legendar programas de TV pré-gravados e ao vivo.

falantes e efeitos sonoros (ARAÚJO, 2004; 2005; FRANCO; ARAÚJO, 2003). A identificação dos falantes é necessária, pois muitas vezes a mudança de fala de um sujeito para outro não é percebida pelos surdos. Isso se agrava se houver dois ou mais personagens em cena, comprometendo a compreensão da obra audiovisual (ARAÚJO, 2004; 2008).

Atualmente, há várias formas de inserir as informações adicionais. Na Europa, ela é feita através das cores: por exemplo, na Espanha, a identificação de personagens se dá por uma cor específica para cada um; já em Portugal, é usada a cor amarela para a identificação dos falantes em cena e a cor branca para as falas. Também, as legendas são posicionadas sobre o falante indicando de quem é a vez. No Brasil, essas informações são acrescentadas entre colchetes.

O posicionamento de Nascimento (2013) sustenta que a identificação de efeitos sonoros é um parâmetro usado para preencher uma lacuna na LSE deixada pelos elementos acústicos não verbais presentes no meio audiovisual, quando esses mesmos elementos colaboram para a construção de sentido. Essas legendas se apresentam entre colchetes e contribuem para a compreensão do enredo. Por exemplo, o uso de legendas que descrevem a música de fosso: [música alegre], [música triste] e [música de suspense].

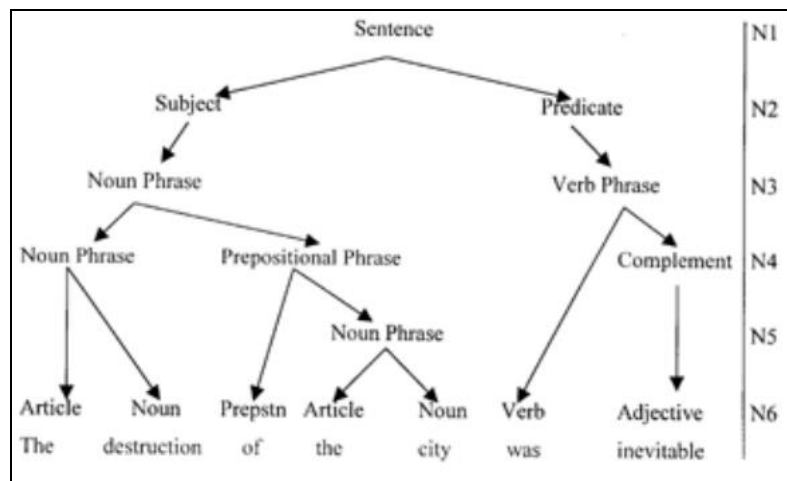
Além de apresentarem as características mencionadas anteriormente, a proposta de LSE do grupo LEAD chama a atenção para a forma como estas são bem segmentadas entre quebra de linhas, obedecendo a um *layout* simétrico, privilegiando a ausência de quebras de linhas entre os constituintes dos sintagmas e entre as orações.

2.3 SEGMENTAÇÃO LINGUÍSTICA

A segmentação em legendagem diz respeito à divisão da fala traduzida em porções de texto escrito, normalmente, na parte inferior da tela. A segmentação pode acontecer em dois níveis: quebra de linha (*line break*), divisão da legenda em duas ou mais linhas; e segmentação entre duas ou mais legendas, divisão das legendas ao longo da legendagem (DIAZ CINTAS; REMAEL, 2007). Como já foi dito, a segmentação pode ser feita seguindo três critérios: o visual, o retórico e o linguístico (REID, 1990). O visual segue as mudanças de tomada no filme, o retórico segue o fluxo da fala e o linguístico é determinado pelas regras da sintaxe, facilitando a leitura da legenda por parte do espectador. Neste trabalho, iremos considerar apenas os problemas de segmentação linguística entre linhas.

Os trechos resultantes da segmentação linguística devem apresentar uma estrutura sintática e semântica completa e possibilitar uma leitura rápida e eficiente (KARAMITROGLOU, 1998). Entretanto, isso nem sempre é possível, sendo necessário, então, que lhe seja dada uma maior atenção. Segundo Karamitroglou (1998), a divisão de uma legenda deve ser feita no mais alto nível sintático possível. A Figura 1 traz o exemplo de uma oração⁸ e os nódulos sintáticos nos quais ela pode ser subdividida para explicar como se faz a segmentação de uma legenda.

Figura 1 – Árvore sintática usada por Karamitroglou (1998) para explicar a segmentação linguística



Fonte: Karamitroglou (1998, p. 9).

A árvore sintática proposta pelo autor divide a oração em seis nódulos: no primeiro, N1, a oração aparece completa: *the destruction of the city was inevitable*. No segundo, N2, constituído de sujeito e de predicado, os quais são formados de sintagmas nominais e verbais respectivamente, alocados em N3. O sujeito *the destruction of the city* é constituído de dois sintagmas, um nominal *the destruction* e o outro preposicional *of the city*, enquanto o predicado tem um complemento, no patamar N4. O sintagma preposicional

⁸ Araújo e Assis (2014, p. 165) dizem o seguinte sobre a distinção entre sentença e oração: “Segundo o autor [Castilho], a “designação da sentença não é pacífica na literatura” ([2012,] p. 58), ao passo que podemos encontrar “[...] termos tais como oração, frase, período (conjunto de orações) etc.” (ibidem). No Glossário de sua obra, Castilho, define o vocábulo ‘sentença’ da seguinte forma: “1. Sentença ou oração é a unidade da sintaxe estruturada por um verbo que seleciona o seu sujeito e seus complementos. Os adjuntos também integram uma sentença, mas não são selecionados pelo verbo [...]” ([2012,] p. 691). Ao que nos parece, o autor considera os dois termos intercambiáveis, apesar de preferência pela utilização de ‘sentença’. Esta pressuposição confirma-se pelo fato de que no vocábulo ‘oração’, há uma remissão ao vocábulo ‘sentença’ expressa por ‘Veja **Sentença**’ ([Castilho, 2012,] p. 686). Em nossa análise preferimos utilizar o termo ‘oração’, pois, se ‘sentença’ e ‘oração’ são intercambiáveis, a utilização deste último se torna menos obscura. Inclusive, o termo ‘sentença’ nem mesmo consta ‘Nomenclatura Gramatical Brasileira’, ao contrário de ‘oração’.”

também tem entre seus constituintes um sintagma nominal, *the city*, N5. Por fim, temos o léxico e a função dentro da estrutura gramatical, N6.

Quando a divisão de uma oração não for possível no mais alto nóculo sintático possível (a oração inteira numa única linha de legenda), como sugere o autor, uma segunda opção seria a separação entre sintagma nominal e sintagma verbal, evitando a quebra interna de outros sintagmas que possam fazer parte desses sintagmas. Nesta opção, a quebra aconteceria no nóculo N2, como exibida no exemplo abaixo:

*the destruction of the city
was inevitable.*

Esta quebra de linha está ocorrendo no mais alto nível sintático possível. Se esta premissa não fosse respeitada, tal que a quebra de linha acontecesse num nível sintático inferior, por exemplo, em N4, a legenda seria exibida assim:

*the destruction of the city was
inevitable.*

Assim, também, a segmentação deve evitar a separação dos elementos que constituem os sintagmas verbais, nominais, adjetivais, adverbiais e preposicionais (ARAÚJO; ASSIS, 2014; CHAVES, 2012; DIAZ CINTAS; REMAEL, 2007; PEREGO, 2008).

Segundo Araújo e Assis (2014, p. 162), a segmentação também pode ser feita em três formatos de legendas, seguindo os números de caracteres que elas apresentem. No primeiro, ocorre uma distribuição do (ou quase) mesmo número de caracteres para as duas linhas, tendo um formato semelhante ao de um retângulo. No segundo, a distribuição dos caracteres é maior na linha de cima do que na de baixo. No terceiro tipo, a distribuição é o oposto da segunda.

Quadro 1 – Formato das legendas na tela

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

Fonte: Araújo e Assis (2014, p. 162).

Para exemplificar os problemas de segmentação linguística, vejamos os dois exemplos do Quadro 2, contendo problemas de segmentação no nível dos sintagmas verbal e nominal:

Quadro 2 – Legendas com problemas de segmentação linguística

Quebra do sintagma verbal	(1) Agora sou eu que <u>estou</u> <u>lendo</u> seus pensamentos, doutor
Quebra do sintagma nominal	(2) Os seres precisam de <u>alguma</u> <u>espiritualidade</u> , seja ela qual for.

Fonte: Chaves (2012, p.81 e 87).

Em primeiro lugar, temos o exemplo (1) da segmentação ocorrida no sintagma verbal, em que o sintagma verbal composto (estou lendo) teve seu núcleo (lendo) separado. Em seguida, em (2), a segmentação ocorreu no sintagma nominal, em que o especificador (alguma) foi separado do seu núcleo (espiritualidade).

É relevante reconhecermos a estrutura dos sintagmas, a fim de evitar possíveis quebras ou separações entre seus constituintes durante a confecção de legendas. Para isso, foi preciso um embasamento teórico que melhor se adequasse aos objetivos desta pesquisa, que por estar inserida no projeto CORSEL, seguiu as mesmas escolhas, utilizando-se de um viés funcionalista, pois, como afirmam Araújo e Assis (2014, p.164) “essa vertente dos estudos linguísticos possui uma maior aproximação epistemológica com a metodologia baseada em *corpus*”. Encontramos na utilização da ‘Nova Gramática do Português Brasileiro’ (CASTILHO, 2012) o referencial necessário para a descrição dos sintagmas e das orações.

2.4 A DESCRIÇÃO DO PORTUGUÊS SEGUNDO CASTILHO

Em um segmento conversacional, podem ser identificados unidades descritivas, como o fonema, a sílaba, o morfema, a palavra, o sintagma e a sentença (preferimos o termo oração no lugar de sentença). Para este estudo, iremos nos deter nos sintagmas e nas orações como constituintes de legendas, pois são essas as unidades fundamentais para a análise dos problemas de segmentação linguística ocorrentes na LSE.

2.4.1 A Oração

A oração é um somatório de estruturas, uma unidade gramatical que apresenta as propriedades de regularidade que retratam a conexão existente no interior de cada unidade (CASTILHO, 2012, p. 59). Podemos distinguir as orações coordenadas das subordinadas, por apresentarem características distintas. Na coordenação, as estruturas apresentam equivalência, portanto, são simétricas e não funcionam como constituintes de outras orações (*Mário chegou e André saiu*). Já na subordinação, não há equivalência, portanto, são assimétricas e funcionam como constituintes de outra oração (*A rua está molhada **porque choveu***) (CASTILHO, 2012, p. 346).

2.4.1.1 Oração Coordenada

Segundo Castilho (2012), os períodos compostos por coordenação e subordinação apresentam expansões de uma oração simples. Naqueles compostos por coordenação, os elementos ou orações que os compõem são simétricos, não se modificam e podem mudar de lugar no enunciado sem alterar sua interpretação semântica, como mostram os exemplos retirados de Castilho (2012, p. 347):

(3) *Mário chegou e André saiu.*

(4) *Amanhã chega Mário ou André?*

As estruturas coordenadas têm seu núcleo em seu interior; já as subordinadas têm seu núcleo fora dos próprios limites, na oração principal. As coordenadas são ligadas por conjunções e podem servir para adicionar informações a algo mencionado anteriormente, manter um valor de inclusão, ou ainda, contrariar uma informação anterior, como nos exemplos:

(5) *Convenceu a todos e não precisou mais gritar.*

(6) *Não saio nem te deixo entrar.*

(7) *Pensei que ia dar certo, mas me enganei.*

2.4.1.2 Oração Subordinada

Segundo Castilho (2012), os períodos compostos por subordinação, assim como os períodos compostos por coordenação, apresentam expansões de uma oração simples. Para distingui-los, podemos focalizar os primeiros por clivagem⁹, visto que os coordenados não têm essa propriedade.

(8)

- a. *Disse que voltaria logo.*
- b. *Disse é que voltaria logo.*

(9)

- a. *Se eu quiser eu saio.*
- b. *Eu saio quando quiser.*
- c. *É se eu quiser que eu saio.*
- d. *É quando eu quiser que eu saio.*

Segundo Castilho (2012), há três tipos de relações de subordinação: as substantivas, as adjetivas e as adverbiais. As substantivas podem ser conjuncionais (*o menino falou que o professor tinha saído*), com o verbo no indicativo ou no subjuntivo; ou não conjuncionais (*o menino falou ter saído o professor*), com o verbo no infinitivo, no gerúndio ou no particípio. Ainda, quando o verbo da matriz for interrogativo, em lugar da conjunção que aparece a conjunção *se* (*o menino perguntou se o professor tinha saído*).

As subordinadas adjetivas aparecem encaixadas num sintagma nominal, em que atuam como complementadores¹⁰. Nesses tipos ocorre uma relação entre dois sintagmas nominais correferenciais, ou seja, que remetem ao mesmo indivíduo, à mesma coisa ou ao mesmo fenômeno. Como mostra Castilho, (2012, p. 366): [*o aluno atento passa de ano.*]; [*O aluno estudioso passa de ano.*]; [*O aluno atento que é estudioso*] *passa de ano*. Sobre as adjetivas, vale ressaltar que elas são introduzidas pelos pronomes relativos: que, qual, cujo, quanto, onde. Contudo, na língua falada o “que” é “uma espécie de pronome relativo universal, que está ocupando espaço dos outros” (CASTILHO, 2012, p. 366). O relevante é

⁹ Clivagem é utilizado como um recurso que divide uma oração em duas, dando ênfase a determinados elementos.

¹⁰ Complementadores são constituintes de um sintagma, e serão vistos mais detalhadamente quando tratarmos dos tipos de sintagmas e suas partes.

sabermos que raramente ouviremos uma sentença do tipo *o livro de história cuja capa está rasgada* no português brasileiro, optando, claro, pela forma coloquial.

As subordinadas adverbiais acrescentam informações adicionais, predicando uma informação que ocorre no verbo, ou ainda, essas orações verbalizam novas informações adicionais de causa, de condição e de tempo. Como exemplifica Castilho (2012, p. 371):

(10) *Ficarei porque Maria vem.*

(11) *Ficarei mais tempo do que Maria pensa.*

As subordinadas descritas anteriormente são do tipo conjuncional. Castilho (2012, p. 380) acrescenta a estas as não conjuncionais: infinitivas, gerundiais e participiais, que detêm seu estatuto de subordinação através dos morfemas do infinitivo, do gerúndio e do participípio. Na tradição gramatical, são conhecidas como orações reduzidas. Como exemplifica Castilho, (2012, p. 380-384):

(12) *Ordenei fechar a porta.*

(13) *Reclamando do barulho, acabou arranjando encrenca com o vizinho.*

(14) *Aborrecido com aquilo tudo, decidiu abandonar o projeto.*

2.4.2 Os sintagmas

Segundo Castilho (2012, p. 55), “[...] o sintagma é a quarta unidade gramatical na hierarquia descritivista. Trata-se de uma associação de palavras articuladas à volta de cinco dentre elas: o verbo, o substantivo, o adjetivo, o advérbio e a preposição”. Para o estruturalismo, o sintagma é especificamente definido como um grupo de palavras que forma uma unidade sintática hierarquizada maior que uma palavra, resultado de uma associação de palavras, e menor que a oração de que é constituinte. O sintagma recebe o seu nome a partir do seu núcleo; ou seja, se o núcleo é um substantivo ou um nome, o sintagma recebe o nome de sintagma nominal (SN). E, da mesma forma, teremos o sintagma adverbial (SAdv), o sintagma verbal (SV), o sintagma adjetival (SAdj), o sintagma preposicional (SP).

Como mostra Castilho (2012), os sintagmas exemplificam a propriedade de “constituência”, isto é, a capacidade linguística de organizar expressões dotadas de uma margem esquerda, um núcleo e uma margem direita. Para simplificar, o mesmo denominou essa regularidade através da seguinte fórmula:

Sintagma = (Especificadores) + Núcleo + (Complementadores)

Explorando essa fórmula básica, iremos detalhar os sintagmas nos seus constituintes principais. Além disso, abordaremos com maior especificidade os sintagmas: verbal, nominal, adjetival, adverbial e preposicional. Os exemplos apresentados a seguir foram todos retirados da Nova Gramática do Português Brasileiro (CASTILHO, 2012).

2.4.2.1 Sintagma verbal

Segundo Castilho (2012, p. 391), “[...] o sintagma verbal é a construção nucleada pelo verbo”. Assemelha-se a uma oração, pois é o próprio verbo que articula seus argumentos, diferenciando-se apenas pela ausência do sujeito. O sintagma verbal segue a fórmula: (especificadores) + verbo + (Complementadores). Primeiro, trataremos do núcleo do sintagma verbal, seguindo os especificadores e, por último, os complementadores.

Seguem os exemplos de núcleo verbal¹¹:

(15) Infelizmente, pobres [**não vivem**], pobres [**vegetam**].¹²

(16) A fita [**é a base do inquérito**].¹³

Podemos observar que no exemplo 15 há dois sintagmas verbais, o primeiro sintagma verbal é composto por um advérbio e um verbo, que constituem o próprio núcleo; assim como no segundo sintagma verbal, composto por um verbo que também constitui o seu núcleo. Já no exemplo 16, o sintagma verbal é composto por um verbo (núcleo) e seu complementador.

Ainda, sobre o sintagma verbal, é relevante considerar algumas das colocações propostas por Castilho (2012). Em primeiro lugar, a propriedade do verbo de maior importância é a transitividade, pois estrutura a oração ao selecionar seus argumentos, constituindo o predicado. Outro aspecto a ser considerado é acerca de seus tipos: sintagma verbal simples, sintagma verbal composto e sintagma verbal complexo. O primeiro tem como

¹¹ Os exemplos de sintagmas verbais estão separados por colchetes e seu núcleo em negrito.

¹² Castilho (2012, p.391).

¹³ *Ibid.*, p. 391.

núcleo um verbo pleno. Os verbos plenos funcionam como núcleo e escolhem seus argumentos atribuindo-lhes papéis temáticos.

O sintagma verbal composto, por sua vez, também tem como núcleo um verbo pleno, mas em sua forma nominal especificado por um verbo auxiliar: infinitivo, gerúndio e participípio. Por exemplo:

(17) O rapaz disse que [**ia contar**] toda a verdade.

(18) Os preços [**estão subindo**] pelo país todo.

(19) [**Tenho visto**] o fulano ultimamente.

Para identificar os especificadores que fazem parte do sintagma verbal, deve-se saber distinguir claramente entre verbos auxiliares e verbos plenos. O primeiro organiza uma perífrase, em que os dois verbos compartilham o mesmo sujeito, sendo assim classificado por Castilho (2012) como especificadores. *Tenho visto o fulano*. Segundo Castilho (2012), os especificadores verbais são classificados em quatro tipos: especificadores de tempo (**Tem-se calado** desde que tomou pé na situação), especificadores de aspecto (**Está acabado**, coitado!); especificadores de modo (**saber falar** em público); especificadores de voz (A notícia **foi difundida** pela estação de rádio).

Os complementadores do sintagma verbal são os constituintes que aparecem à direita do núcleo. Eles são formados por outros sintagmas, como SN, SAdv etc.

2.4.2.2 Sintagma nominal

Segundo Castilho (2012, p. 452), “[...] o sintagma nominal é uma construção sintática que tem por NÚCLEO um substantivo ou um pronome”, somando-se a este um ESPECIFICADOR e um COMPLEMENTADOR. Dessa forma, o mesmo propõe a seguinte fórmula para a estrutura do sintagma nominal:

SN= (Especificadores) + Núcleo + (Complementadores)
--

Podemos observar os seguintes exemplos de sintagma nominal, aparecendo os mesmos entre colchetes e negritos para sua melhor identificação. Por exemplo:

- (20) [**Fogo**] queima.
 (21) [**Isso**] dói.
 (22) [**Este menino**] exige [**tudo**], [**aquele** \emptyset] já não quer [**nada**].
 (23) Ainda bem que [**as encomendas**] chegaram [**todas**].

Os especificadores que podem constituir um sintagma nominal são: artigos, demonstrativos, possessivos, quantificadores, expressões qualitativas, delimitadores. O núcleo pode ser preenchido por um substantivo ou ainda, por pronomes: pessoais, neutros e adverbiais. Os complementadores, por sua vez: os sintagmas adjetivais, os sintagmas preposicionais, as orações relativas.

A estrutura do sintagma nominal pode ser identificada em: a) sintagma nominal simples; b) sintagma nominal composto pelo Especificador + núcleo; c) sintagma nominal composto pelo núcleo + Complementador e d) sintagma nominal composto pelo Especificador + núcleo + Complementador.

Há ainda os quantificadores indefinidos que nucleiam um sintagma nominal, mas não se combinam com um substantivo, dado seu estatuto privativamente pronominal. Podendo funcionar como sujeito (**Alguém** aqui não entendeu isso?), objeto direto e agente da passiva (... prometi também... que diria a vocês se... eu iria exigir cobrar... **algo** do que vocês já fizeram... Então o próprio leite que ela... vamos dizer produzia... era consumido pelo bezerro... e... por **ninguém** mais...).

Para finalizar sobre os possíveis núcleos de um sintagma nominal, Castilho (2012) inclui as expressões dêiticas, que indiciam o lugar e o tempo em que se passa o estado de coisas expresso pelo verbo, podendo, ademais, desempenhar papel argumental. Como no os seguintes exemplos:

- (24) Sua mulher está aí junto, vou **lá** chamar ela.
 (25) Mãe, então quer dizer que **hoje** é o **amanhã** de **ontem**?

Como dito anteriormente, os complementadores, constituintes do sintagma nominal, podem ser representados pelos sintagmas adjetivais, sintagmas preposicionais e orações adjetivas. Os referidos sintagmas serão abordados a seguir.

2.4.2.3 Sintagma adjetival

Assim como os demais sintagmas, o sintagma adjetival é constituído por especificador, núcleo e complementador. Sendo sua forma assim apresentada:

$$\text{SADJ} = (\text{Especificadores}) + \text{Núcleo} + (\text{Complementadores})$$

Segundo Castilho (2012, p. 516): “[...] o sintagma adjetival tem por núcleo o adjetivo, que é uma classe basicamente predicadora, funcionando como adjunto adnominal enquanto constituinte do sintagma nominal, ou como predicativo, enquanto constituinte do sintagma verbal.” Quando esses sintagmas fazem parte do sintagma nominal, eles podem aparecer como nos seguintes exemplos¹⁴:

- (26) Fizeram um barulho **terrível** por causa de um mero acidente
 (27) Pioraram as relações **franco-americanas**.

Já quando os sintagmas adjetivais são parte do sintagma verbal, eles podem se apresentar como núcleo de miniorações:

- (28) A sala estava extremamente **cheia** de curiosos.
 (29) Visto que todos ficaram **calados**, eu também fiquei.
 (30) Ele quer as coisas muito **rápidas**.

Sobre os adjetivos, podemos descrever alguns dos aspectos sintáticos que eles apresentam (CASTILHO, 2012). Em primeiro lugar, os adjetivos concordam com o substantivo em gênero e número e acompanham o substantivo, pospondo-se a ele (situação **brasileira**). Apresentam um Especificador quantificador à sua esquerda (extremamente **cheia**), e podem apresentar um Complementador à direita (**cheia** de curiosos, **repleto** de mentiras). Ainda podem apresentar uma forma deverbal (mente **ocupada**), uma forma simples (barulho **terrível**) ou composta (relações **franco-americanas**).

Sobre os especificadores, Castilho (2012) afirma que são os predicativos qualificadores que atuam como Especificadores do sintagma adjetival (indivíduo [**fortemente**

¹⁴ Exemplo tirado de Castilho (2012, p.518).

marcado pelo destino]), principalmente como advérbios, que podem aparecer à sua esquerda (passagem [**muito mais** cara]).

Segundo Castilho (2012, p. 538), “[...] os adjetivos argumentais selecionam sintagmas preposicionais e orações para atuar como seus complementadores”. Os sintagmas preposicionais podem aparecer acompanhando um adjetivo deverbal (acomodado **com as coisas**) ou acompanhando um adjetivo não deverbal (bom **para isso**). As orações que funcionam como complementadores são, em geral, orações substantivas objetivas (Ficou surpreendido **de que conseguiria o emprego**). Há ainda certos Complementadores de sintagmas adjetivais, como alguns sintagmas preposicionais e expressões comparativas, que organizam expressões idiomáticas (doido **de pedra**, feio **de matar**, amargo **como jiló**) como orações relativas (CASTILHO, 2012).

2.4.2.4 Sintagma adverbial

Para Castilho (2012), o sintagma adverbial segue a mesma fórmula descritiva usada para os sintagmas que mencionamos anteriormente.

$$\text{SADV} = (\text{Especificadores}) + \text{Núcleo} + (\text{Complementadores})$$

Assim, os sintagmas adverbiais podem ser constituídos apenas por um núcleo ([**provavelmente**] [**hoje**] vai chover) ou por um especificador e um núcleo (**muito bem**) ou por um núcleo e um complementador ([**relativamente a sua proposta**]; devo dizer que ela foi acolhida por todos). Sabemos que o advérbio é uma palavra invariável que funciona como modificador do verbo. Castilho (2012) considera modificação o mesmo que predicação, e atribui ao advérbio o papel de adjetivar e de substituir. Além de o advérbio estar relacionado ao verbo, pode também estar relacionado ao adjetivo ou a outro advérbio.

Considerando a Nomenclatura Gramatical Brasileira, Castilho (2012) apresenta 14 tipos de advérbios: afirmação (*sim, certamente, efetivamente*), dúvida (*acaso, porventura, possivelmente*), intensidade (*assaz, bastante, bem*), lugar (*abaixo, acima, adiante, aí, além*), modo (*assim, bem, de balde, depressa, devagar*), negação (*não, nunca, jamais, sequer*), tempo (*agora, ainda, amanhã, anteontem, antes*), ordem (*primeiramente, ultimamente*), inclusão (*inclusive, somente, mais*), designação (*eis*), realce (*lá (em sei lá), cá*), retificação (*aliás, ou antes, isto é,*), situação (*afinal, agora, então, mas*), advérbios interrogativos (*por quê?, onde?, como?*).

Sobre a descrição do núcleo, Castilho (2012) entende que os advérbios podem ser considerados “quase argumentais” por classificarem seu escopo semântico. Esses advérbios são quase “argumentos do verbo” (atender **rapidamente**, permanecer **globalmente** no campo). Ainda, os advérbios operam como adjuntos adverbiais (Discutiu **apenas** francamente) e como adjuntos adsentenciais¹⁵ (**Provavelmente** hoje vai chover). No que concerne à sua transitividade, advérbios derivados de verbos e adjetivos transitivos reproduzem, frequentemente, a respectiva estrutura argumental (Agiu **vinculadamente** ao estrangeiro).

Para a descrição dos especificadores e complementadores, Castilho (2012, p. 582) diz: “[...] visto que o advérbio é a classe que incide sobre as outras classes, inclusive sobre ela própria, outros advérbios podem funcionar como Especificadores do sintagma adverbial”. Além disso, como vimos anteriormente, os advérbios podem ser argumentais ou transitivos, e não argumentais ou intransitivos. Os primeiros tipos terão complementadores (**Contrariamente às expectativas**, os preços baixaram).

2.4.2.5 Sintagma preposicional

A estrutura do sintagma preposicional segue a mesma dos sintagmas já descritos:

SP = (Especificadores) + Núcleo + (Complementadores)

Segundo Castilho (2012, p. 583),

[...] as preposições são palavras invariáveis que atuam como núcleo do sintagma preposicional, desempenhando as seguintes funções: (i) função sintática: ligação de palavras e de oração; (ii) função semântica; atribuição ao seu escopo de um sentido geral de localização no espaço; (iii) função discursiva: acréscimo de informação secundárias ao texto e organização do texto, no caso das construções de tópico preposicionado.

As preposições podem ser classificadas do ponto de vista morfológico em simples e complexas. As simples são formadas por apenas uma palavra. As complexas são definidas como um advérbio ou um substantivo antecedido opcionalmente por certas preposições e seguidos obrigatoriamente por *de* ou *a*.

¹⁵ Como adjuntos adsentenciais, os sintagmas adverbiais precisam ser preenchidos pelos advérbios em *-mente* que pode ser transformado em uma oração nucleada pelo verbo *ser* + o adjetivo. Por exemplo: *é provável = provavelmente*.

Castilho (2012) apresenta um quadro das preposições complexas, o qual ele divide em ADV + PREP (dentro de, fora de, perto de etc.), PRE + SUB + PREP (a cabo de, a par de, ao redor de etc.), PREP + ADV + PREP (diante de, por debaixo de etc.), PREP + PREP (por trás de, para com, por entre).

Quadro 3 – Preposições complexas

ADV+PREP	PREP+SUB+PREP	PREP+ADV+PREP	PREP+PREP
dentro de	a cabo de	diante de	por trás de
fora de	a par de	por/debaixo de	para com
perto de	ao redor de	por/em cima de	por entre
longe de	por amor de	acerca de	de a
antes de	em meio de		
depois de	em vez de		
quanto a	a respeito de		
Junto de	defronte de		
através de	em presença de		
ademais de	á beira de		
	em prol de		

Fonte: Castilho (2012, p.588).

Além do aporte possibilitado pela gramática de Castilho (2012), esta pesquisa, seguindo na esteira do projeto CORSEL, fez uso da metodologia da Linguística de *Corpus*, detalhada na seção abaixo.

2.5 LEGENDAGEM E LINGUÍSTICA DE *CORPUS*

Os estudos baseados em *corpus* significam um grande avanço para as pesquisas em Linguística e Linguística Aplicada, tornando-se populares em diversas áreas, como a tradução, a lexicografia e a tradução audiovisual. Segundo Baker (1996), os estudos feitos a partir de *corpora*, analisados em computador foram, de fato, uma resposta à necessidade de uma metodologia descritiva que pudesse fornecer uma base teórica para o fenômeno tradutório.

Segundo Baker (1995), *corpus* é uma coletânea eletrônica de textos falados ou escritos por diferentes autores que incluem uma grande variedade de fontes e tratam de

diversos tópicos e assuntos variados. Os mesmos são agrupados para um determinado propósito e seguem critérios pré-estabelecidos, a fim de torná-los passíveis de serem analisados de maneira automática ou semiautomática. Baker (1993) preconizou relevantes transformações nos estudos da tradução, propondo a subárea “Estudos de Tradução Baseados em *Corpus* (ETBC)”. Segundo a autora, esses estudos proveriam os teóricos da tradução com meios de observação de seu objeto de estudo, visando a elucidação dos princípios que regem o comportamento tradutório.

Posteriormente, a Tradução Audiovisual em interface com a acessibilidade, ou seja, a audiodescrição e a legendagem, tiveram seus estudos fomentados por essa nova metodologia, o que lhe permitiu um avanço considerável. Na audiodescrição, os estudos de Salway (2007), Hurtado (2007), Hurtado *et al.* (2010) e Marques (2012) investigaram peculiaridades da audiodescrição, como aspectos linguísticos dos roteiros de audiodescrição. Na legendagem, os trabalhos de Perego (2008), Kalantzi (2008), Chaves (2009; 2012), Diniz (2012), Assis (2013), Gabriel (2013) investigaram os parâmetros da segmentação linguística.

Perego (2008) realizou uma análise descritiva e qualitativa por meio de uma análise manual de *corpus*, investigando a segmentação linguística em legendas para ouvintes de um *corpus* heterogêneo de legendas de filmes para DVD e para cinema. Mais especificamente, investigou em que parte da legenda ocorria a quebra de linha. Para sua pesquisa, a autora utilizou os estudos de Karamitroglou (1998). Seus resultados estabeleceram importantes categorias indicativas dos problemas de segmentação: quebra do sintagma nominal, quebra do sintagma verbal, quebra do sintagma preposicional e das orações coordenadas e subordinadas. A partir dessas categorias¹⁶, Chaves (2012) criou etiquetas linguísticas que foram usadas para analisar os problemas de segmentação linguística em LSE.

Kalantzi (2008) teve como objetivo descrever e analisar os parâmetros linguísticos e técnicos da LSE através de uma análise de *corpus*. A autora utilizou um *corpus* composto de 360 LSEs, retiradas de três programas da British Broadcasting Company (BBC) – dois documentários e um programa de atualidades e analisou principalmente as características da segmentação e da redução nessas legendas. Além dos parâmetros linguísticos, foram analisados os parâmetros técnicos de legendagem: tempo de marcação, tempo de exibição, número de caracteres por legenda, velocidade da legenda, restrição de tempo, intervalo entre

¹⁶ As categorias indicativas de problemas de segmentação deram início à pesquisa de Chaves, pois foi o trabalho de Perego que motivou e serviu como ponto de partida para a produção de etiquetas de Chaves (2009) e para a análise de aspectos técnicos (número de linhas e caracteres, velocidade, tempo inicial e final da legenda) e, principalmente, de problemas de segmentação na LSE do filme brasileiro ‘Nosso Lar’ (CHAVES, 2012; CHAVES; ARAÚJO, 2014; ARAÚJO; ASSIS, 2014).

legendas, número de palavras por legenda, número de linhas por legenda e problemas de segmentação. Porém, somente a segmentação foi analisada eletronicamente, através do *software* de legendagem *Claw5*. Esse trabalho foi pioneiro no uso de *corpus* eletrônico para segmentação.

Diniz (2012) analisou as categorias da segmentação linguística das legendas do filme ‘Irmãos de Fé’, disponibilizadas em DVD – tanto as legendas para ouvintes, quanto as LSEs. Em seu aporte teórico, a autora fez uso da gramática gerativista de Lemle do Português Brasileiro, viabilizando a identificação das categorias de segmentação (DINIZ, 2012). Para a análise de dados quantitativos, a mesma fez anotação manual dos problemas de segmentação encontrados em seu *corpus* e em seguida utilizou o programa *WordsSmith Tools 5.0* e suas ferramentas. Na comparação entre ambos os tipos de legenda, a autora chegou à conclusão de que não havia discrepâncias quanto à segmentação. Os problemas de segmentação mais frequentes ocorreram no sintagma verbal.

Além de Kalantzi (2008) e Diniz (2012), podemos citar outras cinco pesquisas que trazem contribuições relevantes para a análise de LSE através da Linguística de *Corpus*: Chaves (2009; 2012), Assis (2013), Gabriel (2013) e Arraes (2015).

Chaves (2009) verificou a viabilidade do uso de *corpora* na análise e descrição de parâmetros de LSE em trechos de gêneros variados da TV Globo. A autora expandiu a aplicabilidade da Linguística de *Corpus* para a análise de LSE, pois, ao contrário de Kalantzi (2008), a qual analisou via *software* apenas os parâmetros linguísticos da legendagem, Chaves (2009) utilizou meios eletrônicos para analisar também os parâmetros técnicos.

A autora (*ibid.*) compilou um *corpus* com as legendas utilizadas pelo grupo LEAD-LATAV (ARAÚJO, 2008), as quais foram baseadas em sugestões de um grupo de consultores surdos do Centro de Atendimento ao Surdo (CAS). Os parâmetros analisados foram: densidade lexical (relacionada à quantidade de palavras por minuto de uma legendagem), nível de condensação (percentual de fala traduzido), segmentação, explicitação por adição e redução de palavras. Para a análise do *corpus*, primeiramente foram feitos os procedimentos de anotação e alinhamento. A Anotação foi manual e feita através de algumas etiquetas específicas para a LSE, como, por exemplo: para a ordem numérica da legenda, legenda 1, fica <L1>; identificação dos falantes, falante 1, fica <fal1>; identificação dos efeitos sonoros fica <ES>. Chaves (2009) utilizou o programa de análise linguística *Wordsmith Tools 5.0*, precisamente as ferramentas *Wordlist* e *View and Aligner*.

Os resultados de Chaves (2009) mostraram que as legendas da Rede Globo de televisão são mais densas e menos condensadas. Além disso, a autora confirmou a viabilidade

do uso de *corpora* para a análise e descrição de legendas, assim como para a sistematização da descrição dos parâmetros de legendagem. Essa pesquisa foi substancial para a realização do meu trabalho, pois foi através da leitura de Chaves (2009), notadamente de sua metodologia, que este foi vislumbrado.

Chaves (2012), dando continuidade aos seus estudos sobre legendagem, tratou da segmentação na LSE via Linguística de *corpus* e buscou estabelecer parâmetros para a análise da segmentação que descrevesse os casos de PROSEGL no âmbito do português brasileiro. Assim, a autora propôs, originalmente, etiquetas que serviriam para a análise eletrônica de aspectos técnicos e linguísticos, sobretudo de problemas de segmentação na legendagem do filme brasileiro ‘Nosso Lar’ (2010). Para a elaboração das etiquetas linguísticas, a autora utilizou-se do referencial teórico encontrado na gramática de Mario Perini, por contemplar a descrição do português falado informal e enfatizar as dimensões sintática e semântica (CHAVES, 2012). Para o alcance dos seus objetivos, a autora fez a anotação manual do *corpus*, identificando os PROSEGL nos diferentes tipos de sintagmas e orações; posteriormente ela utilizou o programa de análise linguística *Wordsmith Tools 5.0* suas ferramentas *Wordlist* e *Concord*, essas ferramentas possibilitaram a análise quantitativa do seu *corpus*.

Assis (2013) analisou a segmentação das legendas da novela brasileira ‘Amor Eterno Amor’ e propôs novas etiquetas de identificação dos problemas de segmentação linguística com subsídios do trabalho de Chaves (2012) e das considerações funcionalistas da gramática de Castilho (2012) sobre os sintagmas do Português Brasileiro Falado. O autor reformulou as etiquetas de PROSEGL propostas por Chaves (2012), algumas em relação à nomenclatura, já outras foram unificadas numa única categoria e outras acrescentadas, dado os casos inéditos de PROSEGL encontrados no *corpus* de seu estudo. Para a identificação dos problemas de segmentação linguística, Assis (2013) totalizou 12 subetiquetas, contra as 19 de Chaves. A metodologia utilizada seguiu os mesmos procedimentos adotados por Chaves (2012).

Utilizando as etiquetas propostas por Assis¹⁷ (2013), Gabriel (2013) descreveu a segmentação linguística de um capítulo das legendas da telenovela ‘Cheias de Charme’, verificando a ocorrência dos principais problemas de segmentação linguística no seu *corpus* e analisando os parâmetros técnicos de legendagem: velocidade, número de caracteres e número de linhas das legendas. A metodologia foi a mesma do trabalho de Chaves (2012) e Assis

¹⁷ As etiquetas elaboradas por Chaves (2012) e Assis (2013) foram refinadas mais uma vez pelo LEAD, e foram essas as utilizadas nesta pesquisa.

(2013), o que possibilitou o levantamento quantitativo dos casos de PROSEGL e em seguida a análise dos dados e a sua relação com os parâmetros técnicos de velocidade e número de linhas. Além disso, a autora propôs algumas sugestões de ressegmentação de LSE para os casos encontrados em seu *corpus*.

Recentemente, Arraes (2015) analisou, também, via Linguística de *Corpus* as LSEs de filmes exibidos na televisão, mais especificamente do filme ‘Virada Radical’, exibido no programa Sessão da Tarde pela Rede Globo. Os resultados desse estudo mostraram uma ocorrência de PROSEGL no sintagma verbal proporcionalmente superior (70,6%) às ocorrências do mesmo tipo de sintagma nos estudos de Chaves (2012), Assis (2013) e Gabriel (2013).

Essas pesquisas apresentaram semelhanças e diferenças: Diniz (2012) e Chaves (2012) analisaram a LSE de filmes de DVDs, Arraes analisou um filme exibido na TV, e Assis (2013) analisou um capítulo da telenovela ‘Amor Eterno Amor’. Gabriel (2013), tal qual Assis, analisou um capítulo da telenovela, especificamente de ‘Cheias de Charme’. Já os resultados quantitativos das pesquisas supracitadas mostraram que a ocorrência de PROSEGL ocorreu em sua maioria no sintagma verbal.

Recentemente, Araújo (2015) traz os resultados de seu projeto CORSEL, que investigou a segmentação em LSE em gêneros diversos, como filmes, séries e telenovelas. Sendo, para o propósito desta pesquisa, relevante nos termos ao gênero das telenovelas. Para esse gênero, foram gravados cinco capítulos de quatro telenovelas brasileiras exibidas pela Rede Globo: *Amor Eterno Amor*, *Cheias de Charme*, *Chocolate com Pimenta* e *Avenida Brasil*. A autora fez uma análise quanti-qualitativa dos problemas de segmentação presentes em seu *corpus*. Essa pesquisa revelou as ocorrências de PROSEGL, de legendas e de tokens, calculados em valores absolutos e índice de frequência Simples. Esses dados estão expostos a seguir:

Tabela 4 – IFS e VA dos PROSEGL encontrados no gênero das telenovelas

TELENOVELA	Nº DE LEGENDAS	TOKENS	PROSEGL	
			VA	IFS %
Amor Eterno Amor	1849	14.856	332	22,3
Cheias de Charme	1725	13.112	242	19,4
Chocolate com Pimenta	1282	9.744	324	34,1
Avenida Brasil	2847	22.998	309	13,4
TOTAL	7703	60.710	1207	19,8

Fonte: Arquivo do projeto CORSEL.

Os valores apresentados mostraram que o *corpus* constituído de 7.703 legendas apresentou características distintas. Em termos de valores absolutos, *Amor eterno Amor* apresentou 332 casos de PROSEGL, *Chocolate com Pimenta*, 324; *Avenida Brasil*, 309; e *Cheias de Charme*, 224. Contudo, não são os valores absolutos que determinam a maior ou menor ocorrência de PROSEGL quando se trata de diferentes *corpora*. Para esse propósito, usou-se o Índice de Frequência Simples (IFS), proposto por Praxedes Filho e Magalhães (2013) e Praxedes Filho e Silva (2014), que é um recurso estatístico adotado para o controle de variáveis; assim, com o uso desse recurso, as médias de PROSEGL em valores de IFS¹⁸ passam a ser aproximadas, possibilitando a comparação entre as mesmas.

Os resultados em IFS, apresentaram uma variação de IFS entre 13,4 e 34,1. *Chocolate com Pimenta* apresentou um IFS de 34,1, em seguida *Amor eterno Amor*, com 22,3; *Avenida Brasil*, 19,8, e *Cheias de Charme*, 19,4.

Sobre os tipos de PROSEGL encontrados nas LSEs para o gênero das telenovelas, foram identificadas 530 ocorrências no sintagma verbal, 44,3%, quase a metade se concentraram entre os constituintes desse sintagma. Esse resultado corrobora outras pesquisas já resenhadas. (CHAVES, 2012; DINIZ, 2012; ASSIS, 2013; GABRIEL, 2013; ARAÚJO; ASSIS, 2014).

Este capítulo procurou localizar a pesquisa no campo disciplinar da Tradução Audiovisual. Em seguida, descreveu as peculiaridades da legendagem para ouvintes, nos seus parâmetros técnicos e linguísticos, bem como apresentou a LSE. Depois, discutiu as considerações gramaticais de Castilho (2012) sobre os sintagmas e as orações visando esclarecer sua importância na identificação dos PROSEGL. Por último, apresentou a importância dos Estudos Descritivos da Tradução e o papel da Linguística de *Corpus*. Logo, a resenha desses trabalhos se fez essencial para situar a presente pesquisa em relação ao estado da arte.

O próximo capítulo traz a descrição dos procedimentos metodológicos adotados nesta pesquisa.

¹⁸ O Índice de Frequência Simples (IFS) corresponde ao número de ocorrências de uma categoria por cada 1.000 palavras de texto, calculado pela divisão do número ou valor absoluto (VA) de uma subetiqueta e o VA total de tokens do sub*corpus*; esse resultado é multiplicado por 1.000.

3 METODOLOGIA

Este capítulo apresenta a metodologia utilizada para desenvolver a presente pesquisa, que teve o suporte teórico-metodológico nos Estudos da Tradução, nos estudos em TAV-LSE e na Linguística de *Corpus*.

3.1 TIPO DA PESQUISA

Esta pesquisa apresenta-se como uma pesquisa descritiva de natureza quanti-qualitativa baseada em *corpus*, pois procurou descrever e analisar as ocorrências dos casos de PROSEGL na LSE da telenovela ‘Cheias de Charme’.

3.2 CONSTITUIÇÃO DO *CORPUS*

O *corpus* é do tipo especializado, constituído por 1725 legendas do tipo *closed caption pop-on*, pertencentes a cinco capítulos da telenovela ‘Cheias de Charme’ e gravados no período entre 16 e 27 de abril de 2012. Durante este período, foram capturados um total de dez capítulos, porém não foi possível extrair a legenda de todos os dez. Sendo assim, as legendas constituintes do *corpus* são provenientes dos cinco capítulos aqui analisados.

Essa telenovela brasileira, escrita por Filipe Grimberg e Izabel de Oliveira, teve direção de Allan Fiterman, Maria de Médicis e Natália Grimberg e direção geral de Carlos Araújo. A mesma foi produzida em 143 capítulos e exibida pela Rede Globo de Televisão entre o período de 16 de abril e 28 de setembro de 2012.

O enredo conta a história de três empregadas domésticas (ver Figura 2) que, após se conhecerem, mudaram completamente suas vidas. As “empreguetes” Maria da Penha, Maria do Rosário e Maria Aparecida são as protagonistas que tiveram seus papéis respectivamente interpretados pelas atrizes Taís Araújo, Leandra Leal, e Isabelle Drummond.

Figura 2 – Imagens da vinheta de abertura da telenovela ‘Cheia de Charme’ e de suas protagonistas



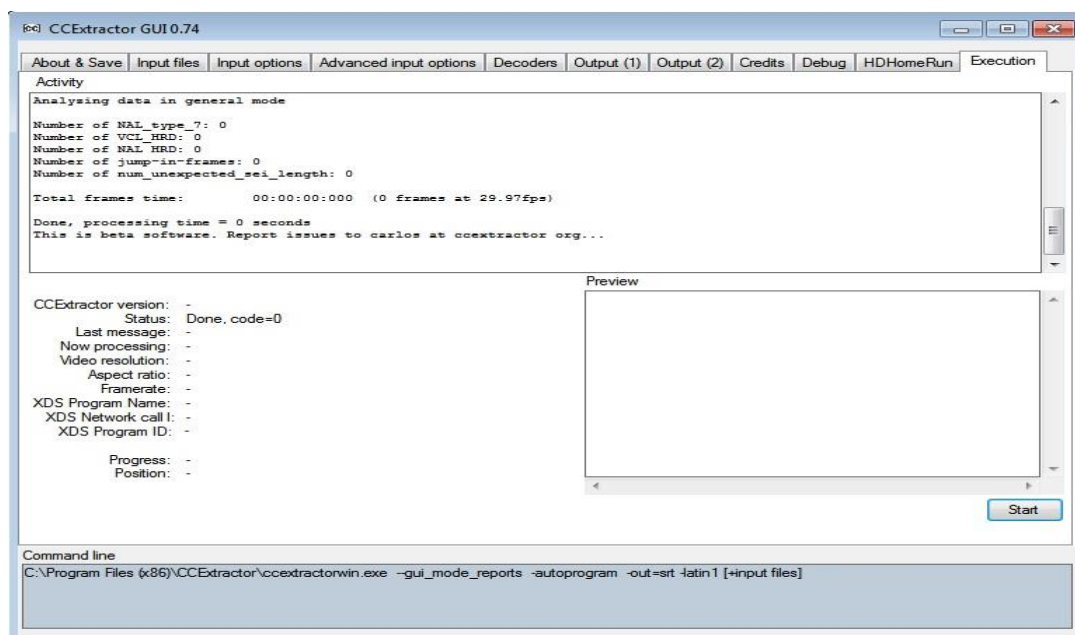
Fonte: <https://www.google.com.br/search?q=figuras+de+cheia+de+charme>

3.3 PROCEDIMENTOS

3.3.1 Extração de legendas

Como mencionei, esta pesquisa faz parte do projeto CORSEL, que teve seu *corpus* extraído automaticamente através do *software CCEXtractor 0.74*. Esse programa é um *software* gratuito, que pode ser encontrado na internet e baixado facilmente. O uso dessa ferramenta ofereceu vantagens para o processo de extração, por ter uma interface simples, de fácil operação.

Figura 3 – CCEXtractor

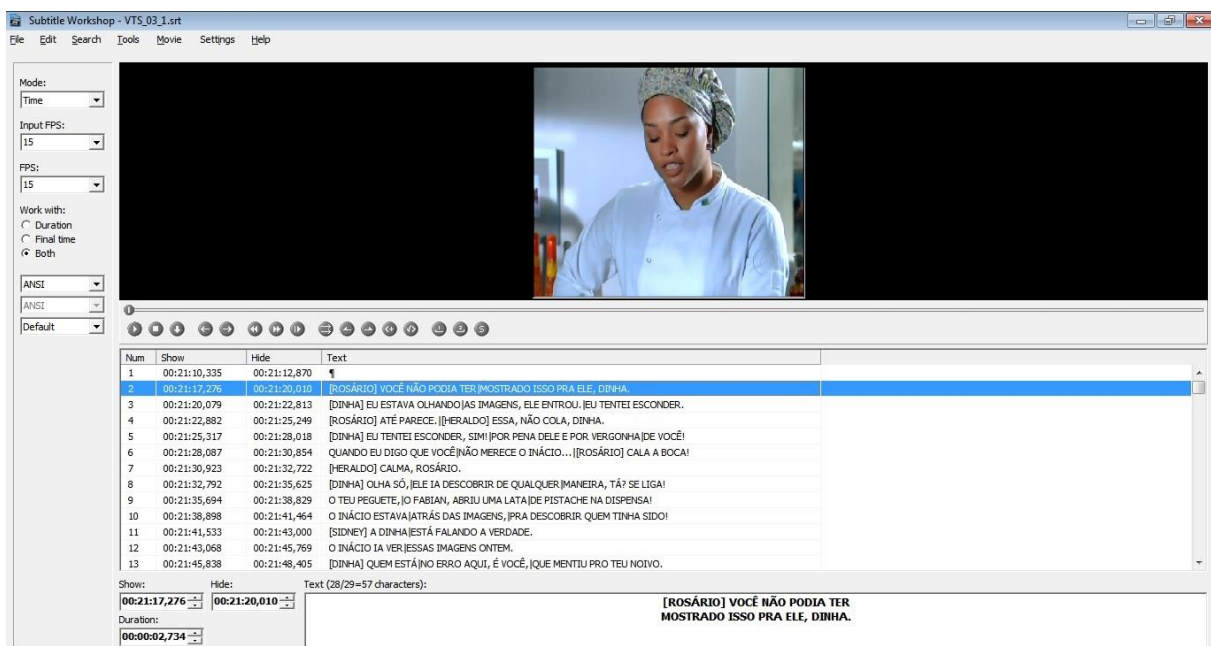


Fonte: Elaborado pela autora.

A primeira etapa desse processo é o registro digital dos capítulos da telenovela exibidos ao vivo, o qual produz um arquivo matriz com três componentes: um de vídeo, um de áudio e uma de legenda do tipo *closed caption*. O *CCE extractor* é utilizado para abrir o arquivo supracitado, para então extrair apenas o componente ‘legenda’, gerando um arquivo de legenda srt.com o número de inserções das legendas, assim como a marcações e as sincronizações das mesmas. Esse tipo de arquivo é um dos formatos utilizados por programas específicos para a criação e edição de legendas. A Figura 4 mostra o arquivo srt. aberto no programa *Subtitle Workshop*, utilizado nesta pesquisa para as análises de parâmetros técnicos das LSEs. Ademais, o *Subtitle Workshop* permite que sejam feitas, entre outras coisas, edições no tempo de exibição, bem como adição de novas legendas.

A utilização deste *software* na pesquisa ocorreu durante dois momentos desta pesquisa: o processo de etiquetagem do *corpus* (vide Subseção 3.3.2 a seguir) e a ressegmentação das legendas. Durante a etiquetagem, o mesmo foi utilizado para verificar a duração das legendas e as informações necessária para o cálculo de suas velocidades. Mais adiante, esse programa viabilizou a ressegmentação das legendas, durante esse processo, o programa permitiu que tivéssemos um panorama das características técnicas, do layout e da segmentação das LSEs. Assim, resultando em uma nova LSE que fosse apropriada aos parâmetros preconizados pelos estudiosos da área.

Figura 4 – Subtitle Workshop



Fonte: Elaborado pela autora.

3.3.2 Etiquetação

A etiquetação é o processo de anotação do *corpus*, o que possibilitou a análise da segmentação da LSE. As etiquetas técnicas e linguísticas usadas para analisar os dados do *corpus* desta pesquisa foram as mesmas desenvolvidas e utilizadas pelo grupo LEAD para o projeto CORSEL, encontradas mais adiante nos quadros três e quatro, presentes nas páginas 51 e 52. A Figura 5 apresenta um exemplo de arquivo de legendas aberto no programa Bloco de Notas depois de anotado com as etiquetas relativas aos parâmetros técnicos e aos problemas de segmentação linguística, os quais serão detalhados mais adiante.

Figura 5 – Arquivo de legenda em formato srt convertido para txt - *corpus* anotado

```

NCDC2012_3_anotadorevisado.txt - Blo...
Arquivo  Editar  Formatar  Exibir  Ajuda
|sub1><1L>1
<t>00:21:32,123 --> 00:21:35,259</t>
<cp11>¶
<veloc_leg_baixa 1c/3,1s></sub1>

<sub2><2L>2
<t>00:21:35,326 --> 00:21:37,795</t>
<cp131>[ELANO] VOCÊ É MUITO BEM-VINDA,
<cp131>CIDA. TODO MUNDO AQUI TE ADORA.
<veloc_leg_alta 62c/2,4s></sub2>

<sub3><3L>3
<t>00:21:37,862 --> 00:21:41,098</t>
<cp117>[CIDA] É, MAS VAI<PROSSEGL><SV_composto>
<cp123>SER MAIS UMA MULHER PRA<PROSSEGL>SP_prep+SV>
<cp130>DIVIDIR BANHEIRO COM VOCÊ, NÉ?
<veloc_leg_alta 70c/3,2s></sub3>

<sub4><2L>4
<t>00:21:41,166 --> 00:21:43,233</t>
<cp125>[ELANO] NÃO TEM PROBLEMA.
<cp117>ESTOU ACOSTUMADO.
<veloc_leg_alta 42c/2s></sub4>

<sub5><3L>5
<t>00:21:43,301 --> 00:21:46,770</t>
<cp125>O QUE EU QUERO É QUE VOCÊ
<cp127>SE SINTA EM CASA COMO NUNCA
<cp130>SE SENTIU NA CASA DO SARMENTO.
<veloc_leg_alta 82c/3,4s></sub5>

```

Fonte: Arquivo do projeto CORSEL.

Ao tomarmos uma inserção de legenda da Figura 5, por exemplo, a primeira, o leitor notará na primeira linha o seguinte: <sub1><1L>1. O número 1 no final da linha sinaliza para o programa de legendagem que esta deve ser a primeira legenda a ser exibida; todavia, para que o pesquisador possa utilizar o mecanismo de busca do Bloco de Notas e achar rapidamente uma legenda, insere-se a etiqueta <sub1>, identificando dentro da etiqueta qual a posição da legenda no arquivo. Na segunda linha, temos a indicação do tempo de

entrada daquela legenda, '00:21:32, 123', bem como o tempo de saída '00:21:35,259', definindo, assim, a marcação e sincronização da legenda 1, ou seja, o tempo em que ela permanecerá na tela. Na terceira linha, temos a indicação de caracteres por linha <cpl>, indicando o número total de caracteres presentes por linha de legenda. Na quarta linha, temos a indicação de velocidade, que pode ser baixa, alta e média. No exemplo, <veloc_leg_baixa 1c/3,1s>, temos uma legenda que se apresentou em uma velocidade baixa, pois continha apenas um caráter exibido em um tempo de 3,1 segundos.

A etiquetagem manual foi realizada através do quadro de etiquetas elaborado pelo grupo LEAD (Legendagem e Audiodescrição), aperfeiçoadas ao longo de 2014 em relação ao quadro de etiquetas utilizado nas pesquisas sobre segmentação em LSE realizadas anteriormente (i.e.: CHAVES, 2012; ASSIS, 2013; GABRIEL, 2013). Foram utilizadas 7 etiquetas para os parâmetros técnicos da legendagem, uma etiqueta para identificação dos problemas de segmentação e 14 subetiquetas para análise dos problemas de segmentação linguística (ver Quadros 4 e 5). Vale lembrar que as etiquetas são o resultado de vários trabalhos e pesquisas realizadas sob a coordenação da professora Dra. Vera Lúcia Santiago Araújo, pioneiramente elaboradas por Chaves (2012), em seguida, adaptadas por Assis (2013) e, mais recentemente, aprimoradas por membros do grupo de pesquisa LEAD que estão envolvidos no projeto CORSEL.

Quadro 4 – Quadro sinótico de etiquetas do projeto CORSEL relativas aos parâmetros técnicos da legendagem

ETIQUETAS DE ANÁLISE DE PARÂMETROS TÉCNICOS DA LEGENDAGEM	
Número da Legenda <sub1>legenda 1</sub1>	<sub1>legenda1</sub1>
Linhas por legenda	<1L>, <2L> e <3L>
Tempos inicial e final de cada legenda	<t>início final</t>
Número de caracteres por linha	<cpl>
Velocidade da legenda baixa (145ppm) 14 e 15cps	<veloc_leg_baixa>
Velocidade de legenda média (160ppm) 16cps	<veloc_leg_média>
Velocidade de legenda alta (180ppm) 17 e 18cps	<veloc_leg_alta>

Fonte: Arquivo do Projeto CORSEL.

As etiquetas usadas para esta pesquisa são divididas em dois tipos: as de análise de parâmetros técnicos e as indicativas de problemas de segmentação linguística. O primeiro tipo (Quadro 4) são as relativas: (i) ao número de linhas de cada legenda, uma linha<1L>, duas linhas <2L> e três linhas<3L>; (ii) ao número de caracteres por linha, <cpl>; e (iii) a velocidade das legendas, <veloc_leg_baixa>, <veloc_leg_média> e <veloc_leg_alta>.

Já o segundo tipo (Quadro 5) pode ser explicado assim: primeiramente há a etiqueta <PROSEGL>, que identifica os problemas de segmentação linguística e é utilizada para encontrar o total de problemas de segmentação no *corpus*. Em segundo lugar, existem aquelas dentro de cada um dos cinco sintagmas: quatro relativas ao sintagma verbal, quatro ao sintagma nominal, duas ao sintagma preposicional, uma ao sintagma adjetival e uma ao sintagma adverbial. Ademais, há também as concernentes às orações subordinadas e coordenadas.

Quadro 5 – Quadro sinótico de etiquetas do projeto CORSEL relativa aos problemas de segmentação linguística

ETIQUETA INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO LINGUÍSTICA
<PROSEGL>
ETIQUETAS DE ANÁLISE DE SINTAGMA VERBAL (SV)
<SV_composto>
<SV_verbo+SAdv>
<SV_verbo+SP>
<SV_(verbo)+oblíquo+SV>
ETIQUETAS DE ANÁLISE DE SINTAGMA NOMINAL (SN)
<SN_especif+SN>
<SN_subst+SAdj>
<SN_subst+SP>
<SN_núcleo>
ETIQUETAS DE ANÁLISE DE SINTAGMA PREPOSICIONAL (SP)
<SP_prep+SN>
<SP_prep+SV>
ETIQUETAS DE ANÁLISE DE SINTAGMA ADJETIVAL (SAdj)
<SAdj_especif+SAdj>
ETIQUETAS DE ANÁLISE DE SINTAGMA ADVERBIAL (SAdv)
<SAdv_adv+SAdv>
ETIQUETAS DE ANÁLISE DE ORAÇÃO SUBORDINADA (SUBORD)
<SUBORD_conj/pron_rel+oração>
ETIQUETAS DE ANÁLISE DE ORAÇÃO COORDENADA (COORD)
<COORD_conj/pron_rel +oração>

Fonte: Arquivo do Projeto CORSEL.

As etiquetas que indicam problemas no sintagma verbal são: <SV_composto>, indicando a quebra que ocorre entre os verbos que fazem parte do núcleo do sintagma verbal; <SV_verbo+SAdv>, quando houver quebra entre verbo e sintagma adverbial; <SV_verbo+SP>, quando houver quebra entre verbo e sintagma preposicional; e <SV_(verbo)+oblíquo+SV>, em que o verbo se separa do oblíquo.

No sintagma nominal, as etiquetas são: <SN_especif+SN>, para quebras entre os especificadores e o SN; <SN_subst+SAdj>, para quebras entre o substantivo e o sintagma adjetival; <SN_subst+SP>, para quebras entre o substantivo e o sintagma preposicional; e <SN_núcleo>, para quebras dentro do próprio núcleo do SN.

Também as ocorrências de PROSEGL podem acontecer no sintagma preposicional (SP). Nesse caso, a quebra ou separação dos constituintes ocorre entre a preposição e o sintagma nominal <SP_prep+SN>, ou entre a preposição e o sintagma verbal, <SP_prep+SV>. Nos dois casos, a preposição é o núcleo do sintagma preposicional.

Além desses tipos de PROSEGL, temos ainda, <SAdj_especif+SAdj>, ocorrência de quebra entre o especificador e o sintagma adjetival, que pode ser seu próprio núcleo, e <SAdv_adv+SAdv>, ocorrência de quebra entre advérbios.

Para os problemas de PROSEGL em períodos compostos por orações, coordenadas e subordinadas, foram elaboradas uma etiqueta para cada tipo: <COORD_conj/pron_rel+oração> e <SUBORD_conj/pron_rel+oração>, respectivamente. A primeira etiqueta deve ser usada quando a quebra de linha ocorre entre a conjunção (ou o pronome) e uma oração, dentro de um período composto por coordenação. A segunda deve ser usada quando a conjunção ou o pronome é separado da oração, dentro do período composto por subordinação.

Além do mais, essas etiquetas seguem a recomendação de que não se deve ocorrer quebras entre os sintagmas constituintes de uma dada linha de legenda, como os verbais (SVs), nominais (SNs), adjetivais (SAdj), adverbiais (SAdv) e preposicionais (SPs), nem nas estruturas internas das orações coordenadas e subordinadas (ARAÚJO; ASSIS, 2014).

3.4 ANÁLISE DOS DADOS

A análise dos dados foi realizada com o auxílio do programa *Wordsmith Tools*, criado por Mike Scott (*Oxford University Press*). O *Wordsmith Tools* é um conjunto integrado de programas para verificar como as palavras se comportam em textos. Esse programa possui três ferramentas principais: *WordList*, *Concord* e *KeyWords*. O *WordList* produz a lista de todas as palavras dos arquivos selecionados: o total de *tokens* (palavras corridas) no texto e *tokens* usado para a lista de palavras. O primeiro traz o número total de palavras presentes no *corpus*, mesmo repetidos; indicando a extensão do *corpus*. o segundo fornece o número de palavras corridas sem a inclusão dos números.

Figura 6 – Tela do Wordlist

	N	Overall	1
text file		Overall	ópia.srt
file size	46,141	46,141	
tokens (running words) in text	9,398	9,398	
tokens used for word list	4,140	4,140	
sum of entries	0	0	
types (distinct words)	1,033	1,033	
type/token ratio (TTR)	24.95	24.95	
standardised TTR	21.82	21.82	
standardised TTR std.dev.	68.50	68.50	
standardised TTR basis	1,000	1,000	
mean word length (in characters)	3.13	3.13	
word length std.dev.	1.82	1.82	
sentences	4,829	689	
mean (in words)	6.01	6.01	
std.dev.	1.36	3.60	
paragraphs	1	1	
mean (in words)	140.00	140.00	
std.dev.			
headings	0	0	
mean (in words)			
std.dev.			
sections	1	1	

frequency alphabetical statistics filenames notes

77 Type-in 4,140

Fonte: Arquivo do projeto CORSEL.

Conforme a Figura 6 ilustra, a terceira linha fornece o total de *tokens*, enquanto a quarta linha apresenta apenas o total de *tokens* apropriados para a pesquisa, porque nela estão apenas o número de palavras corridas sem a inclusão dos números. Por meio da ferramenta *Concord*, foram identificados e quantificados os problemas de segmentação linguística (PROSEGL), como mostra a Figura 7:

Figura 7 – Tela com resultados da ferramenta *Concord*

N	Concordance	Set	Tag	Word #	t.	os.	os.	os.	os.	File	%
1	ele é o reinizio de todo esse<PROSEGL><SN_especif+SN>			653	1.7%	0.9%		0.9%	0.9%	lida e Daniel.txt	28%
2	<cpl28>[Sérgio Chapelin] O povo que<PROSEGL><SUBORD_conj			674	2.5%	0.0%		0.0%	0.0%	lida e Daniel.txt	29%
3	<cpl28>é capaz de criar um dos mais<PROSEGL><SAAdj_especif+SAAdj>			710	2.2%	0.2%		0.2%	0.2%	lida e Daniel.txt	31%
4	<cpl25>Eu também vou amarrar uma<PROSEGL><SN_especif+SN>			596	1.5%	0.7%		0.7%	0.7%	lida e Daniel.txt	26%
5	<cpl30>Bonito, é lindo, é comovete a<PROSEGL><SN_especif+SN>			521	0.9%	0.3%		0.3%	0.3%	lida e Daniel.txt	23%
6	esse e se sentir parte, não<PROSEGL><SV_verbo+SAAdv>***			539	0.2%	0.4%		0.4%	0.4%	lida e Daniel.txt	23%
7	<cpl28>No final da cerimônia, muita<PROSEGL><SN_especif+SN>			566	0.7%	0.5%		0.5%	0.5%	lida e Daniel.txt	25%
8	<cpl23>[Glória Maria] Você vai<PROSEGL><SV_composto>			746	2.8%	0.3%		0.3%	0.3%	lida e Daniel.txt	33%
9	percorrendo devagar essa<PROSEGL><SN_especif+SN>			1,043	4.5%	0.6%		0.6%	0.6%	lida e Daniel.txt	46%
10	<cpl27>A vida aqui segue com ritmo<PROSEGL><SN_subst+SAAdj>			1,057	4.9%	0.7%		0.7%	0.7%	lida e Daniel.txt	46%
11	banho é fora de casa, e para<PROSEGL><intersub><C			1,073	4.4%	0.8%		0.8%	0.8%	lida e Daniel.txt	47%
12	<cpl26>Fica tudo exposto pra quem<PROSEGL><SUBORD_conj			1,016	4.7%	0.5%		0.5%	0.5%	lida e Daniel.txt	44%
13	<cpl25>feita nos antigos teários<PROSEGL><SN_subst+SAAdj>			772	2.0%	0.4%		0.4%	0.4%	lida e Daniel.txt	34%
14	não ficaria encantado com<PROSEGL><SP_prep+SN>			883	2.0%	0.9%		0.9%	0.9%	lida e Daniel.txt	39%
15	aqui que vamos conhecer uma<PROSEGL><SN_especif+SN>			912	4.5%	0.1%		0.1%	0.1%	lida e Daniel.txt	40%
16	<cpl26>[Glória Maria] Ai, é muito<PROSEGL><SAAdj_especif+SAAdj>			506	0.6%	0.3%		0.3%	0.3%	lida e Daniel.txt	22%
17	<cpl27>[Glória Maria] A família se<PROSEGL><SV_(verbo)+oblquo+SV>			179	0.1%	0.8%		0.8%	0.8%	lida e Daniel.txt	7%
18	em volta dessa espécie de<PROSEGL><intersub><S			185	0.2%	0.8%		0.8%	0.8%	lida e Daniel.txt	8%
19	<cpl20>a comida simboliza a<PROSEGL><SN_especif+SN>			207	0.5%	0.9%		0.9%	0.9%	lida e Daniel.txt	9%
20	um bebezinho que acabou de<PROSEGL><intersub><SV_composto>			105	0.8%	0.5%		0.5%	0.5%	lida e Daniel.txt	4%
21	<cpl30>tradicional, o Baci, que ela é<PROSEGL><SV_composto>			53	0.9%	0.2%		0.2%	0.2%	lida e Daniel.txt	2%
22	<cpl25>feita pra afastar os maus<PROSEGL><intersub><S			58	0.0%	0.3%		0.3%	0.3%	lida e Daniel.txt	2%
23	<cpl23>Agora vai acontecer uma<PROSEGL><SN_especif+SN>			81	0.4%	0.4%		0.4%	0.4%	lida e Daniel.txt	3%
24	<cpl28>[Glória Maria] Quem conduz a<PROSEGL><SN_especif+SN>			244	0.2%	0.1%		0.1%	0.1%	lida e Daniel.txt	11%
25	<cpl28>Cada nozinho que eles dão na<PROSEGL><intersub><SP_prep+SN>			450	0.7%	0.0%		0.0%	0.0%	lida e Daniel.txt	19%
26	<cpl25>fitinha eles expressam um<PROSEGL><SN_especif+SN>			460	0.9%	0.0%		0.0%	0.0%	lida e Daniel.txt	20%
27	peessoa faz um pedido, um<PROSEGL><SN_especif+SN>			473	0.1%	0.1%		0.1%	0.1%	lida e Daniel.txt	20%
28	<cpl29>É para garantir que ele tenha<PROSEGL><SUBORD_oração+oração>			402	0.9%	0.8%		0.8%	0.8%	lida e Daniel.txt	17%

wordance collocations plot patterns clusters filenames follow up source text notes

Fonte: Arquivo do projeto CORSEL

Em seguida, esse mesmo procedimento é adotado para identificar e quantificar cada um dos problemas relacionados à quebra dos sintagmas e das orações por meio das subetiquetas correspondentes. Foram utilizados dois tipos de quantificações para os resultados encontrados no *corpus*. O primeiro relacionou os problemas ao número de legendas por meio de valores absolutos. O segundo relacionou os dados ao número de *tokens*.

Como cada capítulo apresenta dimensão diferente, ou seja, possui um número total desigual de *tokens*, adotou-se a solução proposta por Praxedes Filho e Magalhães (2013) e Praxedes Filho e Silva (2014), o uso do Índice de Frequência Simples (IFS), recurso estatístico adotado para o controle dessa variável. O objetivo é mostrar que as médias de problemas de cada programa são estatisticamente comparáveis, dessa maneira, se os episódios apresentarem os IFS próximos entre si, eles podem ser considerados como tal. Um IFS corresponde ao número de ocorrências de uma categoria por cada 1.000 palavras de texto. Para se chegar a um IFS, o número ou valor absoluto (VA) de ocorrências de uma subetiqueta é dividido pelo VA total de *tokens* do *subcorpus* e o resultado é multiplicado por 1.000.

Depois de colhidos esses resultados, foi avaliada a relação entre os parâmetros técnicos e os problemas de segmentação e, finalmente, foram propostas novas segmentações para os itens do *corpus* com o intuito de demonstrar como os resultados da pesquisa podem auxiliar o legendista em formação. Para as propostas, foram utilizadas as velocidades, que, segundo Diaz Cintas e Remael (2007, p. 96), são as velocidades com que um espectador consegue ler as legendas confortavelmente e harmonizá-las com as imagens. As estratégias usadas para a ressegmentação foram a condensação ou edição por redução (uso de palavras mais curtas), omissão (eliminação de palavras redundantes) e paráfrase.

Findado o detalhamento sobre a metodologia da pesquisa, o capítulo seguinte discorrerá sobre as análises e discussão dos resultados.

4 RESULTADOS E DISCUSSÃO DOS DADOS

Este capítulo apresenta os resultados e a discussão dos dados obtidos. Em primeiro lugar, estão expostos os dados quantitativos referentes ao número de *tokens*, ao IFS, os resultados, com exceção do IFS, foram obtidos pelo programa *WordSmith Tools*, mais especificamente da ferramenta *WordList*. Em segundo lugar, a análise primordial desta pesquisa, que tratou do parâmetro linguístico da segmentação. Em terceiro lugar, a relação entre os PROSEGL e os parâmetros técnicos. Finalmente, os resultados da triangulação dos dados, o que me permitiu um olhar mais detalhado sobre a ocorrência de PROSEGL no *corpus* em estudo, comparando-os com as demais telenovelas do projeto CORSEL: Amor Eterno Amor, Chocolate com Pimenta e Avenida Brasil.

4.1 RESULTADOS QUANTITATIVOS DE ‘CHEIAS DE CHARME’

A Tabela 5 traz as ocorrências de PROSEGL, o número de legendas e *tokens* da LSE as quantificações em valor absoluto e por índice de frequência simples de cada capítulo da novela.

Tabela 5 – VA e IFS dos PROSEGL em ‘Cheias de Charme’

Programa	Nº de legendas	<i>Tokens</i>	PROSEGL	
			VA	IFS
Capítulo 1	252	1927	23	11,9
Capítulo 2	410	2776	46	16,5
Capítulo 3	345	2801	66	23,6
Capítulo 4	403	3331	58	17,4
Capítulo 5	315	2277	49	21,5
TOTAL	1725	13112	242	18,4

Fonte: Arquivo do projeto CORSEL.

Em geral, podemos dizer que os problemas de segmentação linguística encontrados foram distintos para cada capítulo, variando entre 23 e 49 PROSEGL em termos de valores absolutos. O IFS para cada um dos cinco capítulos analisados também apresentou uma variação entre 11,9 e 23,6. O primeiro apresentou números inferiores aos demais, tanto para o número de legendas e *tokens* quanto para o valor absoluto e IFS. É possível que o número menor de legendas e *tokens* presente no primeiro capítulo justifique uma ocorrência menor de PROSEGL.

Para os demais capítulos, essa variação não é tão grande, apresentam valores absolutos entre 46 e 66, e IFS entre 16,5 e 23,6. O que se destaca nessa análise é que embora

o menor capítulo tenha apresentado menos ocorrências de PROSEGL, o maior capítulo não apresentou maiores ocorrências de PROSEGL, como poderíamos esperar. Talvez, essa discrepância se explique se considerarmos que não é o tamanho do *corpus* que define a quantidade de ocorrências de PROSEGL. Portanto, faz-se relevante, também, tomarmos o total geral do *corpus*, constituído de 1725 legendas, apresentou 242 ocorrências de PROSEGL, representando um IFS de 18,4.

4.2 TIPOS DE PROBLEMA DE SEGMENTAÇÃO LINGUÍSTICA (PROSEGL)

Os resultados obtidos e expostos na Tabela 6 mostram a ocorrência dos problemas de segmentação linguística por tipos de sintagma e de orações encontrados em cada capítulo da telenovela pertencente ao *corpus*. Ainda, na tabela, podemos ver o total geral de ocorrência de PROSEGL por tipo de PROSEGL, em números quantitativos e percentuais.

Tabela 6 – VA dos tipos de PROSEGL em ‘Cheias de Charme’

Programa	SV	SN	SAdj	SAdv	SP	COORD	SUBORD
Capítulo 1	8	8	1	0	4	1	2
Capítulo 2	16	15	0	0	6	0	9
Capítulo 3	23	17	1	0	5	6	9
Capítulo 4	19	20	0	0	11	5	3
Capítulo 5	19	20	0	3	5	0	5
TOTAL 1	85	80	2	6	31	12	27
(%)	35%	33%	1%	2%	13%	5%	11%

Fonte: Elaborado pela autora.

4.2.1 PROSEGL no sintagma verbal

Dentre essas ocorrências de PROSEGL, foram identificados 47 casos de PROSEGL do tipo SV_composto (19,4 %); 21 casos do tipo SV_verbo+SAdv (8,6%) 13 casos do tipo SV_(verbo)+oblíquo+SV (5,4%); e 4 casos do tipo SV_verbo+SP (1,6%).

Sabemos, então, que a maioria dos PROSEGL ocorreu no sintagma verbal do tipo composto. Essas ocorrências são quebras linhas entre os constituintes do núcleo do sintagma verbal – verbo auxiliar e verbo pleno – como ilustra a Figura 8, a qual apresenta uma legenda de três linhas.

Figura 8 – Imagem com exemplo de PROSEGL do SV_composto



Fonte: Rede Globo, Cheias de Charme, legenda 128, capítulo 5.

Podemos tomar como exemplo de PROSEGL do tipo **SV_composto** a legenda 128 do Capítulo 5 e a 350, do Capítulo 2, respectivamente. No primeiro caso, a quebra ocorre entre os verbos “vou” e “ver” e no segundo, entre os verbos “precisa” e “cumprir”.

Quadro 6 – PROSEGL do SV_composto

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
128	00:27:22,006 00:27:27,210	3	5,2s	180 ppm (Alta)	PÔ, TU ACHA QUE EU <u>VOU</u> <u>VER</u> TU FAZENDO BURRADA E NÃO VOU CAGUETAR PRA TUA MÃE?!?
350	00:36:09,533 00:36:14,003	2	4,4s	145 ppm (Baixa)	[PENHA] O QUÊ QUE <u>NÃO</u> <u>PRECISA</u> <u>CUMPRIR</u> , SANDRO?

Fonte: Elaborado pela autora.

A legenda 128 se apresenta em três linhas e com um número elevado de caracteres (76 caracteres para 5,2s), mas mesmo apresentando muitos caracteres, ainda assim, estaria dentro do máximo permitido para uma legenda de duas linhas se considerarmos o padrão europeu; isso anula uma possível justificativa para a exibição dessa legenda em três linhas. No entanto, no Brasil, a legenda precisaria ser condensada ou transformada em duas legendas. A primeira opção seria, então, pela condensação, evitando a quebra do sintagma verbal composto. Para quatro segundos, tempo máximo de permanência de uma legenda no Brasil, a

nova legenda deveria ter entre 58 e 70 caracteres. Considerando estas restrições, proponho ressegmentar da seguinte maneira¹⁹:

**Achas que vou ver tuas burradas
e não dizer pra tua mãe?**

A legenda ficou com 68 caracteres, logo, está dentro dos padrões, mas como teve que ser bastante condensada, o idioleto da personagem foi totalmente apagado, descaracterizando-a totalmente. A fala da personagem, com uma variedade popular, foi ressignificada na norma culta. Apesar de os manuais de legendagem preconizarem o uso da norma culta, concordamos com os legendistas da novela, que procuraram repetir a variedade linguística usada pela personagem.

Portanto, o melhor nesse caso, parece ser a segunda opção, que seria dividir a fala em duas legendas, a primeira com duas e a outra com uma linha. Para o tempo de duração dessa legenda, poderíamos ter entre 75 e 78 caracteres, portanto, não necessitando editar as legendas:

**Pô, tu acha que eu vou ver
tu fazendo burrada**

e não vou caguetar pra tua mãe?

Assim, a primeira legenda poderia permanecer na tela por 3 segundos, pois ficou com 45 caracteres, seguindo as sugestões de Diaz Cintas e Remael (2007) e mantendo o formato triangular. Já a segunda, teria o tempo de exibição de 2,2s e 31 caracteres.

Para a legenda 350, apresentada no Quadro 6, não é necessário fazer nenhum tipo de edição, porque os 45 caracteres cabem com folga dentro do tempo de duração da legenda (4,4s):

**[Penha]
O quê não precisa cumprir, Sandro?**

¹⁹ Todas as propostas de ressegmentação expostas neste capítulo são de minha autoria.

A proposta de ressegmentação coloca a oração inteira numa linha e a identificação do falante noutra. Quando possível, essa é uma boa maneira de destacar a personagem e sua fala em uma LSE.

A segunda maior ocorrência de casos de PROSEGL do tipo sintagma verbal ocorreu em quebras de linhas com a separação entre o verbo e o advérbio, como ilustra a figura 9:

Figura 9 – Imagem com exemplo de PROSEGL do SV_verbo+SAdv



Fonte: Rede Globo, Cheias de Charme, legenda 223, capítulo 5.

Foram encontrados 21 casos desse tipo **SV_verbo+SAdv**, como nas legendas 223, Capítulo 5, e 308 do Capítulo 3.

Quadro 7 – PROSEGL do SV_verbo+SAdv

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
223	00:32:20,632 00:32:22,705	3	2s	180 ppm (Alta)	[ROSÁRIO] NÃO, NÃO PRECISA, NÃO. NÃO PRECISA, NÃO. <u>EU NÃO VOU</u> DEMORAR. EU VOU NUM MINUTO
308	00:37:51,101 00:37:53,302	3	3,7s	180 ppm (Alta)	[LYGIA] O RAPAZ QUE ME ENTREVISTOU DISSE QUE <u>GOSTOU BASTANTE</u> DO MEU CURRÍCULO.

Fonte: Arquivo do projeto CORSEL.

Na primeira legenda, temos a quebra entre o advérbio “não” e o verbo “vou”, como mostra a Figura 9. Também há um excesso de repetições, portanto me parece que a saída está em reduzir essas repetições, eliminando as redundâncias (“não vou demorar”) e

omitindo o sujeito. Essas estratégias são necessárias, não apenas para solucionar o caso de PROSEGL, mas também, para que através do parâmetro da redução, essa legenda se torne apropriada ao tempo de exibição de 2s e ao número de caracteres por linha que ela se apresenta.

[Rosário]

Não precisa. Vou num minuto.

Na segunda legenda, a quebra de linha ocorreu entre o verbo “gostou” e o advérbio “bastante”. Também. A mesma se apresenta com um pouco mais de “folga” na edição, pois a legenda ultrapassa o número de apenas 8 caracteres. Para reduzir a legenda para 68 caracteres, foram eliminadas a palavra redundante (o rapaz), além da nominalização de “o rapaz que me entrevistou” e da redução pela troca de “bastante” por um advérbio mais curto “muito”.

[Lygia] O entrevistador me disse

que gostou muito do meu currículo.

Para o sintagma verbal, ainda tivemos 13 ocorrências de PROSEGL em que a quebra ocorreu na subetiqueta **SV_(verbo)+oblíquo+SV**, como na legenda 201, Capítulo 3, e a legenda 133 do Capítulo 1, respectivamente.

Quadro 8 – PROSEGL do SV_(verbo)+oblíquo+SV

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
201	00:31:56,547 00:31:58,615	2	2s	160 ppm (Média)	OBRIGADO POR <u>TER</u> <u>ME AVISADO</u> , VIU?
133	00:32:38,489 00:32:43,493	3	5s	145 ppm (Baixa)	[PENHA] Ó, GENTIL, TU <u>PODE</u> <u>ME CHAMAR</u> SÓ DE PENHA, TÁ? EU ATÉ PREFIRO.

Fonte: Arquivo do projeto CORSEL.

Legenda 201 encontra-se dentro dos padrões (33 caracteres para 2s), exceto pelo PROSEGL, no qual o verbo “ter” está separado do pronome oblíquo “me” e do verbo “avisado”. Diaz Cintas e Remael (2007) sugerem que, quando houver espaço, é preferível manter uma legenda em apenas uma linha, evitando quebras desnecessárias e legendas muito curtas. Então, a proposta de ressegmentação seria como exposta abaixo:

Obrigado por ter me avisado, viu?

A legenda acima teria 33 caracteres em uma única linha, sem que seja necessário segmentar algum dos sintagmas pertencentes à mesma. O tempo de 2 segundos mantém a velocidade ideal.

Para a legenda 133, temos a quebra do verbo “pode”, do pronome oblíquo “me” e do verbo “chamar”. Como a legenda para os padrões brasileiros (com quatro segundos) pode ter até 70 caracteres, não é preciso condensar a legenda, mas somente observar a quebra em questão:

[Penha] Ó, Gentil, tu pode me chamar só de Penha, tá?

Eu até prefiro.

Pode-se também dividir a fala em duas legendas:

[Penha] Ó, Gentil,

tu pode me chamar só de Penha, tá?

Eu até prefiro.

A última subetiqueta relacionada ao sintagma verbal **SV_verbo+SP** apresentou apenas quatro ocorrências, como nas legendas 30 e 16 dos Capítulos 3 e 2, respectivamente.

Quadro 9 – PROSEGL do SV_verbo+SP

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
30	00:28:19,897 00:28:22,199	3	2,3s	180 ppm (Alta)	[ALEJANDRO] O QUÊ? MAS GENTE, <u>EU ESTOU</u> <u>NA MINHA CASA</u> , AONDE QUE EU TÔ?
16	00:37:51,101 00:37:53,302	2	2,2s	145 ppm (Baixa)	[INACIO] <u>DÁ</u> <u>PRA ESSE ASSUNTO</u> , ROSÁRIO?

Fonte: Arquivo do projeto CORSEL.

As legendas mostradas no Quadro 9 aparecem com PROSEGL entre os constituintes do sintagma verbal e tiveram verbo (“estou” e “dá”) e sintagma preposicional

(“na minha casa” e “pra esse assunto”) separados em linhas diferentes. Na legenda 30, a mesma foi exibida em três linhas e velocidade alta, 68 caracteres para dois segundos. Então, além de eliminar a quebra indevida, é necessário condensar por redução, eliminando as repetições e redundâncias. A legenda, com 37 caracteres, ficaria assim:

[Alejandro]

Gente, tô em casa? Onde tô?

Para a legenda 16, apesar de não apresentar problemas em relação ao número de caracteres (32) para a velocidade da legenda (2,2s), o PROSEGL em questão ocorreu entre legendas. Então, para fazer uma proposta de ressegmentação, foi preciso retomar a legenda anterior:

[Inácio]

**Eventos por dia, vendo mil
quentinhas. mas isso, não**

Dá pra esse assunto, ROSÁRIO?

O advérbio “não”, proveniente da legenda 15, deveria estar na 16. Então os textos das duas legendas ficariam assim:

[Inácio] Eventos!

Vendo mil quentinhas por dia.

[Inácio]

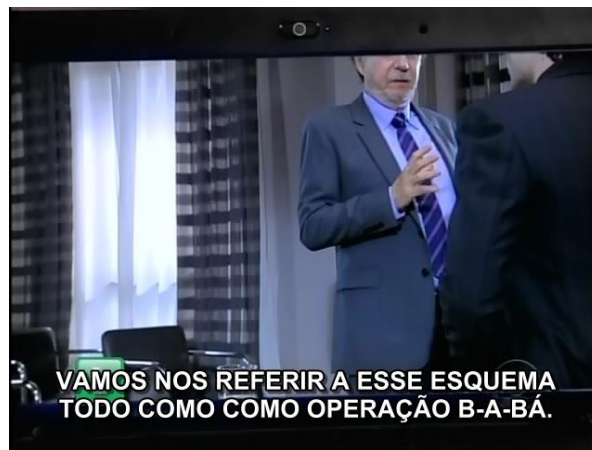
Mas não dá pra isso, Rosário?

A primeira legenda ficaria com duas linhas de 08 e 35 caracteres, respectivamente. A segunda legenda ficaria com 18 e 26 caracteres respectivamente.

4.2.2 PROSEGL no sintagma nominal

Em ‘Cheias de Charme’, o número de ocorrências de quebra no sintagma nominal quase se iguala à do sintagma verbal; temos 80 de SN contra 85 ocorrências de SV. Dos 80 PROSEGL de quebra do sintagma nominal foram encontrados 34 (14%) do tipo **SN_especif+SN**, 25 (10,3%) do tipo **SN_subst+SP**, 15 (6%) do tipo **SN_subst+SAdj** e 5 (2%) do tipo **SN_núcleo**. A figura 10 mostra a ocorrência de um PROSEGL na legenda durante a sua exibição na TV.

Figura 10 – Imagem com exemplo de PROSEGL do SN_especif+SN



Fonte: Rede Globo, Cheias de Charme, legenda 5, capítulo 1.

Para a análise, tomemos como exemplo de quebra do **SN_especif+SN** as legendas 5 (Capítulo 1) e 129 (Capítulo 5).

Quadro 10 – PROSEGL do SN_especif+SN

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
5	00:20:41,172 00:20:46,343	2	5,1	145 ppm (Baixa)	VAMOS NOS REFERIR A <u>ESSE ESQUEMA</u> TODO COMO OPERAÇÃO B-A-BÁ
129	00:27:27,278 00:27:29,712	2	2,4	180 ppm (Alta)	SERIA ATÉ <u>IRRESPONSABILIDADE</u> <u>MINHA</u> NÃO FAZER ISSO

Fonte: Arquivo do projeto CORSEL.

Nos casos acima, a quebra ocorreu entre o especificador (“todo” e “minha”) e o núcleo (“esquema” e “irresponsabilidade”). Em ambos os exemplos, os especificadores

aparecem pospostos aos substantivos. Como na legenda 5 temos tempo suficiente para os 63 caracteres da legenda, basta realocar o especificador (todo) para a primeira linha.

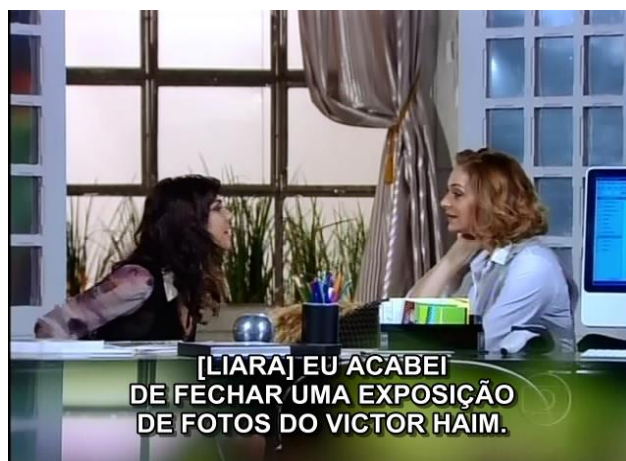
**Vamos nos referir a esse esquema todo
como operação b-a-bá**

Já na legenda 129, o tempo reduzido de exibição da legenda exige que o número de caracteres seja reduzido de 49 para, no máximo, 39, de forma a atender aos padrões preconizados por pesquisadores da área. Portanto, não foi possível contar com uma solução meramente linguística com a utilizada no exemplo anterior. Assim, utilizei a estratégia da condensação.

**Seria irresponsabilidade
não fazer isso.**

Para os casos de segmentação no sintagma nominal, foram encontrados 25 casos de quebra do subtipo **SN_subst+SP**. Podemos ver na figura 11 esse tipo de quebra de linha, em que o Sintagma Nominal aparece com o seu núcleo (substantivo) separado da Sintagma Preposicional.

Figura 11 – Imagem com exemplo de PROSEGL do SN_subst+SP



Fonte: Rede Globo, Cheias de Charme, legenda 147, capítulo 5.

Para a análise, tomemos como exemplo de quebra do **SN_especif+SP** as legendas 147 e 85 (Capítulo 5), expostos no quadro abaixo:

Quadro 11 – PROSEGL do SN_subst+SP

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
147	00:28:23,033 00:28:25,935	3	2,9s	180 ppm (Alta)	[LIARA] EU ACABEI DE FECHAR <u>UMA EXPOSIÇÃO</u> <u>DE FOTOS</u> DO VICTOR HAIM.
85	00:24:35,80 00:24:39,042	2	3,2s	165 ppm (Média)	<u>ELA TÁ COM O MESMO CORPINHO</u> <u>DE SEREIA</u> DE TODA A VIDA

Fonte: Arquivo do projeto CORSEL.

Na primeira legenda, a quebra de linha ocorreu entre o substantivo “exposição” e os sintagmas preposicional “de fotos” , na segunda entre o substantivo “corpinho” e o sintagma preposicional “de sereia”. Nos exemplos, os problemas principais são a quebra do sintagma nominal e a velocidade.

Para solucionar o PROSEGL da legenda 147, que tem 64 caracteres para quase três segundos, podemos omitir, além do sujeito, a informação sobre a exposição de fotos, porque ela pode ser inferida pelo contexto. Com isso, a legenda passaria a ter 52 caracteres, quantidade ideal para o tempo de quase três segundos.

**[Liara] Acabei de fechar
uma exposição do Victor Haim.**

Para a legenda 85, precisamos somente resolver a questão da quebra do sintagma nominal, mas junto ao sintagma nominal em questão há um sintagma preposicional que também não pode ser separado, a solução veio através da condensação por redução com a retirada do especificador “mesmo”.

**Ela tá com o corpo de sereia
de toda vida.**

Para o PROSEGL do tipo **SN_subst+SAdj**, encontramos no *corpus* 16 ocorrências no *corpus* como ilustra a figura 12:

Figura 12 – Imagem com exemplo de PROSEGL do SN_subst+SAdj



Fonte: Rede Globo, Cheias de Charme, legenda 82, capítulo 5

Para a análise desse tipo de PROSEGL tomemos como exemplo de quebra de linha no **SN_subst+SAdj** as legendas 82 (Capítulo 5) e 76 (capítulo 4), expostos no quadro a seguir:

Quadro 12 – PROSEGL do SN_subst+SAdj

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
82	00:24:24,795 00:24:28,065	3	3,2s	180 ppm (alta)	[LAÉRCIO] CULPA DESSA <u>COMIDA</u> <u>ENGORDURADA</u> DO PIAUÍ QUE VOCÊ ESTÁ DEVORANDO, NÉ CHAYZINHA?
76	00:21:35,593 00:21:38,762	3	3,1s	180 ppm (alta)	[OTTO] AH, SÃO AMERICANOS, COM QUEM EU TENHO <u>RELAÇÕES</u> <u>COMERCIAIS</u> .

Fonte: Arquivo do projeto CORSEL.

Para a legenda 82, o problema está na quebra entre os constituintes do sintagma nominal, o substantivo “comida” e o adjetivo “engordurada”, no número de linhas e de caracteres, que são 85, quando deveria ser, no máximo, 55. Como podemos ver na seguinte proposta de uma LSE bem segmentada:

[Laércio]

Culpa da comida engordurada do Piauí?

Na proposta acima, teríamos uma legenda distribuída em duas linhas, com uma redução de 85 para 45 caracteres, com a omissão de informações redundantes como “Né,

Chayzinha”. Pelo contexto, é possível inferir com quem a personagem está falando. Para a legenda 76, o problema está na quebra entre os constituintes do sintagma nominal, o substantivo “relações” e o adjetivo “comerciais”, no número de linhas e de caracteres, que são 72, quando deveria ser, no máximo, 55. Como ressegmentação, levando em conta tais fatores, teríamos:

**[Otto] São americanos,
com quem tenho relações comerciais.**

Além da solução do PROSEGL, do número de linhas e da velocidade da legenda, teríamos também uma oração em cada linha, de acordo com a recomendação de segmentarmos a legenda no mais alto nível sintático (KARAMITROGLOU, 1998).

Para o subtipo **SN_núcleo**, foram encontrados apenas 05 casos de PROSEGL, a legenda 134 (Capítulo 1) e 147 (Capítulo 3), nas quais os núcleos dos sintagmas “Maria” e “dor” estão separados de seus especificadores “da Penha” e “de cabeça”.

Quadro 13 – PROSEGL do SN_subst+núcleo

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
134	00:32:43,561 00:32:46,529	3	2,9s	180 ppm (Alta)	[GENTIL] AH, EU ACHO <u>MARIA DA PENHA</u> UM NOME TÃO BONITO.
147	00:29:05,643 00:29:07,777	2	2,1s	165 ppm (Média)	[SÔNIA] <u>ESSA DOR DE CABEÇA</u> VAI ME MATAR.

Fonte: Arquivo do projeto CORSEL.

A legenda 134 tem o número de caracteres, 53, os quais podem garantir uma boa recepção. Então, a proposta corrige apenas a quebra **SN_núcleo**.

**[Gentil] Ah, acho Maria da Penha
um nome tão bonito.**

Para a legenda 147, para que a legenda seja adequada aos padrões discutidos aqui, basta apenas uma condensação por paráfrase para a redução de 39 para 34 caracteres.

**[Sônia]
Essa dor de cabeça me mata.**

4.2.3 PROSEGL no sintagma adjetival

Em ‘Cheias de Charme’ foram encontradas apenas duas ocorrências de PROSEGL no sintagma adjetival. Para a análise tomemos o PROSEGL que ocorreu na legenda 90, capítulo 5.

Quadro 14 – PROSEGL do SAdj

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
90	00:26:51,242 00:26:55,645	3	4,4s	180 ppm (Alta)	A DONA LYGIA NÃO. TÁ DEMORANDO PRA FICAR BOA. TÔ ACHANDO <u>TÃO</u> ESQUISITO. TU VAI LÁ AGORA?

Fonte: Arquivo do projeto CORSEL

Nessa legenda o sintagma adjetival teve o seu núcleo “esquisito” separado do seu especificador “tão”. O especificador intensifica o núcleo. Aqui temos o desafio de além de evitar uma quebra no sintagma adjetival, reduzir o número de linhas, de três para duas, e caracteres, de 91 para 68.

A proposta abaixo foi obtida por omissão e manteve as orações distribuídas no maior nível sintático, colocando uma oração em cima e duas em baixo.

Dona Lygia tá demorando a ficar boa.

É tão esquisito. Tu vai lá agora?

4.2.4 PROSEGL no sintagma adverbial

Foram encontrados 6 casos de ocorrência de PROSEGL no sintagma adverbial. O sintagma adverbial segue a mesma regra descritiva usada para os sintagmas mencionados anteriormente, ou seja, pode ser constituído de (Especificador) + Núcleo + (Complementador). A legenda 126 do Capítulo 5 ilustra esse caso de PROSEGL no sintagma Adverbial.

Quadro 15 – PROSEGL do SAdv

TCR	Nº de linhas	Duração	Velocidade	Legenda
00:27:15,032 00:27:17,300	2	2,2s	145 ppm (Baixa)	[SAMUEL] AINDA POR CIMA É DEDO DURO

Fonte: Arquivo do projeto CORSEL.

A legenda acima tem apenas 35 caracteres e exibição de 2,2 segundos. A mesma está dentro dos parâmetros de legendagem aqui discutidos, portanto a proposta resolve a questão da quebra de linha, e coloca, para melhor visualização a identificação do falante em cima e a tradução total da fala embaixo.

[Samuel]
Ainda por cima é dedo duro.

4.2.5 PROSEGL no sintagma preposicional

Dos 31 PROSEGL de quebra do sintagma preposicional, foram encontrados 22 OCORRÊNCIAS do tipo **SP_prep+SN**, e 09 do tipo **SP_prep+SV**. Para as análises foram selecionadas as legendas 22 (capítulo 2) e 115 (capítulo 4) .

Quadro 16 – PROSEGL do SP

Nº da legenda	TCR	Nº de linhas	Duração	Velocidade	Legenda
22	00:22:23,942 00:22:27,044	2	3,1s	180 ppm (Alta)	EU QUERO VIRAR ESSA PÁGINA NA MINHA VIDA DE UMA VEZ POR TODAS!
115	00:26:39,497 00:26:43,567	2	4 s	145 ppm (Baixa)	PENHA] Ó, TÔ NA TUA COLA <u>PRA</u> <u>NÃO DEIXAR</u> TU FAZER BESTEIRA!

Fonte: Arquivo do projeto CORSEL.

No primeiro exemplo de segmentação no Sintagma Preposicional, a separação ocorreu na legenda 22 entre a preposição e o substantivo. Nessa legenda, o número de caracteres não ultrapassou o máximo permitido por linha, portanto não havendo necessidade alguma de separar os constituintes do sintagma preposicional. Porém se sabe que essa legenda apresenta uma velocidade alta, o que poderia ser reduzida a medida que se reduz o número de caracteres para o tempo de exibição. Assim, ocultamos o sujeito na oração reduzindo para 59

o número de caracteres. A proposta a seguir tentou respeitar os parâmetros de segmentação, evitando a quebra de linha sem perda de significado:

**quero virar essa página
na minha vida de uma vez por todas.**

No segundo exemplo mostrado no Quadro 16, temos uma LSE de 2 linhas, com velocidade baixa, com 58 caracteres para 4s. A mesma LSE não apresenta características técnicas que justifiquem a quebra de linha ocorrida na LSE, mas que necessita ser ressegmentada. Para isso preferimos um layout geométrico com a primeira linha menor que a segunda, sendo necessário a redução da legenda, através da reformulação e cuidando para que não houvesse nenhuma perda de significado. O número de caracteres da LSE sofreu uma leve redução de 59 caracteres para 49 na LSE. O sintagma preposicional passou a ficar todo na segunda linha da LSE, evitando, é claro, a segmentação indevida do SP.

**[Penha] tô na tua cola
pra tu não fazer besteira!**

4.2.6 Orações coordenadas

No que diz respeito aos períodos compostos por coordenação e subordinação, temos a quebra em 12 coordenadas e 27 subordinadas. Quanto às **coordenadas**, temos a legenda 278 do Capítulo 4, a qual acima apresenta uma quebra indevida na oração coordenada, em que a conjunção (“ou”) se segmentou da segunda oração, quando deveria acompanhá-la. A legenda tem 56 caracteres, quando deveria no máximo ter 35 para quase dois segundos.

Quadro 17 – PROSEGL de Orações Coordenadas

TCR	Nº de linhas	Duração	Velocidade	Legenda
00:39:08,778 00:39:10,746	2	1,9s	180 ppm (Alta)	QUE OU PARECE O FABIAN, <u>OU</u> <u>EMPREGUETE</u> , OU OS DOIS JUNTOS

Fonte: Arquivo do projeto CORSEL.

A ressegmentação sugerida teve seu número de caracteres reduzidos para 46, além da eliminação de palavras redundantes, como a conjunção “que” e o substantivo “juntos”.

**ou parece o fabian,
ou empreguete, ou os dois.**

4.2.7 Orações subordinadas

Quanto ao PROSEGL nas orações **subordinadas**, temos o exemplo da legenda 21 (Capítulo 1), na qual ocorre separação entre parte da oração subordinada (*que eu*) na segunda linha e parte na terceira. Além do PROSEGL, essa legenda se apresenta em 3 linhas e 74 caracteres.

Quadro 18 – PROSEGL de Orações Subordinadas

TCR	Nº de linhas	Duração	Velocidade	Legenda
00:22:54,939 00:22:58,408	3	3,4	180 ppm (Alta)	EU POSSO IR AGORA MESMO. EU JÁ PAGUEI TUDO <u>QUE EU</u> <u>TINHA QUE PAGAR NESSA CASA.</u>

Fonte: Arquivo do projeto CORSEL.

Para essa legenda, precisamos reduzir o número de caracteres de 74 para 61 caracteres. A oração subordinada se manteve completa na segunda linha, solucionando a quebra. Ainda nessa nova legenda, manteríamos um formato apropriado, além de permanecer dentro dos parâmetros já preconizados e já discutidos em nossas análises anteriores.

**Posso ir agora, já paguei tudo
que tinha que pagar nessa casa.**

4.3 RELAÇÃO ENTRE PROSEGL E PARÂMETROS TÉCNICOS

A Tabela 7 traz a relação entre os PROSEGL encontrados e os parâmetros técnicos analisados: primeiramente, a velocidade, na qual estão diretamente relacionados, o número de caracteres e o número de linhas por legenda.

Tabela 7 – Relação entre PROSEGL e parâmetros técnicos em ‘Cheia de Charme em VA

Programa	Legenda de 2L	Legenda de 3L	145 ppm Vel. Baixa	160 ppm Vel. Média	180 ppm Vel. Alta
Capítulo 1	12	12	7	0	16
Capítulo 2	23	23	33	1	12
Capítulo 3	32	32	10	4	50
Capítulo 4	18	40	11	5	42
Capítulo 10	31	20	14	1	36
TOTAL 1	115	127	75	11	156
TOTAL 2 (%)	47%	53%	41%	4,5%	54,5%

Fonte: Arquivo do projeto CORSEL.

Entre os 242 problemas de segmentação linguística encontrados no *corpus* analisado, 115 problemas (47%) ocorreram em legendas de 2 linhas e 127 (53%) em legendas de 3 linhas e de velocidade alta, corroborando os resultados de ‘Amor Eterno Amor’ (ASSIS, 2013; ARAÚJO; ASSIS, 2014), em que a maioria de PROSEGL ocorreu em legendas de 3 linhas e de velocidade alta.

Esses resultados indicam a necessidade de se estabelecer e obedecer aos parâmetros técnicos em harmonia com os parâmetros linguísticos, para que assim se alcance uma legenda de qualidade, que atenda às necessidades do seu público-alvo. Já se sabe que o *corpus* em estudo faz parte de um projeto bem mais audacioso, e também que para a viabilidade do referido projeto, e por sua extensão, o seu *corpus* foi fragmentado por pesquisas menores, mas nem menos importantes. Neste momento, o grupo LEAD, já realizou a tarefa de analisar “esses fragmentos” do CORSEL, mas os resultados ainda estão sendo compilados para futura publicação. Coube a mim, expor e comparar previamente esses resultados obtidos.

4.4 A SEGMENTAÇÃO LINGUÍSTICA NO GÊNERO TELENOVELA

Após a análise dos dados encontrados para o *corpus* desta telenovela, percebemos que seria relevante uma comparação entre os outros trabalhos do grupo, afinal percebo a minha pesquisa como uma parte importante de um “quebra-cabeça” que, quando montado, poderá trazer conclusões mais sólidas e significantes. Assim, para o gênero telenovela, foram analisadas quatro telenovelas distintas. Além de ‘Cheias de Charme’, tivemos ‘Amor Eterno Amor’, ‘Chocolate com Pimenta’ e ‘Avenida Brasil’.

4.4.1 Quantidade de problemas de segmentação

Os primeiros resultados quantitativos mostrados abaixo caracterizam os números de ocorrências de PROSEGL, de legendas e de *tokens*. Esses dados são calculados em termos de valor absoluto e de índice de frequência simples de cada telenovela.

Tabela 8 – VA e IFS dos PROSEGL encontrados no gênero das telenovelas

TELENOVELA	Nº DE LEGENDAS	TOKENS	PROSEGL	
			VA	IFS
Amor Eterno Amor	1849	14.856	332	22,3
Cheias de Charme	1725	13.112	242	18,4
Chocolate com Pimenta	1282	9.744	324	33,2
Avenida Brasil	2847	22.998	309	13,4
TOTAL	7703	60.710	1207	19,8

Fonte: Arquivo do projeto CORSEL.

Ao observarmos os valores apresentados na Tabela 8, podemos ter uma visão geral dos dados obtidos, e então caracterizar o gênero da telenovela dentro do Projeto CORSEL. Para isso, foram analisadas 7.703 legendas de quatro telenovelas brasileiras, para cada telenovela foram analisados cinco capítulos, totalizando 20 capítulos de quatro telenovelas diferentes. Contudo, os capítulos analisados apresentaram valores quantitativos diferentes, apesar de números de capítulos iguais. As quantidades de legendas, *tokens* e valores de PROSEGL encontrados e analisados se distinguiram umas das outras.

Em geral, podemos dizer que os problemas de segmentação linguística encontrados também foram distintos, sendo que a maioria das ocorrências de PROSEGL, em termos de valores absolutos presentes no *corpus* da telenovela, apresentou-se em ‘Amor eterno Amor’ (332), seguida por ‘Chocolate com Pimenta’ (324), ‘Avenida Brasil’ (309), e ‘Cheias de Charme’ (224), respectivamente. Lembrando que, para que a comparação desses resultados se sustente, foi preciso tomarmos os dados em IFS, pois assim *corpus* de tamanhos distintos se tornam comparáveis: ‘Amor eterno Amor’ (22,3), seguida por ‘Chocolate com Pimenta’ (18,4), ‘Avenida Brasil’ (13,4), e ‘Cheias de Charme’ (18,4). Temos uma média de 21,8.

4.4.2 Tipos de problemas de segmentação

A Tabela 9 traz os dados quantitativos de PROSEGL por tipos de sintagmas e de orações para o gênero em questão.

Tabela 9 – VA dos tipos de PROSEGL do Gênero Novela

Novela	SV	SN	SAdj	SAdv	SP	COORD	SUBORD
AMOR ETERNO AMOR	146	109	6	4	32	6	30
CHEIAS DE CHARME	85	80	2	6	31	12	27
CHOCOLATE COM PIMENTA	150	61	8	1	27	22	55
AVENIDA BRASIL	149	71	10	3	36	5	36
TOTAL	530	321	26	14	126	45	148
(%)	44,3	26,8	2,1	1,1	10,5	3,7	12,3

Fonte: Arquivo do projeto CORSEL.

Dos 1207 casos de PROSEGL encontrados nas legendas das telenovelas que foram analisadas, 530 casos ocorreram na separação dos constituintes do sintagma verbal, 44,3%, quase metade das ocorrências se concentrou nesse tipo de sintagma. Esse resultado corrobora o resultado de outras pesquisas que também estão inseridas nesse tipo de pesquisa. (CHAVES, 2012; DINIZ, 2012; ASSIS, 2013; GABRIEL, 2013; ARAÚJO; ASSIS, 2014).

Contudo, a ocorrência de PROSEGL no sintagma nominal foi a segunda maior no *corpus* analisado. No universo de 1207 ocorrências de PROSEGL, 321 foram no sintagma nominal, correspondendo a 26,8% do total. Significa dizer que, assim como o sintagma verbal, esse tipo de sintagma não pode ser considerado de menor relevância durante a edição de legendas.

Além das evidências de que o maior número de PROSEGL no gênero das telenovelas ocorreu nos tipos de sintagma verbal e sintagma nominal, os resultados obtidos ainda demonstraram que o terceiro maior número de ocorrências se concentrou nas orações subordinadas, 12,3%, seguido pelo sintagma preposicional, 10,5%. São dados que indicam a necessidade de uma análise qualitativa para além dos sintagmas verbais e nominais, são quebras de linhas causadas por esses tipos de PROSEGL que também merecem atenção, pois acredito que se tratando de LSE, não se pode negligenciar orações nem sintagmas, respeitando suas estruturas sintagmáticas e, se possível, manter uma oração mais completa e coerente, como sugerem os estudiosos da área.

4.4.3 Relação entre problemas de segmentação e parâmetros técnicos

A ocorrência de PROSEGL nos diferentes tipos de sintagmas e orações coordenadas e subordinadas deve ser analisada também em consonância com os parâmetros técnicos da legendagem, pois sabemos que legendas não são apenas estruturas linguísticas. Portanto, faz-se necessária a “ligação” dessas ocorrências com as características técnicas que elas se apresentam; queremos dizer que não se pode analisar problemas de LSE isoladamente.

A Tabela 10 traz os resultados das ocorrências de PROSEGL em relação aos parâmetros técnicos de velocidade, relacionados diretamente ao número de caracteres e de número de linhas por legenda.

Tabela 10 – Relação entre PROSEGL e parâmetros técnicos do gênero telenovela

Novela	Legenda de 2L	Legenda de 3L	145 ppm Vel. Baixa	160 ppm Vel. Média	180 ppm Vel. Alta
AMOR ETERNO AMOR	150	182	143	14	175
CHEIAS DE CHARME	115	127	75	11	156
CHOCOLATE COM PIMENTA	161	163	170	16	138
AVENIDA BRASIL	131	178	77	23	209
TOTAL	557	650	465	64	678
(%)	46,1	53,8	38,5	5,3	56,1

Fonte: Arquivo do projeto CORSEL.

Como podemos ver, a maioria das ocorrências de PROSEGL para o gênero das telenovelas ocorreram em legendas de 3 linhas, 53,8% do total de legendas, e com velocidade alta, 56,1% do total de legendas. Entretanto, o número de ocorrências de PROSEGL em legendas de duas linhas também foi alto, 46,1%. Esses resultados mostram certa uniformidade para o gênero em questão: apenas a telenovela ‘Chocolate com Pimenta’ se diferenciou em relação à velocidade, pois apresentou maior ocorrência de PROSEGL em legendas com velocidade baixa e não em legendas de velocidade alta, como as demais.

Esses resultados corroboram com os das pesquisas anteriores, Assis (2013) e Gabriel (2013), que analisaram um único capítulo das telenovelas ‘Amor Eterno Amor’ e ‘Cheias de Charme’, respectivamente, e se defrontaram com os mesmos resultados expostos aqui. Todavia, os resultados obtidos por Chaves (2012) se diferenciam por uma menor ocorrência de PROSEGL, caracterizados por velocidade alta e duas linhas. Chaves (2012) investigou a LSE em filmes em DVD, cuja legendas são de uma ou duas linhas e devem ter sido feitas por profissionais, portanto, justificando seus resultados.

Em resumo, podemos concluir que a ocorrência de PROSEGL para o gênero das telenovelas ocorreu, em sua maioria, em legendas de duas e três linhas com velocidade alta. No entanto, nem a velocidade baixa e nem a velocidade média são fatores de exclusão de PROSEGL, como comprovamos através dos resultados das pesquisas do CORSEL que demonstraram também ocorrências para essas diferentes velocidades. Da mesma forma que apenas a decisão de manter a edição de LSE em legendas de no máximo duas linhas não irão assegurar uma legenda bem segmentada, sem PROSEGL nos tipos de sintagmas e orações.

Vale observar ainda que os resultados indicam que a maioria das ocorrências de PROSEGL está presente nos tipos de sintagma verbal e de sintagma nominal como mostrado nas análises. Isso nos remete a pensar que, como consequência, uma maior e mais precisa atenção deve ser dada pelos legendistas a esses tipos de sintagmas durante sua edição. É provável que, se assim fosse, haveria uma menor incidência de PROSEGL na quebra de linhas das LSEs, pois o conhecimento acerca dos parâmetros da segmentação para a LSE se faz necessário e imprescindível.

5 CONSIDERAÇÕES FINAIS

Esta pesquisa dedicou-se a uma das modalidades da TAVa – cuja área disciplinar vem possibilitando estudos e discussões acerca da produção de produtos audiovisuais que promovam acessibilidade sensorial a PcDVs e a surdos e ensurdecidos – e teve como objetivo principal descrever os problemas de segmentação linguística (PROSEGL) na LSE da telenovela ‘Cheias de Charme’, utilizando a Linguística de *corpus*. Para o alcance desse objetivo, utilizamos as etiquetas propostas por Chaves (2012), Assis (2013) e, posteriormente, refinadas pelo grupo LEAD. O uso dessas etiquetas facilitou a execução das análises eletrônicas, visto que a maioria dos problemas encontrados pôde ser descrito através das mesmas. O *corpus* desta pesquisa foi analisado através do programa *Wordsmith Tools 5.0*. Para tal fim, o *corpus* foi constituído de 1725 legendas, presentes em cinco capítulos da supracitada telenovela.

Como vimos anteriormente, os resultados da análise mostraram que os PROSEGL ocorreram, em sua maioria, entre os constituintes dos sintagmas verbal (35%) e nominal (33%). No sintagma verbal, essa separação ocorreu principalmente no sintagma verbal composto. No sintagma nominal, essa separação ocorreu principalmente entre os constituintes do sintagma nominal: especificador mais sintagma nominal, mas também ocorreu entre substantivo mais SAdj, substantivo mais SP e no núcleo nominal.

O resultado geral demonstrou que a legendagem da telenovela ‘Cheias de Charme’ apresentou um número significativo de problemas de segmentação linguística, representando um total de 14,2% do total de 1725 legendas analisadas. Esses resultados sugerem que as legendas do tipo *closed caption* presentes nos cinco capítulos da telenovela ‘Cheias de Charme’ apresentam muitas ocorrências de PROSEGL.

Além disso, comparando os parâmetros técnicos de velocidade e números de linhas, vimos que a maioria dos PROSEGL ocorreu em legendas de três linhas com velocidade alta. Não obstante, compreendo que não há regras para a edição e exibição da LSE, pois, apesar de a maioria de ocorrências de PROSEGL acontecerem em LSEs de três linhas, os resultados para as LSEs de duas linhas se aproximaram dos de três linhas, ficando com 47% para duas linhas e 57% para três linhas. Quanto à velocidade, a maioria dos casos de PROSEGL ocorreu nas legendas de velocidade alta 54,5%, e velocidade baixa, 41 % apresentando velocidade média, apenas 4,5% do total de PROSEGL presente no *corpus* em estudo. Sendo assim, a LSE do tipo *closed caption* da telenovela ‘Cheias de Charme’ não segue parâmetros de velocidade nem de número de linhas.

Esta pesquisa revela, em consonância com as demais pesquisas sobre segmentação e LSE, uma situação promissora quanto aos estudos de LSE baseados em *corpus*, favorecendo os estudos em TAV. Contudo, trata-se de uma situação ainda incipiente, porque elas ainda representam um pequeno número diante da ampla possibilidade que temos com o uso da Linguística de *Corpus*, pois viabiliza a análise quantitativa de um número significativo de dados. O resultado quantitativo tem sua relevância, pois fortalece as conclusões acerca desse objeto de estudo, possibilitando uma caracterização dos casos de PROSEGL, de modo a saber as plausíveis razões para as quebras de linhas ou separação de legendas. Somando-se a isso, ressalto a importância da análise qualitativa dos casos de PROSEGL do *corpus* em estudo, que, além disso, apresentou nas propostas de ressegmentação algumas possibilidades de se “renovar” essas legendas evitando ou corrigindo os problemas identificados durante o processo de legendagem. Não se deve esquecer que concomitantemente à análise dos casos de PROSEGL foi dada a devida atenção às características técnicas que envolveram a execução dessas LSEs, seguindo a maioria das orientações sugeridas pelos principais preconizadores da área.

Esses resultados indicam a necessidade de se estabelecer parâmetros de segmentação linguística em harmonia com os parâmetros técnicos da legendagem para que assim se alcance uma legenda de qualidade que atenda às necessidades do seu público-alvo. Assim, através das análises e das propostas de ressegmentação, esta pesquisa indica parâmetros para a LSE: legendas de no máximo duas linhas, com o número máximo de 35 caracteres por linhas, os sintagmas devem ser mantidos em uma mesma linha, quando se tratar de um sintagma composto de vários constituintes, essa legenda pode ser condensada e o sintagma pode ter um dos seus constituintes omitidos ou ainda essa legenda pode ser modificada e ter seus constituintes substituídos por outros compostos de um número menor de caracteres. Ainda, quando houver tempo suficiente, uma legenda mal segmentada que se apresente em mais de duas linhas poderá ser expandida em duas legendas, priorizando os critérios linguísticos, como evitar a separação do sintagma em linhas diferentes sempre que possível, e também, mantendo um *layout* retangular, como os sugeridos por Araújo e Assis, 2015.

À guisa de conclusão, quero enfatizar que este trabalho não pretende se encerrar aqui, ao contrário, acredito que o mesmo possibilita a continuidade de novos estudos que envolvam a LSE, apontando caminhos que podem ser seguidos por outros pesquisadores, e também, explorando outros gêneros, além das telenovelas, como filmes, seriados, jornais, propagandas, etc. Também, envolvendo outras dimensões, além da descritiva baseada em

corpus, mas exploratória e estudos de casos. Sobretudo vejo neste estudo o resultado de uma relevante tarefa que procurou atingir seus objetivos e que pode permitir novas problematizações, como e questionamentos. Questionamentos esses capazes de suscitar passos mais audaciosos, como, por exemplo, através de uma pesquisa experimental de recepção com o uso da tecnologia do rastreador ocular. Isto possibilitaria resultados mais conclusivos, pois envolveria sujeitos, elemento fundamental para os estudos da segmentação em LSE. Ainda, sem esquecer nas implicações práticas para a formação do legendista, este estudo possibilita uma reflexão sobre a forma de legendar, principalmente no que se refere as questões de segmentação linguística. Abrindo espaço, para que o legendista compreenda que segmentar vai além de dividir as legendas, que é necessário compreender que constituintes de um sintagma ou de uma oração não podem ser separados aleatoriamente, mas seguindo critérios que possivelmente são determinantes para a recepção de uma LSE.

REFERÊNCIAS

ALVARENGA, Lina. **Subtítler: legendador ou Legendista?** In: **Anais do CIATI** – Congresso Ibero-Americano de Tradução e Interpretação. São Paulo, 1998, 214-216.

ARAÚJO, Vera Lúcia Santiago. A legendagem para surdos no Brasil. In: LIMA, Paula Lenz Costa; ARAÚJO, Antonia Dilamar (Orgs.). **Questões de linguística aplicada: miscelânea**. Fortaleza: EdUECE, 2005.

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

_____. Closed subtitling in Brazil. In: ORERO, Pilar (Org.). **Topics in Audiovision Translation**. Amsterdã: John Benjamins Publishing Company, v. 1. 2004.

_____. O processo de legendagem no Brasil. **Legenders**. Notícias e matérias, arquivos, comentários sobre o mundo da legendagem e traduções. dez/2009. Disponível em: <http://legenders.com.br/Docs/legendagem_no_brasil.pdf>. Acesso em: 23 out. 2012.

_____. Por um modelo de legendagem para Surdos no Brasil. In VERAS, Viviane. (Org.). **Tradução e Comunicação: Revista Brasileira de Tradutores**, São Paulo: UNBERO, n. 17. 2008.

_____. Subtitling for the deaf and hard-of-hearing in Brazil. In: ORERO, Pilar; REMAEL, Aline. (Org.). **Media for all: subtitling for deaf, audio description and sign language**. Kenilworth: Nova Jersey, EUA: Rodopi, 2007.

_____; ASSIS, Ítalo Pinto. A segmentação linguística na legendagem para surdos e ensurdecidos (LSE) de ‘Amor Eterno Amor’: uma análise baseada em *corpus*. **Letras & Letras**, v. 30, n. 2, jul/dez. 2014.

_____; FRANCO, Eliana. Paes Cardoso; ALVARENGA, Lina. Audiovisual translation research in Brasil and in Europe. **Revista Brasileira de Linguística Aplicada**, Belo Horizonte, v. 2, n.2, 2002.

_____; NASCIMENTO, Ana Katarinna Pessoa. Investigando parâmetros de legendas para surdos e ensurdecidos no Brasil. **Tradução em Revista**, v. 2. 2011.

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

ASSIS, Italo Pinto. **A segmentação na LSE de ‘Amor eterno amor’:** uma análise baseada em *corpus*. 2013. Monografia (Bacharelado em Letras Inglês) – Universidade Estadual do Ceará, Fortaleza, 2013.

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

_____. *Corpus linguistics and translation studies: implications and applications*. In: _____; FRANCIS, Gill; TOGNINI-BONELLI, Elena. (Orgs.). **Text and technology: in honour of John Sinclair**. Philadelphia, Amsterdam: John Benjamins, 1993.

_____. *Corpus-based translation studies: the challenges that lie ahead*. In: SOMERS, Harold. (ed.). **Terminology, LSP and translation**. Amsterdã, Filadélfia: John Benjamins, 1996.

BRASIL. **Lei nº. 10.098 de 19 de dezembro de 2000**. Presidência da República, Brasília, DF. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L10098.htm>. Acesso em: 16 set. 2014.

CASTILHO, Ataliba de. **Nova Gramática do Português brasileiro**. São Paulo: Contexto, 2012.

CHAVES, Élide Gama. **Legenda para surdos no Brasil: uma análise baseada em corpus**. Monografia (Bacharelado em Letras Inglês) – Universidade Estadual do Ceará, Fortaleza, 2009.

_____. **Legendagem para surdos e ensurdecidos: um estudo baseado em corpus da segmentação nas legendas de filmes brasileiros em DVD**. 2012. Dissertação (Mestrado em Linguística Aplicada) – Universidade Estadual do Ceará, Fortaleza, 2012.

_____; ARAÚJO, Vera Lúcia Santiago. Segmentation tags: a proposal for the analysis of subtitles. In: TAGNIN, S. O. **New Language Technologies and Linguistic Research: A Two-Way Road**. Newcastle upon Tyne: Cambridge Scholars Publishing, 2014

D'YDEWALLE, Géry; PRAET, Caroline; VERFAILLIE, Karl; VAN RENSBERGEN, Joan. 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, 1987.

DIAZ CINTAS, Jorge; REMAEL, Aline. **Audiovisual translation: subtitling**. Manchester, UK: St. Jerome Publishing, 2007.

DINIZ, Nina Soares Lopes. **A Segmentação em Legendagem para Surdos e Ensurdidos: um Estudo Baseado em Corpus**. 2012. 149f. Dissertação (Mestrado em Linguística) – Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

FRANCO, Eliana Paes Cardoso; ARAÚJO, Vera Lúcia Santiago. Questões terminológico-conceituais no campo da tradução audiovisual (TAV). **Tradução em Revista**, v. 2. 2011.

_____. Reading television: checking deaf people's reaction to closed subtitling in Fortaleza, Brazil. **The translator**, v. 9, n. 2, 2003.

GABRIEL, Maria Helena Clarindo. **A segmentação linguística na legendagem para surdos e ensurdidos (LSE): uma análise baseada em corpus**. 2013. 59f. Trabalho de conclusão de curso (Especialização em Estudos da Tradução) – Universidade Federal do Ceará, Fortaleza, 2013.

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

HURTADO, Catalina Jiménez. Una gramática local del gui3n audiodescrito. Desde la semántica a la pragmática de un nuevo tipo de traducción. In: Hurtado, C. J. (ed.). **Traducción y accesibilidad**. Frankfurt: Peter Lang Internationaler Verlag der Wissenschaften, 2007.

_____; RODRÍGUEZ, Ana; SEIBEL, Claudia. (Eds.). **Por um corpus de cine: teoria y práctica de la audiodescripción**. Granada: Ediciones Tragacanto, 2010.

IVARSSON, Jan; CARROL, Mary. **Subtitling**. Simrishamn, Sweden: GraphoTryck AB, 1998.

JAKOBSON, Roman. Aspectos lingüísticos da tradução. In: JAKOBSON, R. **Linguística e comunicação**. São Paulo: Cultrix, 1995.

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, Fotios. A Proposed Set of Subtitling Standards in Europe. **Translation Journal**, v. 2, n. 2, 1998.

MARQUES, Osmina Maria. **A audiodescrição dos personagens de filmes: um estudo baseado em corpus**. 2012. Dissertação (Mestrado em Linguística Aplicada) – Universidade Estadual do Ceará, Fortaleza, 2012.

NASCIMENTO, Ana Katarinna Pessoa. **Linguística de corpus e legendagem para surdos e ensurdecidos (LSE): um estudo baseado em corpus da tradução de efeitos sonoros segmentação na legendagem de filmes brasileiros em DVD**. 2013. Dissertação (Mestrado em Linguística Aplicada) – Universidade Estadual do Ceará, Fortaleza, 2013.

NOSSO Iar. Direção: Wagner de Assis. Produção: Iafa Britz. Roteiro: Wagner de Assis. Elenco: Renato Pietro, Fernando Alves Pinto, Rosanne Mulholland. Brasil: Cinética Filmes, 2010. 1 DVD (105 min), color.

PEREGO, Elisa. Evidence of explicitation in subtitling: towards a characterization. **Across Languages and Cultures**, v. 4, n. 1, 2003.

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

_____. What would read best? hypothesis and suggestion for the location of line breaks in films subtitles. **The Sign Language Translator and Interpreter**, v. 2, n. 1, 2008.

PRAXEDES FILHO, Pedro Henrique Lima. MAGALHÃES, Célia Maria. A neutralidade em audiodescrições de pinturas: resultados preliminares de uma descrição via teoria da avaliatividade. In: Vera Lúcia Santiago Araújo; Marisa Ferreira Aderaldo. (Org.). **Os novos rumos da pesquisa em audiodescrição no Brasil**. 1ed. Curitiba: CRV, 2013.

_____; SILVA, Cristiene Ferreira. A (in)existência de neutralidade: um estudo de caso baseado em corpus com roteiros de audiodescrições francesas de filmes via Teoria da Avaliatividade. **Letras & Letras**, v. 30, 2014.

REID, Helen. Literature on the screen: subtitle translation for public broadcasting. In: BART, Westerweel; D'HAEN, T. (Eds.). **Something understood**: studies in anglo-dutch literary translation. Amsterdam: Rodopi, 1990.

SALWAY, Andrew. A *corpus*-based analysis of audio description. In: DIAZ CINTAS, Jorge. **Media for all**: subtitling for deaf, audio description and sign language. Amsterdam, New York: Rodopi, 2007.

SCOTT, Mike. **WordSmith Tools**. Version 5. Oxford: Oxford University Press, 2011. Disponível em: < <http://www.lexically.net/wordsmith/version6/index.html>>. Acesso em 10 ago. 2014.

ANEXOS

ANEXO A – SINTÁGMA VERBAL COMPOSTO

Cap	leg	TCR	<SV_composto>	L	cpl	Velocidade
1	35	00:25:50,114 00:25:52,115	[ELANO] CALMA. A GENTE NÃO PODE DEIXAR DE TOCAR O DIA A DIA TAMBÉM.	3	65c/2s	Alta
1	115	00:31:28,486 00:31:31,688	DR. SARMENTO, ME DESCULPA, MAS O SENHOR PRECISA ASSINAR ESSA PETIÇÃO...	3	69c/3,2s	Alta
1	166	00:34:25,529 00:34:27,129	BOM, EU É QUE QUERIA CONVERSAR COM ELA.	2	38c/1,6s	Alta
1	167	00:34:32,703 00:34:35,939	[IVONE] TÔ TE ACHANDO PREOCUPADO, HEIN.	2	38c/3,2s	Baixa
2	10	00:00:38,370 00:00:41,105	PÔ, SÓ QUE VOCÊ ESTAVA AQUI, NÉ, QUE O BUFFET ESTÁ PRECISANDO DE GRANA.	3	69c/2,7s	Alta
2	34	00:02:38,090 00:02:49,067	QUE EU JÁ ESTOU PERDENDO A LINHA!	2	32c/10,9 s	Baixa
2	75	00:07:09,094 00:07:10,962	MAS A GENTE PODE SAIR DE NOVO, NÃO PODE?	2	39c/1,8s	Alta
2	145	00:15:09,841 00:15:15,012	[ISADORA] DEIXA DE SER SONSA, EMPREGUETE Ô! O QÁC	2	47c/5,1s	Baixa
2	188	00:19:29,600 00:19:32,369	[LYGIA] COMO É QUE EU POSSO RECUSAR UMA PROPOSTA DESSA?	2	54c/2,7s	Alta
2	270	00:28:02,613 00:28:11,921	[EMÍLIA] MAS OLHA SÓ, QUEM TEMOS POR AQUI! FINALMENTE EU CONSIGO <cpl31>FALAR COM O ARTISTA DO MOMENTO!	3	94c/9,3s	Baixa
2	350	00:36:09,533 00:36:14,003	[PENHA] O QUÊ QUE NÃO PRECISA CUMPRIR, SANDRO?	2	45c/4,4s	Baixa
2	389	00:39:26,930 00:39:33,002	É PRECISO SER UM HOMEM DE VERDADE.	2	33c/6s	Baixa
3	3	00:21:37,862 00:21:41,098	[CIDA] É, MAS VAI SER MAIS UMA MULHER PRA DIVIDIR BANHEIRO COM VOCÊ, NÉ?	3	70c/3,2s	Alta
3	16	00:18:04,849 00:18:07,084	[GENTIL] BOM, SE VAI SER RECORDE DE AUDIÊNCIA, SÓ ME RESTA ESPERAR UM POUCO, NÉ?	3	78c/2,2s	Alta
3	22	00:22:45,529 00:22:47,530	[ELANO] ADOREI TE VER TIRANDO O UNIFORME.	2	40c/2s	Alta
3	25	00:22:52,036 00:22:55,07 2	ESTAVA SENDO RESGATADA PELOS MEUS AMIGOS.	2	40c/3s	Baixa
3	52	00:24:28,866 00:24:33,136	[VALDA] A TUA FILHA SE PIRULITOU, FOI CUIDAR DA VIDA DELA.	3	56c/4,2s	Baixa
3	67	00:25:45,043 00:25:49,312	[CIDA] EU JÁ TERMINEI DE LAVAR A LOUÇA. VOCÊ QUER QUE EU LEVE O PATRICK NA ESCOLA?	3	80c/4,2s	Alta
3	84	00:26:35,025 00:26:36,626	O QUÊ QUE TU VAI FAZER NA DONA LYGIA?	2	36c/1,6s	Alta

3	159	00:29:42,846 00:29:45,982	MAS COMO É QUE EU VOU ENCARAR AS MINHAS AMIGAS?	2	46c/3,1s	Baixa
3	165	00:30:04,902 00:30:06,703	ACHO QUE VOCÊ VAI TER QUE SE VIRAR SOZINHA.	2	42c/1,8s	Alta
3	193	00:31:35,459 00:31:37,761	TEM ALGUÉM QUE VAI ADORAR SABER DISSO.	2	37c/2,3s	Média
3	205	00:32:15,466 00:32:16,999	[SÔNIA] EU NÃO VIM FALAR COM VOCÊ, VALDELICE.	2	44c/1,5s	Alta
3	268	00:36:39,496 00:36:42,865	MAS LOGO DEPOIS, JÁ ESTAVA TRABALHANDO, ASSIM, NORMAL.E? DO,	2	58c/3,3s	Alta
3	284	00:37:32,682 00:37:35,317	[SANDRO] MINHA NEGUINHA NÃO VAI SAIR COM GENTIL NENHUM, RAPÁ.	2	60c/2,6s	Alta
3	307	00:38:46,190 00:38:48,057	[INÁCIO] NÃO VAI SAINDO ASSIM, NÃO? VEM CÁ!	2	42c/1,8s	Alta
3	317	00:39:20,724 00:39:23,726	QUE QUER GASTAR ESSA GRANA TODA, FAZER UMA BAITA FESTA?	3	53c/3s	Baixa
4	18	00:18:09,420 00:18:10,854	[GENTIL] BOM, ENTÃO, EU VOU TER QUE LIGAR PARA UMA PESSOA, E AVISAR QUE EU TÔ ATRASADO.	3	85c/1,4s	Alta
4	84	00:22:08,293 00:22:12,863	EU IA ADORAR TER A SUA COMPANHIA. SE A PESSOA CHEGAR...	3	53c/4,5s	Baixa
4	182	00:28:04,982 00:28:08,351	[PENHA] GENTE, EU NEM TINHA REPARADO. EU NÃO REPARO MUITO NAS COISAS, NÃO.	3	72c/3,3s	Alta
4	221	00:29:51,756 00:29:54,724	<i>PÔ, O SENHOR NÃO VAI</i> <i>FAZER UMA DESFEITA DESSA.</i>	2	48c/2,9s	Média
4	222	00:29:54,792 00:29:56,760	<i>O SENHOR VAI ENTRAR,</i> <i>PRA TOMAR UM CAFEZINHO</i> <i>COMIGO, NÃO VAI, NAO?</i>	3	66c/1,9s	Alta
4	254	00:31:19,142 00:31:22,178	VOCÊ NÃO QUERIA QUE EU FICASSE CONVERSANDO SOBRE A LIGA <cpl27>DAS PATROAS DO CASA GRANDE?	3	81c/3s	Alta
4	265	00:31:52,009 00:31:55,945	[SARMENTO] ELA DEVE<PROSEGL> ESTAR SE DIVERTINDO, AMOR. ELA ESTÁ NA LUA DE MEL!	3	68c/3,9s	Alta
4	266	00:31:56,013 00:31:59,148	EXATAMENTE O QUE A GENTE DEVIA ESTAR FAZENDO AGORA. VEM CÁ.	2	58c/3,1s	Alta
4	269	00:32:05,389 00:32:09,191	[SARMENTO] MAS JUSTAMENTE! ENTÃO! A GENTE ESTÁ PRECISANDO RELAXAR.	3	64c/3,8s	Média
4	293	00:33:35,212 00:33:38,247	DEVE ACHAR BESTEIRA EU JÁ ESTAR PENSANDO EM PETROQUÍMICA, NÉ? AINDA SOU NOVA.	3	75c/3s	Alta
4	331	00:35:22,853 00:35:24,253	QUEM MANDOU EU ACHAR QUE PODIA TER UMA VIDA NORMAL AO SEU LADO?	2	62c/1,4s	Alta
4	386	00:38:14,357 00:38:17,994	[PENHA] EU VOU LEVAR O DR. OTTO... VOU TE LEVAR LÁ FORA!	3	54c/3,6s	Baixa
5	31	00:21:48,472 00:21:51,141	[INÁCIO] É, MAS CHEGA, EU PRECISO TOCAR MINHA VIDA PRA FRENTE.	3	60c/2,6s	Alta
5	36	00:21:58,883	E AÍ, VOCÊ ACABA	2	36c/1,5s	Alta

		00:22:01,051	EXAGERANDO NO CIÚME.			
5	60	00:23:07,252 00:23:09,452	EU NÃO VOU CONSEGUIR VIVER SEM O INÁCIO.	2	39c/2,2s	Alta
5	128	00:27:22,006 00:27:27,210	PÔ, TU ACHA QUE EU VOU VER TU FAZENDO BURRADA E NÃO VOU CAGUETAR PRA TUA MAE?!	3	76c/5,2s	Baixa
5	147	00:28:23,033 00:28:25,935	[LIARA] EU ACABEI DE FECHAR UMA EXPOSIÇÃO DE FOTOS DO VICTOR HAIM.	3	64c/2,9s	Alta
5	168	00:29:17,588 00:29:21,591	[SÔNIA] IMAGINA, QUERIDA! OLHA, A GENTE TÁ SENTINDO MUITO A SUA FALTA...	3	70c/4s	Alta
5	170	00:29:25,630 00:29:28,498	[SÔNIA] INCLUSIVE, AMANHÃ, NÓS VAMOS FAZER UMA HOMENAGEM A VOCÊ.	3	62c/2,8s	Alta
5	174	00:29:37,875 00:29:40,843	VOU PEDIR PRA MINHA PERSONAL CURICA FAZER, ASSIM, UMAS IGUARIAS...	3	64c/2,9s	Alta
5	197	00:30:58,622 00:31:00,390	[LAÉRCIO] EU TINHA ESQUECIDO DESSE DETALHE.	2	42c/1,7s	Alta
5	220	00:32:14,432 00:32:16,032	O INÁCIO FALOU QUE VAI FICAR DORMINDO NA CASA DA TIA DELE...	2	59c/1,6s	Alta
5	221	00:32:16,099 00:32:17,934	EU NÃO QUERO CORRER O RISCO DE ENCONTRAR.	2	40c/1,8s	Alta
5	310	00:37:28,045 00:37:29,678	VOCÊ ESTAVA QUERENDO FALAR COMIGO, NÉ?	2	37c/1,6s	Alta
5	312	00:37:32,349 00:37:35,885	[PENHA] É, LYGIA. EU PRECISO CONVERSAR COM VOCÊ SOBRE...	2	55c/3,5s	Baixa

ANEXO B – SINTÁGMA VERBAL (VERBO + SINTÁGMA ADVERBIAL)

CAP	LEG	TCR	<SV_verbo+SAdv>	L	CPL/S	Veloc
1	25	00:25:26,657 00:25:29,592	SABE O QUE É? A GENTE QUERIA MUITO UM AUTÓGRAFO SEU	2	51c/2,9s	Alta
1	159	00:34:05,009 00:34:06,643	AH, MAS ISSO NÃO VAI FICAR ASSIM, NÃO, AÍ.	2	41c/1,6s	Alta
2	168	00:17:40,258 00:17:43,527	O QUÊ QUE É? O QUÊ QUE DEU AÍ, QUE TÁ	2	36c/3,2s	Baixa
2	264	00:27:26,677 00:27:31,681	CREDO! AQUELE RAPAZ ALI, NÃO É O ENTREGADOR DO SEU MESSIAS?	2	58c/5s	Baixa
3	13	00:22:19,938 00:22:23,740	CLARO, NÉ? QUE A ISADORA NÃO IA QUERER QUE O MUNDO VISSE QUE ELA É MENOS QUE PERFEITA, NE?	3	88c/3,8s	Alta
3	33	00:23:18,296 00:23:20,364	[ELANO] EU NÃO POSSO TE CONTAR AGORA.	2	36c/2s	Alta
3	40	00:23:40,818 00:23:43,720	SAIR CORRENDO ATRÁS DAS EMPREGUETES	2	35c/2,9s	Baixa
3	116	00:27:48,566 00:27:51,268	PORQUE O SERVIÇO DEMOROU MAIS PORQUE A GENTE ACHOU TAMBÉM QUE IA MAIS SIMPLES...	3	78c/2,7s	Alta
3	185	00:31:00,858 00:31:03,793	[VALDA] MAS, FILHA, EU NÃO VOU MORAR COM VOCÊ DE FAVOR.	2	54c/2,9s	Alta
3	292	00:37:51,101 00:37:53,302	[SANDRO] PÔ, CASA DE RICO TEM SEMPRE UMA COISINHA GOSTOSA PRA COMER.	3	66c/2,2s	Alta
4	5	00:17:37,221 00:17:39,590	[SANDRO] ELE DEVE TÁ ESTOURANDO AÍ, NO ELEVADOR, RAPAZ. TÁ CHEGANDO AÍ.	3	69c/2,3s	Alta
4	22	00:18:17,862 00:18:22,933	[ALANA] OI, GENTIL! NÃO. NAO DÁ PRA PENHA ATENDER AGORA, NÃO. ELA ESTÁ NO BANHO.	3	78c/5s	Baixa
4	163	00:26:28,686 00:26:33,957	PODEM ME SEGUIR, AFINAL, NÃO É TODO DIA QUE A GENTE RECEBE CELEBRIDADES COMO VOCÊS...	3	83c/5,2s	Baixa
4	173	00:27:16,967 00:27:18,935	[INÁCIO] MAS EU NÃO SEI DANÇAR DIREITO, ROSÁRIO. PERAÍ.	2	54c/1,9s	Alta
4	310	00:34:24,562 00:34:26,729	[ELANO] EU SEMPRE QUIS SER ADVOGADO.	2	35c/2,1s	Média
4	325	00:35:01,732 00:35:05,101	ESSA ZOEIRA TODA. DEVE GOSTAR ASSIM, MAIS DA TRANQUILIDADE, NÉ? DO SOSSEGO.	3	73c/3,3s	Alta
4	336	00:35:32,830 00:35:37,700	[ROSÁRIO] UÊ... AGORA NÃO TÁ AQUI, SÓ EU E VOCÊ? HUM?	2	54c/4,8s	Baixa
5	340	00:35:47,044 00:35:52,115	A GENTE PODIA ATÉ FINGIR QUE A NOITE SÓ VAI COMEÇAR AGORA. HUM? QUE TAL?	3	70c/5s	Baixa
5	99	00:25:30,461 00:25:32,896	[LAÉRCIO] SOCORRO, VOCÊ ACHA REALMENTE QUE A CHAYZINHA VAI CAIR NESSE TRUQUE?	3	75c/2,4s	Alta
5	223	00:32:20,638 00:32:22,705	[ROSÁRIO] NÃO, NAO PRECISA, NÃO. NAO PRECISA, NÃO. EU NAO	3	87c/2s	Alta

			VOU DEMORAR. EU VOU NUM MINUTO.			
5	308	00:37:18,435 00:37:22,205	[LYGIA] O RAPAZ QUE ME ENTREVISTOU DISSE QUE GOSTOU BASTANTE DO MEU CURRÍCULO.	3	76c/3,7s	Alta

ANEXO C – SINTÁGMA VERBAL
(VERBO + PRONOME OBLÍQUO + SINTÁGMA ADVERBIAL)

Cap	leg	TCR	<SV_(verbo)+oblíquo+SV>	L	Cpl/s	Velocidade
1	133	00:32:38,489 00:32:43,493	[PENHA] Ó, GENTIL, TU PODE ME CHAMAR SÓ DE PENHA, TÁ? EU ATÉ PREFIRO.	3	67c/5s	Baixa
2	96	00:09:31,170 00:09:42,580	A CIDA VAI ESTUDAR, VAI FAZER VESTIBULAR, VAI SE FORMAR E VAI SER DOUTORA.	3	72c/11,4s	Baixa
2	146	00:24:01,806 00:24:08,312	[ISADORA] ELA VEIO ME PROVOCAR, MÃE! E SÓ DE OLHAR PRA CA	3	55c/2,8s	Alta
2	240	00:24:01,806 00:24:08,312	[OTTO] POR FAVOR, VOCÊ PODE NOS DEIXAR SOZINHOS, QUE EU PRECISO FALAR COM O MEU FILHO?	3	84c/6,5s	Baixa
2	245	00:24:50,554 00:25:05,068	OU ME CASAR, EU ESTAVA ME DESPEDINDO DA BALADA. UÉ, QUAL O PROBLEMA?	3	66c/14,5s	Baixa
2	400	00:40:29,192 00:40:32,361	DA VAI TE SURPREENDER MUITO, MEU GENRO.	2	38c/3,1s	Baixa
3	22	00:22:45,529 00:22:47,530	[ELANO] ADOREI TE VER TIRANDO O UNIFORME.	2	40c/2s	Alta
3	201	00:31:56,547 00:31:58,615	OBRIGADA POR TER ME AVISANDO, VIU?	2	33c/2s	Média
4	69	00:21:17,008 00:21:18,942	[PENHA] NÃO ESTAVA ME RECONHECENDO?	2	34c/1,9s	Alta
5	58	00:23:02,747 00:23:05,148	[SIDNEY] TENTA SE ACALMAR, POR FAVOR. CALMA!	2	43c/2,4s	Alta
5	161	00:28:55,032 00:28:57,166	MAS É QUE OS PAPARATOS ME SEGUEM POR TUDO QUANTO É LUGAR!	2	56c/2,1s	Alta
5	280	00:35:25,656 00:35:28,858	QUE EU TENHO CERTEZA QUE VAI ME FAZER MUITO, MUITO, MUITO FELIZ.	2	63c/3,2s	Alta
5	308	00:37:18,435 00:37:22,205	[LYGIA] O RAPAZ QUE ME ENTREVISTOU DISSE QUE GOSTOU BASTANTE DO MEU CURRÍCULO.	3	76c/3,7s	Alta

ANEXO D – SINTÁGMA VERBAL (VERBO + SINTÁGMA PREPOSICIONAL)

cap	leg	TCR	<SV_verbo+SP>	L	cpl	Velocidade
2	13	00:00:53,218 00:00:54,920	[INACIO] DA PRA ESSE ASSUNTO, ROSÁRIO?	2	32c/1,7s	Alta
3	16	00:22:28,012 00:22:29,980	VENDO VOCÊ ABANDONAR O EMPREGO, SAINDO DE LIMUSINE, VOCÊ VIU, NÉ?	3	63c/1,9s	Alta
3	130	00:28:19,897 00:28:22,199	[ALEJANDRO] O QUÊ? MAS GENTE, EU ESTOU NA MINHA CASA, AONDE QUE EU TÔ?	3	68c/2,3s	Alta

ANEXO E – SINTAGMA NOMINAL (ESPECIFICADOR + SINTÁGMA NOMINAL)

CAP	LEG	TCR	<SN_especif+SN>	L	CPL	VEL
1	175	00:35:09,306 00:35:13,81	MARIA DA PENHA, ESSA SUA BELEZA ME DEIXA AFÔNICO.	2	60c/4,5s	Baixa
2	12	0:00:48,948 00:00:53,151	[INACIO] EVENTOS POR DIA, VENDENDO MIL QUENTINHAS. MAS ISSO, NÃO.	3	58c/4,2s	Baixa
2	112	00:10:51,116 00:10:54,552	E SERVIR NO CASAMENTO DAS MENINAS VAI SER A ÚLTIMA HUMILHAÇÃO QUE EU PASSO AQUI.	3	78c/3,4	Alta
2	144	00:15:00,164 00:15:09,774	[CIDA] NEM VOCÊ SABE ONDE O SEU NOIVO TÁ? EU QUE VOU SABER?	2	58c/9,6s	Baixa
2	150	00:15:30,929 00:15:35,900	HOJE É O SEU CASAMENTO COM UM RAPAZ QUE É UM VERDADEIRO PRÍNCIPE!	3	63c/4,9s	Baixa
2	154	00:15:58,623 00:16:11,068	VOCÊ NÃO É MAIS UMA MENINA. VOCÊ É UMA MULHER INTELIGENTE, ESPERTA...	3	67c/12,4	Baixa
2	399	00:40:25,722 00:40:29,125	[OTTO]NUNCA PENSEI QUE O MEU FILHO FOSSE SE CASAR TÃO CEDO.	3	67c/12,4s	Baixa
3	28	00:23:02,280 00:23:06,616	VOCÊ SE LIVROU DESSE CARA, NA BOA. VOCÊ PODE TER CERTEZA QUE A ISADORA VAI SOFRER NA MÃO DELE.	3	92c/4,3s	Alta
3	38	00:23:32,476 00:23:37,847	DEU TUDO ERRADO! ESSES SEUS AMIGOS, CONRADO, SINCERAMENTE!	2	57c/5,3s	Baixa
3	43	00:23:51,229 00:23:53,663	EU VOU TER QUE PASSAR A MINHA LUA DE MEL VENDENDO CACHOEIRA	2	56c/2,4s	Alta
3	79	00:26:20,211 00:26:23,580	PRA QUE ESSE CAPRICHOS TODO AÍ? DORMIU NA BECA?	2	45c/3,3s	Baixa
3	171	00:30:28,426 00:30:30,159	[CIDA]PODE, OBRIGADA. PARECE QUE ESTAVA TODO MUNDO ESPERANDO ISSO.	3	65c/1,7s	Alta
3	322	00:39:35,105 00:39:37,573	[ROSÁRIO] TA BOM, SEU EXAGERADO! DEIXA EU IR PRA ESSA SESSÃO DE FOTOS	3	68c/2,4s	Alta
3	278	00:37:16,333 00:37:19,668	[PENHA] AI, TÔ NERVOSA COM ESSA SESSÃO DE FOTO, ASSIM, PORQUE EU NÃO SEI NEGOCIO DE POSE, SABE?	3	93c/3,3s	Alta
3	288	00:37:42,359 00:37:43,893	NÃO ME METE NOS SEUS ROLOS	2	19c/1,5s	Alta
3	295	00:38:03,413 00:38:05,848	[DINHA]VOCÊS NÃO CONHECEM AS MALUCAS? NEM PENSAM NO BUFFET	2	59c/2,4s	Alta
4	6	00:17:39,658 00:17:42,927	VOU TE FALAR UM NEGÓCIO, GENTIL, PARECE QUE ESSA NOTÍCIA É UMA BOMBA DE ARRASAR COM QUARTEIRÃO.	3	93c/3,2s	Alta
4	79	00:21:48,306 00:21:52,209	[PENHA] ESTAVA ESPERANDO UMA PESSOA. ACHO QUE LEVEI BOLO.	3	65c/3,9s	Média

			QUE ÓDIO!			
4	111	00:23:38,016 00:23:39,983	[ALANA]GENTE, ATENDI OUTRO TELEFONEMA QUE ERA ENGANDO, HOJE.	2	59c//1,9s	Alta
4	113	00:23:41,852 00:23:44,821	[ROSÁRIO] DEVEM TER ANOTADO O NÚMERO ERRADO. BOM, CIDA,VAMOS!	2	61c/2,9s	Baixa
4	132	00:24:45,683 00:24:47,884	COMO É QUE FALA O NOME DAQUELA MULHER, PELO AMOR DE DEUS?	2	56c/2,2s	Alta
4	174	00:27:19,003 00:27:21,505	[ROSÁRIO] IH, EU SOU UMA ÓTIMA PROFESSORA, TÁ, MEU FILHO? APROVEITA. VEM!	3	71c/2,5s	Alta
4	200	00:28:55,966 00:28:58,468	[KLEITON] Ô, DESCULPA O PALAVRÃO AÍ, CIDA. MAS É QUE O CARA E O MAIOR MALA, NÃO É, NÃO?	3	85c/2,5s	Alta
4	208	00:29:15,018 00:29:17,254	[ELANO]BATATA! BATATA! PEGA A BATATA, POR FAVOR. BATATA FRITA. OBRIGADA, KLEITINHO.	3	82c/2,2s	Alta
4	394	00:38:37,114 00:38:43,419	COMER AQUI, NO CHOPEOKÊ. TEM UM BOLINHO DE BACALHAU, DR. OTTO! ASSIM, SEQUINHO, SABE?	2	83c/6,3	Baixa
10	68	00:23:30,942 00:23:35,012	[TOM] RELAXA, FABIAN! DEVE ESTAR ATÉ AGORA BRIGANDO COM AQUELE OGRO DAQUELE NAMORADO DELA.	3	88c/4s	Alta
10	125	00:27:10,294 00:27:14,964	A TUA MÃE. E SÓ NÃO É ELA OU O SEU ALEJANDRO QUE TÁ AQUI PORQUE EU NÃO CONSEGUI FALAR COM ELES.	3	93c/4,6s	Alta
10	5	00:20:41,172 00:20:46,343	VAMOS NOS REFERIR A ESSA ESQUEMA TODO COMO COMO OPERAÇÃO B-A-BÁ	2	63c/5,1s	Baixa
10	129	00:27:27,278 00:27:29,712	SERIA ATÉ IRRESPONSABILIDADE MINHA NÃO FAZER ISSO.	2	49c/2,4s	Alta
10	248		[ISADORA] QUELLA TERRA DI GENTE ELEGANTE...	2	42c/2,8s	Baixa
10	137	00:27:57,741 00:28:02,445	NÃO TÁ NEM 100% AINDA. ELA NÃO PODE TER ESSE TIPO DE ABORRECIMENTO, NÃO.	3	70c/4,7s	Baixa

ANEXO F – SINTAGMA NOMINAL (SUBSTANTIVO + SINTÁGMA PREPOSICIONAL)

Cap	leg	TCR	<SN subst+ SP>	L	cpl	Velocidade
1	18	00:25:07,372 00:25:10,407	[SANDRO] AÍ, ESSE ALMOFADINHA DO GENTIL PODE PILOTAR O CARRÃO QUE ELE QUISER, MOROU?	3	82c/3	Alta
1	120	00:31:40,831 00:31:45,769	EU ESTAVA COMEMORANDO. ARQUIVARAM O CASO DAS AÉREAS, DR. ELANO!	3	61c/4,9s	Baixa
2	176	00:35:13,877 00:35:17,546	[SARMENTO] MEU AMOR, É UM VERDADEIRO ESTOURO DA BOIADA.	3	53c/3,6s	Baixa
2	151	00:15:35,967 00:15:41,505	FILHO DE UM MAGNATA DA CONSTRUÇÃO CIVIL! O QUÊ QUE TÁ PEGANDO, ISADORA?	3	69c/5,5s	Baixa
2	178	00:18:22,567 00:18:28,572	[PENHA] AH, BOM! ACHEI QUE A SENHORA IA FALAR O NOME DA OUTRA MOÇA QUE TRABALHAVA AQUI.	3	85c/6s	Baixa
2	249	00:25:17,848 00:25:26,423	[CONRADO] FOI UMA DESPEDIDA DE SOLTEIRO, IGUAL A TODO CARA QUE VAI SE CASAR VAI.	3	76c/8,5s	Baixa
2	300	00:30:34,665 00:30:37,801	[BRUNESSA] E TU ACHA QUE É A CASA DOS DONOS DA GALERIE?	2	50c/3,1s	Média
2	352	00:36:15,839 00:36:19,909	[PENHA] ESSE DINHEIRO É DIREITO DO PATRICK, SANDRO. TU É PAI!	2	60c/4s	Baixa
3	42	00:23:47,258 00:23:51,161	AGORA PRA FECHAR COM CHAVE DE OURO, EM VEZ DE MILÃO...	2	53c/3,9s	Baixa
3	184	00:30:57,555 00:31:00,790	OLHA, EU TÔ LÁ NA CASA DA PENHA, JÁ TÔ PROCURANDO UM CANTINHO PRA GENTE MORAR.	3	76c/3,2	Alta
3	333	00:40:34,731 00:40:39,501	AS EMPREGUETES GANHARAM O PRÊMIO "A MAIS PEDIDA DE TRÊS" DESSE BIMESTRE.	3	81c/4,7s	Alta
4	2	00:17:28,747 00:17:32,215	[GENTIL] PESSOAL! PESSOAL! O OLINTINHO LIGOU PRA SEDE DO VASCO...	3	63c/3,5s	Alta
4	13	00:17:58,376 00:18:01,244	MAS, QUE DEVE SER ALGUMA COISA QUE VAI TE DAR UM RECORDE DE AUDIÊNCIA...	3	70c/2,8s	Alta
4	48	00:19:47,251 00:19:51,121	NÓS ESTAMOS MARCANDO UMA NOVA REUNIÃO DA LIGA DAS PATROAS DO CASA GRANDE.	3	71c/3,8s	Alta
10	77	00:24:07,545 00:24:09,779	PORQUE DAQUI A POUCO, O ANIMAL DO INÁCIO SAI DE CENA...	2	54c/2,2s	Alta
10	85	00:24:35,806 00:24:39,042	ELA TÁ COM O MESMO CORPINHO DE SEREIA DE TODA A VIDA.	2	52c/3,2s	Média
10	169	00:29:21,659 00:29:25,562	LÁ NAS REUNIÕES DA LIGA DAS PATROAS, NÃO É, FILHA? [ARIELA] É, SIM, CHAY.	3	71c/3,9s	Alta
10	175	00:29:40,911 00:29:43,146	UMAS COISAS DE MINHA TERRA, UMA DELÍCIA!	2	39c/2,2s	Alta

10	244	00:33:14,625 00:33:17,494	[CONRADO] DIAMANTES, AMOR. UMA LEMBRANCINHA DE VIAGEM, NADA DEMAIS.	3	65c/2,8s	Alta
10	251	00:33:29,039 00:33:30,807	[CONRADO] TERRA DE GENTE ELEGANTE.	2	33c/1,7s	Alta

ANEXO G – SINTAGMA NOMINAL (SUBSTANTIVO + SINTAGMA ADJETIVAL)

Cap	leg	TCR	<SN subst+ SP>	L	cpl	Velocidade
1	18	00:25:07,372 00:25:10,407	[SANDRO] AÍ, ESSE ALMOFADINHA DO GENTIL PODE PILOTAR O CARRÃO QUE ELE QUISER, MOROU?	3	82c/3	Alta
1	120	00:31:40,831 00:31:45,769	EU ESTAVA COMEMORANDO. ARQUIVARAM O CASO DAS AÉREAS, DR. ELANO!	3	61c/4,9s	Baixa
2	176	00:35:13,877 00:35:17,546	[SARMENTO] MEU AMOR, É UM VERDADEIRO ESTOURO DA BOIADA.	3	53c/3,6s	Baixa
2	151	00:15:35,967 00:15:41,505	FILHO DE UM MAGNATA DA CONSTRUÇÃO CIVIL! O QUÊ QUE TÁ PEGANDO, ISADORA?	3	69c/5,5s	Baixa
2	178	00:18:22,567 00:18:28,572	[PENHA] AH, BOM! ACHEI QUE A SENHORA IA FALAR O NOME DA OUTRA MOÇA QUE TRABALHAVA AQUI.	3	85c/6s	Baixa
2	249	00:25:17,848 00:25:26,423	[CONRADO] FOI UMA DESPEDIDA DE SOLTEIRO, IGUAL A TODO CARA QUE VAI SE CASAR VAI.	3	76c/8,5s	Baixa
2	300	00:30:34,665 00:30:37,801	[BRUNESSA] E TU ACHA QUE É A CASA DOS DONOS DA GALERIE?	2	50c/3,1s	Média
2	352	00:36:15,839 00:36:19,909	[PENHA] ESSE DINHEIRO É DIREITO DO PATRICK, SANDRO. TU É PAI!	2	60c/4s	Baixa
3	42	00:23:47,258 00:23:51,161	AGORA PRA FECHAR COM CHAVE DE OURO, EM VEZ DE MILÃO...	2	53c/3,9s	Baixa
3	184	00:30:57,555 00:31:00,790	OLHA, EU TÔ LÁ NA CASA DA PENHA, JÁ TÔ PROCURANDO UM CANTINHO PRA GENTE MORAR.	3	76c/3,2	Alta
3	333	00:40:34,731 00:40:39,501	AS EMPREGUETES GANHARAM O PRÊMIO "A MAIS PEDIDA DE TRÊS" DESSE BIMESTRE.	3	81c/4,7s	Alta
4	2	00:17:28,747 00:17:32,215	[GENTIL] PESSOAL! PESSOAL! O OLINTINHO LIGOU PRA SEDE DO VASCO...	3	63c/3,5s	Alta
4	13	00:17:58,376 00:18:01,244	MAS, QUE DEVE SER ALGUMA COISA QUE VAI TE DAR UM RECORDE DE AUDIÊNCIA...	3	70c/2,8s	Alta
4	48	00:19:47,251 00:19:51,121	NÓS ESTAMOS MARCANDO UMA NOVA REUNIÃO DA LIGA DAS PATROAS DO CASA GRANDE.	3	71c/3,8s	Alta
10	77	00:24:07,545 00:24:09,779	PORQUE DAQUI A POUCO, O ANIMAL DO INÁCIO SAI DE CENA...	2	54c/2,2s	Alta
10	85	00:24:35,806 00:24:39,042	ELA TÁ COM O MESMO CORPINHO DE SEREIA DE TODA A VIDA.	2	52c/3,2s	Média
10	169	00:29:21,659 00:29:25,562	LÁ NAS REUNIÕES DA LIGA DAS PATROAS, NÃO É, FILHA? [ARIELA] É, SIM, CHAY.	3	71c/3,9s	Alta
10	175	00:29:40,911 00:29:43,146	UMAS COISAS DE MINHA TERRA, UMA DELÍCIA!	2	39c/2,2s	Alta

10	244	00:33:14,625 00:33:17,494	[CONRADO] DIAMANTES, AMOR. UMA LEMBRANCINHA DE VIAGEM, NADA DEMAIS.	3	65c/2,8s	Alta
10	251	00:33:29,039 00:33:30,807	[CONRADO] TERRA DE GENTE ELEGANTE.	2	33c/1,7s	Alta

ANEXO H – SINTAGMA NOMINAL (NÚCLEO)

CAP	LEG	TCR	<SN_NÚCLEO>	L	CPL	VELOC
1	134	00:32:43,561 00:32:46,529	GENTIL] AH, EU ACHO MARIA DA PENHA UM NOME TÃO BONITO.	2	54c/2,9s	Alta
3	12		[SÔNIA] ESSA DOR DE CABEÇA VAI ME MATAR.	3	39c/2,1s	Alta
4	307	00:34:16,053 00:34:18,320	[ELANO] ELANO, DO ESCRITÓRIO SARMENTO. EU TRABALHO	2	53c/2,2s	Alta
10	263	00:33:56,333 00:33:58,735	EU PRECISO FALAR COM A DRA. LYGIA. TEM QUE SER PESSOALMENTE.	2	59c/2,4s	Alta
10	300	00:25:07,372 00:25:10,407	COMO QUE FOI A VOLTA DA LUA DE MEL? ME CONTA.	2	44c/2s	Alta

ANEXO I – SINTAGMA ADJETIVAL

CAP	LEG	TCR	< SAdj >	L	CPL/S	VELOC
1	209	00:37:15,199 00:37:18,768	ELAS ATÉ PODEM SER FORTES JUNTAS! MAS, SEPARADAS, ELAS NÃO VALEM NADA!	3	46c/3,3s	Alta
3	90	00:26:51,242 00:26:55,645	A DONA LYGIA NÃO. TÁ DEMORANDO PRA FICAR BOA. TÔ ACHANDO TÃO ESQUISITO. TU VAI LÁ AGORA?	3	86c/4,4s	Alta

ANEXO J – SINTAGMA ADVERBIAL

CAP	LEG	TCR	<SAdv_ adv+adv	L	CPL/S	VELOC
3	34	00:23:20,431 00:23:24,634	MAS DAQUI A POUCO... DAQUI A POUCO, VOCÊ VAI PERCEBER QUE ESSE CARA NÃO PRESTA.	3	77c/4,2s	Alta
3	35	00:23:24,702 00:23:26,937	E VOCÊ SE DEU MUITO BEM EM SE LIVRAR DELE.	2	41c/2,2s	Alta
3	114	00:27:43,661 00:27:46,329	[ALEJANDRO] É, DEPOIS DE 4 DIAS, E UMA PAREDE TOTALMENTE DESTRUÍDA.	3	65c/2,6s	Alta
10	119	00:26:54,912 00:26:56,513	É SÓ A EMPREGADINHA LÁ EM CASA, VOCÊ NÃO TEM NADA VER COM A MINHA VIDA, NÃO.	3	76c/1,6s	Alta
10	126	00:27:15,032 00:27:17,300	SAMUEL] AINDA POR CIMA É DEDO DURO.	2	35c/2,2s	Baixa
10	311	00:37:29,747 00:37:32,281	EU VI QUE VOCÊ LIGOU, MAS EU ESTAVA BEM NO MEIO DA ENTREVISTA.	2	50c/3,3s	Baixa

ANEXO K – SINTAGMA PREPOSICIONAL

CAP	LEG	TCR	<PROSEGL><SP	L	CPL/S	VELOC
1	58	00:27:03,887 00:27:08,758	Ó, A MAMÃE SÓ VAI JANTAR COM O GENTIL E DAQUI A POUCO ESTÁ EM CASA, TA BOM?	3	73c/4,8s	Baixa
1	204	00:37:02,62 00:37:04,91	[LAÉRCIO] CHAYZINHA, VOCÊ VAI VOLTAR COM ESSA MALUQUICE , DE NOVO?	3	63c/2,2s	Alta
1	71	00:27:33,184 00:27:35,152	Ó, NÃO TÔ MUITO ACOSTUMADA COM ESSE NEGÓCIO DE RESTAURANTE,NÃO.	2	62c/1,9s	Alta
1	229	00:38:26,303 00:38:28,871	[ROSÁRIO] VOCÊ TÁ COM UMA CARINHA ABATIDA, INÁCIO.	2	49c/2,5s	Alta
2	22	00:01:39,632 00:01:42,166	[LAÉRCIO] CHAYZINHA, EU ATÉ IMPLICO COM A COMPRIDONA , TAMBÉM.	3	59c/2,5s	Alta
2	106	00:10:10,142 00:10:23,354	[SÔNIA] FILHA, PÃO BRINCA COM A ARIELA DE OUTRA COISA?	2	53c/13,2s	Baixa
2	112	00:10:51,116 00:10:54,552	E SERVIR NO CASAMENTO DAS MENINAS VAI SER A ÚLTIMA HUMILHAÇÃO QUE EU PASSO AQUI.	3	78c/3,4s	Alta
2	178	00:18:22,567 00:18:28,572	PENHA] AH, BOM! ACHEI QUE A SENHORA IA FALAR O NOME DA OUTRA MOÇA QUE TRABALHAVA AQUI.	3	85c/6s	Baixa
2	252	00:25:35,966 00:25:41,104	ALGUMA VEZ VOCÊ JÁ OLHOU PRO LADO E PENSOU EM ALGUMA COISA HÃ?	3	60c/5,1s	Baixa
3	28	00:23:02,280, 00:23:06,616	VOCÊ SE LIVROU DESSE CARA, NA BOA . VOCÊ PODE TER CERTEZA QUE A ISADORA VAI SOFRER NA MÃO DELE.	3	92c/4,3s	Alta
3	72	00:25:58,322 00:26:02,426	QUE EU TENHO QUE ARRUMAR TUDO NA CASA DA DONA LYGIA , DEPOIS IR ÁA PRA AQUELA SESSÃO DE FOTO NOSSA.	3	96c/4,1s	Alta
3	186	00:31:03,861 00:31:07,029	EU VOU CONTINUAR MORANDO LÁ EM CASA , ONDE EU VIVI A VIDA TODA.	2	61c/3,1s	Alta
3	225	00:33:54,197 00:33:57,000	ME DESAFIA, MARIA APARECIDA, PRA VOCÊ VER DO QUE EU SOU CAPAZ!	2	61c/2,8s	Alta
3	279	00:37:19,737 00:37:22,371	TEM UMA POSE QUE EU LEMBRO DO DISCO DA XUXA , QUE EU ADORO.	2	57c/2,6s	Alta
4	39	00:19:07,812 00:19:14,018	[MÁSVOLA] ESPLÊNDIDO! VAI HARMONIZAR PERFEITAMENTE COM O JANTAR QUE SÔNIA NOS PREPAROU.	3	85c/6,2s	Baixa
4	70	00:21:19,010 00:21:22,246	[OTTO] DESCULPA. É QUE DAS OUTRAS VEZES QUE EU TE VI, VOCÊ ESTAVA DE UNIFORME.	3	76c/3,2s	Alta
4	107	00:23:24,135 00:23:29,106	DESISTIU DE DAR O FURO. Ô GENTIL, PÔ , FOI MAL! VALEU PELOS BISCOITOS E PELOS SALGADINHOS!	3	87c/4,9s	Alta
4	115	00:23:47,091 00:23:51,661	[CIDA] AI, NÃO SEI SE DAR PRA DANÇAR E SEGURAR VELA AO MESMO TEMPO.	3	65c/4,5s	Baixa

4	116	00:23:51,729 00:23:54,998	[ROSÁRIO] DEIXA DE SER BOBA, VAI PÔ, A PENHA SAIU PRA COMEMORAR COM O GENTIL.	3	76c/3,2s	Alta
4	166	00:26:38,529 00:26:41,231	[ROSÁRIO] NÃO, NAO, NÃO. PARA, PARA, PARA! PELO AMOR DE DEUS! AGORA A GENTE ESTÁ AQUI!	3	84c/2,7s	Alta
4	227	00:30:03,601 00:30:07,003	AGORA, VOCÊ PODERIA PARAR DE ME CHAMAR DE DOUTOR, DE SENHOR?	2	59c/3,4s	Alta
4	322	00:34:51,421 00:34:56,492	[PENHA] AH, NÃO TÔ FALANDO PRO SENHOR? AI, SENHOR, NÃO. VOCÊ.	2	60c/5s	Baixa
4	355	00:36:38,762 00:36:41,497	PRA FAZER A MACHADÃO, NAO. PRA FAZER A PRÓPRIA GABRIELA, É.	2	58c/2,7s	Alta
4	391	00:38:27,704 00:38:31,240	[PENHA] ESSA PORTA TAMBÉM TÁ COM UM PROBLEMINHA. A GENTE TÁ PRA CONSERTAR. NÃO CONSERTO NUNCA.	3	92c/3,5s	Alta
10	46	00:22:23,942 00:22:27,044	EU QUERO VIRAR ESSA PÁGINA NA MINHA VIDA DE UMA VEZ POR TODAS!	2	61c/3,1s	Alta
10	115	00:26:39,497 00:26:43,567	[PENHA] Ó, TÔ NA TUA COLA PRA NÃO DEIXAR TU FAZER BESTEIRA!	2	58c/4s	Baixa

ANEXO L – ORAÇÕES COORDENADAS

CAP	LEG	TCR	<PROSEGL><COORD>	Linhas	CPL/S	VELOC
1	78	00:26:17,908 00:26:20,143	[PENHA] IH, MAS TÁ FALANDO ATÉ INGLÊS! [ALANA] NOSSA!	33	51c/2,2s	Alta
3	32	00:23:16,961 00:23:18,228	[CIDA] POR QUE VOCÊ ESTÁ FALANDO ISSO?	2	37c/1,2s	Alta
3	120	00:27:57,408 00:27:58,908	SANDRO] PÔ, MAS SENHORA PODE FICAR TRANQUILA.	2	47c/1,5s	Alta
3	202	00:31:58,682 00:32:02,919	[BRANCA] CONTE COMIGO, SÔNIA. PORQUE SE ELAS SÃO UNIDAS, NÓS SOMOS MAIS. VAMOS.	3	77c/4,2s	Alta
4	11	00:17:54,573 00:17:56,073	[GENTIL] ESTRANHO, NÉ? PORQUE O JUNINHO É UM SUJEITO DISCRETO, NÃO SE METE EM CONFUSAO.	3	85c/1,5s	Alta
3	141	00:28:49,460 00:28:51,695	[LYGIA] AGORA QUE VOCÊ CHEGOU, TÁ MELHOR.	2	40c/2,2s	Alta
3	142	00:28:51,762 00:28:54,130	EU VOU LÁ PRA DENTRO, QUE EU TÔ A MINHA CABEÇA ESTOURANDO	2	57c/2,3s	Alta
4	11	00:17:54,573 00:17:56,073	GENTIL] ESTRANHO, NÉ? PORQUE O JUNINHO É UM SUJEITO DISCRETO NÃO SE METE EM CONFUSAO.	3	85c/1,5s	Alta
4	151	00:25:58,056 00:26:01,891	[CIDA] PIOR QUE NÃO. EU IA JANTAR COM AS MENINAS, MAS AÍ FOI CADA UMA PRA UM LADO...		82c/3,8s	Alta
4	160	00:26:18,276 00:26:21,744	[HOSTESS] AH, QUE BOM QUE VOCÊS VIERAM! MAS, CADÊ AS OUTRAS EMPEGUETES?	3	69/3,4s	Alta
4	185	00:28:13,624 00:28:17,527	NORMALMENTE, EU SOU MUITO COMEDIDO, MAS QUANDO A GENTE SE APAIXONA, NÃO É?	3	72c/3,9s	Alta
4	278	00:32:50,534 00:32:52,469	QUE OU PARECE O FABIAN, OU EMPREGUETE, OU OS DOIS JUNTOS.	2	126/1,9s	Alta
4	67	00:23:28,739 00:23:30,874		3	65c/2,1s	Alta

ANEXO M – ORAÇÕES SUBORDINADAS

Cap	leg	TCR	<SUBORD conj/pron_rel+oração>	L	Cpl/S	Veloc
1	112	00:31:21,779 00:31:23,813	[SARMENTO] O QUÊ QUE VOCÊ TÁ FAZENDO AQUI, NA MINHA CASA, RAPAZ?	3	62c/2s	Alta
1	228	00:38:23,333 00:38:26,235	[INÁCIO] EU TÔ TÃO CANSADO, AMOR, QUE EU ACHO QUE EU ATÉ PERDI A FOME.	3	68c/2,9s	Alta
2	43	00:03:30,944 00:03:34,445	ELAS SABEM QUE A SOCORRO É SUA FÃ... NÚME	3	39c/3,5s	Baixa
2	51	00:04:04,877 00:04:07,612	É PORQUE EU ESCUTO O SEU PROGRAMA, DESDE QUE EU SOU MENINA.	3	57c/2,7s	Alta
2	61	00:04:41,147 00:04:47,819	AH! PODE SER AQUELA MÚSICA... AQUELA MÚSICA ANTIGA, QUE TOCA NA NOVELA DAS OITO?	3	78c/6,6s	Baixa
2	110	00:10:45,710 00:10:48,712	DO JEITO QUE A SENHORA SONHAVA PRA MIM.	2	38c/3s	Baixa
2	129	00:12:30,649 00:12:40,124	[OTTO] AGORA, É ISSO MESMO QUE EU ENTENDI? O CONRADO SUMIU?	2	58c/9,4s	Baixa
2	135	00:14:20,959 00:14:24,928	PORQUE EU DETESTO ADA. E EU TENHO QUE MOSTRA PRO SIDNEY OS ARMÁRIOS?	2	62c/3,9s	Baixa
2	178	00:18:22,567 00:18:28,572	[PENHA] AH, BOM! ACHEI QUE A SENHORA IA FALAR O NOME DA OUTRA MOÇA QUE TRABALHAVA AQUI.	3	85c/6s	Baixa
2	229	00:22:54,939 00:22:58,408	EU POSSO IR AGORA MESMO! EU JÁ PAGUEI TUDO QUE EU TINHA QUE PAGAR NESSA CASA!	3	75c/3,4s	Alta
2	381	00:38:59,703 00:39:02,238	[CIDA] A DIFERENÇA É QUE EU NÃO ESTAVA DE UNIFORME, NÉ?	2	54c/2,5s	Alta
3	13	00:22:19,938 00:22:23,740	CLARO, NÉ? QUE A ISADORA NÃO IA QUERER QUE O MUNDO VISSE QUE ELA É MENOS QUE PERFEITA, NE?	3	88c/3,8s	Alta
3	28	00:23:02,280 00:23:06,616	VOCÊ SE LIVROU DESSE CARA, NA BOA. VOCÊ PODE TER CERTEZA QUE A ISADORA VAI SOFRER NA MÃO DELE.	3	92c/4,3s	Alta
3	44	00:23:56,901 00:24:00,103	VOU TE FALAR, HEIN! O QUE MAIS PODE ACONTECER DE RUIM?	2	53c/3,2s	Média
3	63	00:25:09,273 00:25:13,410	BOM, AGORA, EU VOU DORMIR, PORQUE SE ME OUVEM FALANDO SOZINHA...	3	62c/4,1s	Baixa
3	125	00:28:08,219 00:28:10,687	VOCÊ SABE O QUE É ESSA BATEÇÃO NA CABEÇA DE UMA PESSOA QUE ESTÁ CONVALESCENDO?	3	76c/2,4s	Alta
3	141	00:28:49,460 00:28:51,695	[LYGIA] AGORA QUE VOCÊ CHEGOU, TÁ MELHOR.	2	40c/2,2s	Alta
3	222	00:33:39,449 00:33:43,653	[GRACINHA] HUM! PODE IR, QUE EU ME GARANTO. BABÁ MELHOR DO QUE EU, ELA NÃO VAI ARRUMAR MESMO!	3	91c/4,2s	Alta

3	285	00:37:35,385 00:37:38,420	EU VOU FALAR COM A ALANA, QUE ELA GOSTA DO CUNHADINHO AQUI, VOU PEDIR AJUDA PRA ELA.	3	82c/3s	Média
3	298	00:38:16,760 00:38:22,698	A CARA DA SÔNIA SARMENTO QUANDO A CIDA TIROU A ROUPA DE EMPREGADA...FOI UM NEGÓCIO Á!	3	83c/5,9s	Baixa
4	51	00:19:59,698 00:20:04,234	[SARMENTO] BOM, EU TÔ VENDENDO QUE ESSE PAPO DE ASSUNTOS DOMÉSTICOS VAI LONGE...	3	75c/4,5s	Média
4	161	00:26:21,812 00:26:24,514	[ROSÁRIO] AH, ELAS ESTAVAM LOUCAS PRA VIR, MAS É QUE NÃO DEU, DESSA VEZ.	3	70c/2,7s	Alta
4	326	00:35:05,168 00:35:07,536	[OTTO] QUEM DISSE QUE EU PREFIRO, PENHA? QUEM DISSE?	2	51c/2,3s	Alta
5	93	00:25:04,702 00:25:08,838	É ISSO! EU TENHO QUE IMPRESSIONAR ESSE HOMEM...	2	46c/4,1s	Baixa
5	97	00:25:22,153 00:25:27,290	E LEVE SUA EMPREGADA JUNTO, QUE AÍ, NINGUÉM VAI DIZER QUE VOCÊ TRATA MAL UMA EMPREGADA...	3	87c/5,1s	Alta
5	125	00:27:10,294 00:27:14,964	A TUA MÃE. E SÓ NÃO É ELA OU O SEU ALEJANDRO QUE TÁ AQUI PORQUE EU NÃO CONSEGUI FALAR COM ELES.	3	93c/4,6s	Alta
5	173	00:29:34,505 00:29:37,807	ENTÃO, FAÇO QUESTÃO QUE A REUNIÃO SEJA EM MINHA CASA.	2	52c/3,3s	Baixa
5	256	00:33:38,616 00:33:40,116	[CONRADO] VOCÊ SABE O EFEITO QUE VOCÊ CAUSA EM MIM QUANDO VOCÊ FALA ITALIANO, NÃO SABE?	3	85c/1,5s	Alta