



FACULDADE DE LETRAS
NINA SOARES LOPES DINIZ

A SEGMENTAÇÃO EM LEGENDAGEM PARA
SURDOS E ENSURDECIDOS:
UM ESTUDO BASEADO EM CORPUS

BELO HORIZONTE
2012

NINA SOARES LOPES DINIZ

A SEGMENTAÇÃO EM LEGENDAGEM PARA
SURDOS E ENSURDECIDOS:
UM ESTUDO BASEADO EM CORPUS

Dissertação apresentada ao Programa de Pós- Graduação em Estudos Lingüísticos da Faculdade de Letras da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Mestre.

Área de concentração: Linguística Aplicada
Linha de Pesquisa: Estudos da Tradução
Orientador: Prof^a. Dr^a. Célia Maria Magalhães
Co-orientador: Prof^a. Dr^a. Vera Lúcia Santiago Araújo

BELO HORIZONTE

FACULDADE DE LETRAS DA UFMG

2012

Dissertação intitulada: **A Segmentação em Legendagem para Surdos e Ensurdecidos: um Estudo Baseado em Corpus**, de autoria da mestranda Nina Soares Lopes Diniz, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof^a. Dr^a. Célia Maria Magalhães (FALE/UFMG)
(Orientadora)

Pro^a. Dr^a. Vera Lúcia Santiago Araújo (Centro de Humanidades/UECE)
(Co-orientadora)

Prof. Dr. Pedro Henrique Lima Praxedes Filho - UECE/UFMG

Prof^a Dr^a. Soraya Ferreira Alves (UnB)

Prof^a. Dr^a. Deise Prina Dutra (FALE/UFMG)

Belo Horizonte
Setembro de 2012

AGRADECIMENTOS

À Universidade Federal de Minas Gerais e à Faculdade de Letras, pela qualidade no ensino.

Ao Programa de Pós-Graduação em Estudos Linguísticos, pela oportunidade e pela qualidade do curso de Mestrado.

Ao CNPq, por financiar parte desta pesquisa.

À minha orientadora, Prof^a. Dr^a. Célia Maria Magalhães por transmitir o saber com tanta dedicação, zelo e paciência.

À minha co-orientadora, Prof^a. Dr^a Vera Lúcia Santiago Araújo por despertar em mim e dividir comigo a paixão e a luta pela acessibilidade.

Ao LETRA por todo suporte a esta pesquisa. E aos colegas de laboratório pelas várias contribuições ao longo do trabalho.

Um agradecimento especial à companhia e amizade de Flávia Souza e Karina Szapk nas disciplinas de mestrado, nas tardes no laboratório, na biblioteca, na cantina, nas madrugadas em frente a trabalhos, provas e “balõezinhos” que nos escapavam.

Aos colegas da Cultura Inglesa pelas trocas de ideias e experiências, e pelo constante incentivo.

Aos meus queridos amigos, os daqui de perto, e os de bem longe, em especial minha irmã Lari, por orar por mim e entender tanta ausência.

À minha querida, grande e solidária família. Aos meus pais que acreditam em mim, mesmo quando eu mesma duvido. Ao Mateus que cuidou de detalhes importantes, ao Esdras que fez tudo ficar mais bonito, ao Lucas por estar sempre pronto a ajudar.

Ao meu amado Jesus, sempre presente, que preparou todo o caminho.

Muito obrigada.

RESUMO

Em comparação com o desenvolvimento de outras vertentes na trajetória dos estudos da tradução, o interesse na linha de pesquisa em legendagem para surdos e ensurdecidos na subárea tradução audiovisual é relativamente recente, porém promissor. Como um ramo dos estudos da legendagem em geral, os estudos em legendagem para surdos e ensurdecidos ainda em estágio primevo de desenvolvimento, fundamentam-se fortemente nas pesquisas conduzidas na subárea da legendagem para ouvintes. Essas pesquisas têm analisado as várias questões técnicas e textuais envolvidas na legendagem, porém o aspecto da segmentação tem recebido pouca atenção. Buscando contribuir com a consolidação desses estudos, esta pesquisa, portanto, dedica-se a investigar questões relevantes da área, concentrando-se especificamente na análise de categorias da segmentação entre linhas praticada nas legendagens para ouvintes e para surdos e ensurdecidos, produzidas em português brasileiro, do filme *Irmãos de Fé*, dirigido por Moacyr Góes, lançado e distribuído no Brasil em 2004 pela *Columbia Pictures*. As perguntas que nortearam este trabalho baseiam-se na investigação dos critérios que podem motivar a decisão do legendista quanto à quebra de linha da legenda, nos casos em que aspectos linguísticos foram desconsiderados; e no exame de desvios linguísticos na legendagem para surdos e ensurdecidos se comparada à legendagem para ouvintes, devido à necessidade de transmissão de informações adicionais na tela na primeira modalidade de legendagem. A incipiência das pesquisas na área, principalmente sobre o aspecto da segmentação justifica esta pesquisa. O aporte teórico parte dos estudos da tradução baseados em corpus em interface com os estudos da tradução audiovisual (Kalantzi, 2008), além da descrição desenvolvida por Lemle (1989) dos sintagmas do português brasileiro para rastrear categorias de segmentação. Os procedimentos metodológicos adotados incluem a análise de dados quantitativos através de ferramentas do software *Wordsmith Tools*© 5.0, a anotação manual e o rastreamento das categorias de segmentação. Os resultados revelaram a preferência do legendista pelo tipo de legendagem que transcreve grande parte do discurso para a tela, a legendagem verbatim. O rastreamento de segmentações revelou que os desvios linguísticos mais recorrentes neste corpus incidiram na estrutura do sintagma verbal e que a preferência por uma linha de legenda superior mais curta que a inferior é o fator que mais competiu com considerações linguísticas na segmentação das legendas. Foram verificadas poucas escolhas de quebra de linha distintas entre modalidades de legendagem, porém nenhuma delas pôde ser atribuída a peculiaridades do tipo de legendagem envolvida na ocorrência. Os resultados apontam, em geral, que as duas modalidades de legendagens pouco se diferenciam entre si tanto em termos técnicos, quanto textuais; e que receberam o mesmo tratamento por parte do legendista.

Palavras-Chave: Estudos da Tradução Baseados em Corpus; Tradução Audiovisual; Legendagem para Surdos e Ensurdecidos; Legendagem para Ouvintes; Segmentação entre Linhas.

ABSTRACT

Compared to the development of other strands in translation studies trajectory, interest in the line of research on subtitling for the deaf and the hard of hearing is relatively recent, but promising. As a branch from subtitling studies, the studies on subtitling for the deaf and the hard of hearing, still in primeval development, rely heavily on research conducted in the subarea of subtitling aimed at hearing viewers. These researches have analyzed various technical and textual issues involved in subtitling, not enough attention, though, has been drawn to the issue of segmentation. Affiliated to this research strand, the present study investigates relevant discussion points in the area, focusing specifically on the examination of line-breakings adopted in the subtitling aimed at hearing viewers and in the subtitling for the deaf and the hard of hearing elaborated in Brazilian Portuguese for the movie *Irmãos de Fé*, directed by Moacyr Góes, and released as well as distributed in 2004 by *Columbia Pictures*. One of the research questions that motivated the study promote the examination of the criteria that may guide decision making regarding subtitles line-breaking, in those cases where linguistic considerations were ignored. The other research question is whether subtitling for the deaf and the hard of hearing shows more linguistic deviations than subtitling aimed at hearing viewers, due to the need of transmitting additional non-verbal information on the screen. The paucity of research in the area, especially on the aspect of segmentation justifies this study. The theoretical framework is that of Corpus-based Translation Studies in interface with Audiovisual Translation Studies (Kalantzi, 2008); it also consists of the description developed by Lemle (1989) of Brazilian Portuguese syntagmas to inform the tracking of line-break categories. The methodological procedures include quantitative data extracted with the software *Wordsmith Tools*® 5.0, manual annotation and tracking of segmentation categories. The results reveal the subtitler's preference for the transcription of a great amount of speech onto the screen, called *verbatim* subtitling. Line-break tracking shows that linguistic deviations were more frequent in the verb phrase, and that in this corpus the preference for a shorter upper line is the most influential factor that competes with linguistic considerations in subtitle segmentation. Few occurrences of different line breaking choices were found between subtitling modalities, none of which could be assigned to peculiarities of the type of subtitling involved in those occurrences. The results point out that, in general, there are not significant differences between both subtitling modalities either in technical or textual terms, and that both received the same treatment from the subtitler.

Key-Words: Corpus-based Translation Studies; Audiovisual Translation; Subtitling for the Deaf and the Hard of Hearing; Subtitling aimed at Hearing Viewers; Line-Breaking within Subtitles.

LISTA DE ILUSTRAÇÕES

LISTA DE FIGURAS

FIGURA 1: Exemplos de Segmentação.....	27
FIGURA 2: Diagrama de ordens sintagmáticas.....	30
FIGURA 3 - Capa do DVD Irmãos de Fé.....	50
FIGURA 4 – Exemplos de identificação das linhas de legenda.....	54
FIGURA 5 - Pesquisa por <LS> em IF retornada pela ferramenta <i>Concord</i> ...55	
FIGURA 6 - Pesquisa por <SOSN> em IF retornada pela ferramenta <i>Concord</i>	56
FIGURA 7 - Pesquisa por <SSN em IF retornada pela ferramenta <i>Concord</i>	58
FIGURA 8- Pesquisa por <SSV em IF retornada pela ferramenta <i>Concord</i>	59
FIGURA 9 - Pesquisa por <SSAdj retornada pela ferramenta <i>Concord</i>	60
FIGURA 10 - Pesquisa por <SSPrep retornada pela ferramenta <i>Concord</i>	61
FIGURA 11 - Lista de palavras mais frequentes no corpus.....	63
FIGURA 12 - Dados estatísticos do corpus IF.....	64
FIGURA 13 – Mesmo tempo de duração de legenda em LSE_IF e em LO_IF.	70
FIGURA 14 – Lista de palavras comparada LSE_IF e LO_IF.....	73
FIGURA 15 – Pesquisa por <LS> em LSE_IF na ferramenta <i>Concord</i>	78
FIGURA 16 – Diagrama em árvore da ocorrência C15.....	100
FIGURA 17 – Árvore sintática da sentença em C19.....	102
FIGURA 18 – Diagrama em árvore da sentença em C19.....	107
FIGURA 19 – Diagrama em árvore da sentença em C22.....	110
FIGURA 20 – Diagrama em árvore da sentença em C24.....	113
FIGURA 21 – Diagrama em árvore da sentença em C29.....	117
FIGURA 22 – Diagrama em árvore da sentença em C38.....	122
FIGURA 23 – Diagrama em árvore da sentença de C40.....	124
FIGURA 24 – Diagrama em árvore da sentença em C41.....	130
FIGURA 25 – Diagrama em árvore da sentença em C52.....	133

LISTA DE GRÁFICOS

GRÁFICO 1 – Valores absolutos de <i>types</i> e <i>tokens</i> distribuídos em IF.....	66
GRÁFICO 2 – Valores absolutos de linhas de legenda em IF.....	66

GRÁFICO 3 - Distribuição dos textos no corpus em número de <i>tokens</i>	67
GRÁFICO 4 - Distribuição de número de linhas de legenda em IF.....	68
GRÁFICO 5 - Razão <i>type/token</i> na LO_IF, na LSE_IF e em IF.....	72
GRÁFICO 6 – Número de sentenças nos subcorpora.....	73
GRÁFICO 7 – Tamanho médio das sentenças nos subcorpora.....	74
GRÁFICO 8 - Legendas de 1 e 2 linhas em números absolutos.....	75
GRÁFICO 9 - Percentual de legendas de 1 e 2 linhas em LO_IF.....	76
GRÁFICO 10 - Percentual de legendas de 1 e 2 linhas em LSE_IF.....	76
GRÁFICO 11 - Percentual de legendas de 1 e 2 linhas no corpus.....	77
GRÁFICO 12 – Percentual de <LS> nas legendas de 2 linhas em IF.....	80
GRÁFICO 13 – Percentuais da distribuição da Segmentação <LS> em IF.....	81
GRÁFICO 14 – Subordinadas nas legendas de 2 linhas em LO_IF.....	85
GRÁFICO 15 – Subordinadas em legendas de duas linhas LSE_IF.....	86
GRÁFICO16 - Coordenadas nas legendas de duas linhas LO_IF.....	86
GRÁFICO 17 - Coordenadas nas legendas de duas linhas LSE_IF.....	87
GRÁFICO18 - Distribuição das <SSND*> em LO_IF em valores percentuais..	92
GRÁFICO19 - Distribuição das <SSND*> em LSE_IF.....	92
GRÁFICO 20 – Distribuição das <SSVD*> em LO_IF em valores percentuais.....	105
GRÁFICO 21 – Distribuição das <SSVD*> no corpus em valores percentuais.....	105
GRÁFICO 22 – Distribuição de <SSAdjD*> em LO_IF em valores percentuais.....	120
GRÁFICO 23 – Distribuição de <SSAdjD*> em LSE_IF em valores percentuais.....	120
GRÁFICO 24 – Distribuição de <SSPrepD*> em LO_IF em valores percentuais.....	127
GRÁFICO 25 – Distribuição de <SSPrepD*> em LSE_IF em valores percentuais.....	128
GRÁFICO 26 – Desvios de segmentação distribuídos em IF.....	135
GRÁFICO 27 – Desvios distribuídos em IF por subcategoria.....	137
GRÁFICO 28 – Distribuição dos fatores de segmentação em IF.....	139

LISTA DE QUADROS

QUADRO 1 - Possibilidades de Segmentação por Karamitroglou (1998).....	30
QUADRO 2 - Softwares utilizados na pesquisa.....	52
QUADRO 3 - Etiquetas para o Sintagma Nominal.....	57
QUADRO 4 - Etiquetas para o Sintagma Verbal.....	58
QUADRO 5 - Etiquetas para o Sintagma Adjetivo.....	59
QUADRO 6 - Etiquetas para o Sintagma Preposicional.....	60
QUADRO 7 - Ocorrências de <SSND1> em IF.....	94
QUADRO 8 - Ocorrências <SSND1> reformuladas.....	95
QUADRO 9 - Ocorrências de <SSND2>.....	96
QUADRO 10 - Ocorrências <SSND2> reformuladas.....	97
QUADRO 11 - Sentença de C14 em LO_IF.....	98
QUADRO 12 - Exemplos de ocorrências de <SSND3> em IF.....	100
QUADRO 13 - Reformulação de C15.....	101
QUADRO 14 - Reformulação de C16.....	101
QUADRO 15 - Exemplos de ocorrências de <SSVD1> em IF.....	107
QUADRO 16 - Ressegmentação de C19.....	108
QUADRO 17 - Segunda Ressegmentação de C19.....	108
QUADRO 18 - Reformulação de C20.....	109
QUADRO 19 - Segunda Reformulação de C20.....	109
QUADRO 20 - Reformulação tentativa de C21.....	110
QUADRO 21 - Sentença de C22 em LO_IF.....	111
QUADRO 22 - Reformulação de C22.....	111
QUADRO 23 - Reformulação de C24.....	112
QUADRO 24 - Exemplos de ocorrências de <SSVD2> em IF.....	113
QUADRO 25 - Reformulação de C24.....	114
QUADRO 26 - Reformulação de C25.....	114
QUADRO 27 - Reformulação de C26.....	114
QUADRO 28 - Legenda <L275> em LSE_IF.....	115
QUADRO 29 - Legenda <L216> em LO_IF.....	115
QUADRO 30 - Exemplos de ocorrências de <SSVD3> em IF.....	116
QUADRO 31 - Reformulação de C29.....	117
QUADRO 32 - Ocorrências de <SSAdjD1> em IF.....	121

QUADRO 33 - Ressegmentação de C38.....	122
QUADRO 34 - Ocorrência de <SSAdjD2> em IF.....	123
QUADRO 35 - Reformulação de C39.....	124
QUADRO 36 - Ocorrências de <SSPrepD1> em IF.....	129
QUADRO 37 - Reformulações das ocorrências de <SSPrepD1> em IF.....	131
QUADRO 38 - Ocorrências de <SSPrepD2> em IF.....	132
QUADRO 39 - Reformulações das ocorrências de <SSPrepD3> em IF.....	133

LISTA DE TABELAS

TABELA 1 - Dados quantitativos obtidos com o <i>Wordsmith Tools</i>	65
TABELA 2 - Distribuição de tipos de linha de legenda em números absolutos.....	75
TABELA 3 - Distribuição da Segmentação <LS> em IF em números absolutos.....	80
TABELA 4 - Distribuição das Segmentações de Orações em IF em números absolutos.....	84
TABELA 5 - Distribuição da Segmentação <SSN> e <SSND*> em IF em números absolutos.....	91
TABELA 6 - Distribuição da Segmentação de Sintagmas Verbais em IF em números absolutos.....	104
TABELA 7 - Distribuição da Segmentação de Sintagmas Adjetivos em IF em números absolutos.....	119
TABELA 8 - Distribuição da Segmentação de Sintagmas Preposicionais em IF em números absolutos.....	126
TABELA 9 - Desvios de segmentação por categoria.....	134
TABELA 10 - Desvios de segmentação sintagmática por subcategoria.....	136

LISTA DE ABREVIATURAS

DVD – DIGITAL VÍDEO DISC

ETBC – ESTUDOS DA TRADUÇÃO BASEADOS EM CORPUS

ESTRA – CORPUS DE ESTILO EM TRADUÇÃO

FALE – FACULDADE DE LETRAS

FEBEM – FUNDAÇÃO ESTADUAL DO BEM-ESTAR DO MENOR

IF - IRMÃOS DE FÉ

LATAV – LABORATÓRIO DE TRADUÇÃO AUDIOVISUAL

LC - LINGUÍSTICA DE CORPUS

LEAD – LEGENDAGEM E AUDIODESCRIÇÃO

LETRA – LABORATÓRIO EXPERIMENTAL DE TRADUÇÃO

LO – LEGENDAGEM PARA OUVINTES

LSE – LEGENDAGEM PARA SURDOS E ENSURDECIDOS

PB – PORTUGUÊS BRASILEIRO

SAdj - SINTAGMA ADJETIVO

SE - SURDOS E ENSURDECIDOS

SN - SINTAGMA NOMINAL

SPrep - SINTAGMA PREPOSICIONAL

SV - SINTAGMAS VERBAIS

TAV – TRADUÇÃO AUDIOVISUAL

UECE – UNIVERSIDADE ESTADUAL DO CEARÁ

UFMG – UNIVERSIDADE FEDERAL DE MINAS GERAIS

SUMÁRIO

INTRODUÇÃO	14
CAPÍTULO 1	18
REVISÃO DA LITERATURA	18
1.1) ESTUDOS DA TRADUÇÃO BASEADOS EM CORPUS.....	18
1.2) ACESSIBILIDADE E LEGENDAGEM COMO MODALIDADE DE TAV	21
1.3) TAV - LSE NO BRASIL	41
1.4) GRAMÁTICA DOS SINTAGMAS DO PB	43
CAPÍTULO 2.....	49
METODOLOGIA	49
2.1) CORPUS	49
2.1.1) FILME IRMÃOS DE FÉ	50
2.2) PROCEDIMENTOS DE EXTRAÇÃO E COMPILAÇÃO.....	51
2.2.1) EXTRAÇÃO DO FILME	51
2.2.1.1) SOFTWARES UTILIZADOS.....	51
2.2.1.2) COMPILAÇÃO DO CORPUS	53
2.3) PROCEDIMENTOS DE ANÁLISE	53
2.3.1) ANÁLISE QUANTITATIVA.....	53
2.3.2) ANOTAÇÃO DO CORPUS	54
2.3.3) ANÁLISE QUALITATIVA	61
CAPÍTULO 3.....	63
RESULTADOS E DISCUSSÃO	63
3.1) RESULTADOS GERAIS DE IF FORNECIDOS PELO WORDSMITH TOOLS E PELO SUBTITLE WORKSHOP.....	63

3.2) ANÁLISE DOS DADOS DA SEGMENTAÇÃO	77
3.2.1) SEGMENTAÇÃO <LS>	79
3.2.2) SEGMENTAÇÃO <SOSN>/<SOSD> E <SOCN>/<SOCD>	82
3.2.3) SEGMENTAÇÃO <SSN> E <SSND*>	89
3.2.3.1) Subcategoria <SSND1>.....	93
3.2.3.2) SUBCATEGORIA <SSND2>	95
3.2.3.3) SUBCATEGORIA <SSND3>	99
3.2.4) SEGMENTAÇÃO <SSV> E <SSVD*>.....	103
3.2.4.1) SUBCATEGORIA <SSVD1>	106
3.2.4.2) SUBCATEGORIA <SSVD2>	112
3.2.4.3) SUBCATEGORIA <SSVD3>	116
3.2.5) SEGMENTAÇÃO <SSADJ> E <SSADJD*>.....	118
3.2.5.1) SUBCATEGORIA <SSVADJD1>	121
3.2.5.2) SUBCATEGORIA <SSVADJD2>	123
3.2.6) SEGMENTAÇÃO <SSPREP> E <SSPREPD*>	125
3.2.6.1) SUBCATEGORIA <SSPREPD1>	128
3.2.6.2) SUBCATEGORIA <SSPREPD2>	131
3.3) DESVIOS DE SEGMENTAÇÃO EM IF	134
3.4) DISCUSSÃO DOS RESULTADOS	138
CONSIDERAÇÕES FINAIS	143
REFERÊNCIAS BIBLIOGRÁFICAS	147

INTRODUÇÃO

Em comparação com o desenvolvimento de outras vertentes na trajetória dos estudos da tradução, o interesse na linha de pesquisa em Legendagem para Surdos e Ensurdidos (LSE) da subárea tradução audiovisual (TAV) é relativamente recente, porém promissor. Pesquisadores em TAV referem-se à legendagem como uma das áreas mais prósperas dentro dos estudos da tradução. Além disso, a legendagem para surdos é mencionada como uma novidade que está sendo adotada por pesquisadores e profissionais da área de TAV, à medida que as sociedades percebem a necessidade de tornarem-se mais justas e inclusivas em relação às pessoas com deficiências sensoriais (DÍAZ-CINTAS e REMAEL, 2007).

A legendagem pode ser considerada mais um modo¹ semiótico a ser acrescentado num texto já repleto deles, portanto, entende-se que essa inserção deve ser feita da maneira mais sutil possível, de forma a ser verdadeiramente facilitadora e não obstrutiva (DÍAZ-CINTAS e REMAEL, 2007). A busca por essa sutileza tem movido os mais diversos estudos já conduzidos sobre legendagem, porquanto até então a produção de legendagem tem se dado a partir de guias prescritivos elaborados por empresas, redes de televisão e profissionais da área, com pouco ou nenhum fundamento teórico subjacente à tarefa tradutória envolvida na legendagem. Em vista disso, pesquisadores trabalhando em vários países, com diversos pares linguísticos, têm conduzido estudos descritivos, bem como experimentos de recepção com o público alvo, a fim de investigar se necessidades e preferências do público têm sido satisfeitas pela LSE disponibilizada aos interessados.

Como um ramo dos estudos da legendagem em geral, os estudos em LSE ainda em estágio primevo de desenvolvimento, fundamentam-se fortemente em pesquisas conduzidas na subárea da legendagem para ouvintes (LO). Artigos sobre explicitação e segmentação de legendas oferecem

¹ Díaz-Cintas; Remael (2007) usam o termo “canal semiótico”, porém neste trabalho será usado o termo “modo semiótico” em conformidade com o uso nas teorias de multimodalidade e semiótica social (O'TOOLE, 1994; KRESS; VAN LEEUWEN, 1996)

descrições e discussões cruciais sobre a linguagem e a linguística da legendagem (PEREGO, 2003; 2008). Mais especificamente na subárea da LSE há pesquisas que apresentam análises comparativas entre LOs e LSEs de filmes (WURM, 2007) e tratam de questões linguísticas da LSE à luz de teorias dos estudos da tradução, tendo como um dos objetivos auxiliar na formação de legendistas, os quais precisam tomar decisões bem informadas ao lidar com os problemas surgidos na transferência do meio falado para o escrito. Pesquisas de abordagem holística que analisam o conteúdo de legendagens calcadas em teorias linguísticas aliado à observação empírica do comportamento do público surdo e ensurdecido (SE) visam investigar questões que podem emergir a partir da integração dos elementos semióticos de um texto audiovisual (DE LINDE e KAY, 1999). Com base em pesquisas de caráter holístico, Kalantzi (2008) advoga pela necessidade de (i) responder importantes questões da área; e (ii) identificar as necessidades do público alvo. Esta pesquisa, portanto, dedica-se a tentar responder questões relevantes da área por meio de um trabalho de análise da segmentação das LO e LSE de um texto fílmico, desta forma, contribuindo para a área de TAV.

No Brasil, as pesquisas em LSE contam com experimentos de recepção conduzidos com o público alvo a fim de apontar a inadequação da legendagem fechada disponível no país e de sugerir a necessidade de mais pesquisas visando o desenvolvimento de um modelo brasileiro de LSE (ARAÚJO, 2004; FRANCO e ARAÚJO, 2003;). Trabalhos conduzidos no Laboratório Experimental de Tradução da Faculdade de Letras da Universidade Federal de Minas Gerais (LETRA-FALE/UFMG), também marcam o desenvolvimento das pesquisas na área de TAV, dentre eles, DINIZ (2009) propõe o uso da transcrição multimodal (BALDRY e THIBAUT, 2006) como ferramenta para análise de um trecho de texto fílmico para auxiliar na elaboração de LSE e na formação de futuros legendistas.

Para continuar uma tradição de pesquisas em LSE e com vistas a cooperar com a melhoria da acessibilidade de surdos e ensurdecidos à informação e entretenimento veiculados pelos meios audiovisuais brasileiros, este estudo propõe a análise de legendagens – LO e LSE - em língua portuguesa, com foco na segmentação entre linhas de legenda, fundamentada em subsídios de descrição do português brasileiro (PB) da gramática gerativa.

Ainda é insuficiente a quantidade de pesquisa tanto descritiva quanto empírica, especialmente sobre LSE em PB como recurso de acessibilidade sensorial, interessada em medir e descrever a legendagem produzida que permitam que generalizações sejam feitas sobre a modalidade. A presente pesquisa foi desenvolvida como parte das pesquisas em TAV do LETRA-FALE/UFMG em parceria com o Grupo LEAD (Legendagem e Audiodescrição) da Universidade Estadual do Ceará (UECE). O estudo apresentado está inserido no contexto da TAV descrito acima, com foco específico na LSE como meio de acessibilidade descrevendo a LSE do filme *Irmãos de Fé* (IF) de Moacyr Goés, sendo essa comparada à LO do mesmo filme.

Os objetivos gerais desta pesquisa são:

- i. Descrever a segmentação em LSEs e LOs através de subsídios da gramática gerativa, quanto à descrição do português.
- ii. Contribuir com a parceria entre LETRA-FALE/UFMG e Laboratório de Tradução Audiovisual (LATAV) do Grupo LEAD/UECE em TAV, investigando o papel das LSEs na acessibilidade de SEs.

Esta pesquisa tem ainda como objetivos específicos:

- i. Analisar a LSE do filme *Irmãos de Fé* focalizando o aspecto da segmentação de linhas de legenda no nível do sintagma português tendo em vista a descrição do tipo de segmentação e dos desvios gramaticais nela contidos;
- ii. Analisar a LO do filme *Irmãos de Fé* focalizando o aspecto da segmentação de linhas de legenda no nível do sintagma português tendo em vista a descrição do tipo de segmentação e dos desvios gramaticais nela contidos;
- iii. Comparar e discutir a segmentação das LO e LSE do filme *Irmãos de Fé*;

As perguntas de pesquisa que nortearam este trabalho foram:

- a) Que razão de segmentação (visual, retórica, estética, legenda verbatim) teve precedência nas LOs e LSEs de duas linhas com desvio gramatical do filme *Irmãos de Fé*?
- b) A segmentação da LSE apresenta mais desvios que a da LO devido às informações adicionais que devem ser transmitidas?

As legendas sob escrutínio que compõem o corpus neste estudo foram extraídas dos vídeos, corrigidas e analisadas com auxílio de ferramentas eletrônicas de programas como o *Subtitle Workshop*© e o *Wordsmith Tools*©. A etiquetagem do corpus foi manual e rastreou e categorizou as quebras de linhas de legenda dentro do padrão sintático do PB e as quebras de linha com desvios do padrão sintático do PB de acordo com a descrição dos sintagmas do português de Lemle (1989). Foram analisados também os critérios (estético, visual, retórico) que podem ter influenciado os desvios na segmentação das linhas de legendas marcadas com desvio linguístico.

Além da introdução, esta dissertação possui quatro capítulos. O Capítulo 1 faz uma revisão da literatura que fundamenta esta pesquisa. Nesse capítulo, são consideradas as interfaces entre estudos da tradução e linguística de corpus, a legendagem como modalidade de TAV, a TAV em LSE no Brasil e a gramática dos sintagmas do PB. O Capítulo 2 apresenta e descreve o filme de onde foram extraídas as legendas que compõem o corpus da pesquisa, são descritos também os passos metodológicos adotados na execução do trabalho e as etiquetas atribuídas para o rastreamento das segmentações. O Capítulo 3 apresenta os resultados obtidos com a análise do corpus e sua discussão à luz da literatura apresentada no Capítulo 1. Finalmente, nas Considerações Finais apresentam-se as contribuições e limitações desta pesquisa, e fazem-se recomendações para futuros trabalhos interessados em consolidar os estudos de TAV orientados para a LSE e a acessibilidade.

CAPÍTULO 1

REVISÃO DA LITERATURA

Esta pesquisa tem como suporte os subsídios dos estudos da tradução, com foco nas pesquisas em LSE no âmbito da subárea da tradução audiovisual. Portanto, o referencial teórico abrange tópicos da área de estudos da tradução, tais como: estudos da tradução baseados em corpus (ETBC) e tradução audiovisual (TAV), que engloba os temas de acessibilidade, legendagem para ouvintes, e LSE. Está incluída também uma descrição dos sintagmas nominal, verbal, adjetivo e preposicional do português brasileiro [PB] segundo a perspectiva da gramática gerativa.

A próxima subseção traz uma breve revisão de trabalhos relevantes sobre os estudos da tradução baseados em corpus e também dos trabalhos na interface ETBC e TAV, até então conhecidos.

1.1) ESTUDOS DA TRADUÇÃO BASEADOS EM CORPUS

A Linguística de Corpus (LC) emergiu a partir da necessidade de ampliar o alcance das pesquisas linguísticas em geral, visto que anteriormente os estudos linguísticos eram baseados em unidades de texto muito reduzidas. Nesses estudos, a metodologia envolvia uma criteriosa técnica de coletar textos impressos a fim de serem analisados manualmente. Todavia, devido ao grande esforço e longo tempo despendido na análise, os resultados das pesquisas ainda sim eram muito limitados, não favorecendo a identificação de padrões nos textos estudados, o que obstruía o avanço das pesquisas rumo a generalizações linguísticas.

Acompanhando as evoluções tecnológicas, as ferramentas de análise de texto também vêm se modernizando e atualmente *corpus*, do ponto de vista postulado em Baker (1995), é uma coletânea eletrônica de textos – falados ou escritos – provenientes de várias fontes, escritos por diferentes autores, que tratam de tópicos variados, agrupados para um determinado propósito e segundo critérios pré-estabelecidos, passíveis de serem analisados de maneira automática ou semi-automática.

Baker (1993, 1995, 1996) fomenta o início da utilização da LC nos Estudos da Tradução defendendo que

Grandes corpora fornecerão aos estudiosos da tradução uma oportunidade única de observar seu objeto de estudo e explorar o que o faz diferente dos outros objetos de estudo, como a linguagem em geral, ou mesmo qualquer outro tipo de interação cultural. Além disso, irá permitir-nos explorar, em escala maior do que jamais foi possível antes, os princípios que governam o comportamento tradutório e as restrições sob as quais opera. ²(BAKER, 1993, p.235)

Desde então, a LC tem possibilitado a compilação de grandes bancos de dados, que são então aplicados em pesquisas de diferentes áreas da linguística, incluindo os estudos da tradução (BAKER, 1995; KENNY, 1998; LAVIOSA, 2002; OLOHAN, 2004). Esses corpora têm sido de grande importância, uma vez que permitem que estudos - antes conduzidos a partir de casos isolados - mantenham o caráter descritivo, mas também incorporem uma perspectiva empírica que conta com suporte quantitativo e abrangente dos dados fornecidos pelo corpus. Saldanha (2009) destaca a importância dos estudos descritivos da tradução, pois permitem a análise de textos traduzidos dissociados de seus textos fonte, ou seja, consideram a tradução como um fenômeno independente, portanto possuidora de traços distintivos e princípios peculiares, os quais regem a produção de traduções, e assim, tornam-se identificáveis à luz de uma abordagem descritiva.

Na subárea da legendagem, até o momento são poucas as pesquisas documentadas que se utilizam dessa metodologia de análise de corpus (CHAVES, 2009; FEITOSA, 2009; KALANTZI, 2008; PEREGO, 2003, 2008, 2009, 2010) configurando, portanto, uma interface ainda incipiente com a TAV.

Dos trabalhos citados acima, Chaves (2009) - o único que usa o software *Wordsmith Tools*© - tem o objetivo de averiguar o potencial da análise de corpus como metodologia na descrição das legendas para surdos produzidas pela televisão brasileira em comparação com suas respectivas

² Large corpora will provide theorists of translation with a unique opportunity to observe the object of their study and to explore what it is that makes it different from other objects of study, such as language in general or indeed any other kind of cultural interaction. It will also allow us to explore, on a larger scale than was ever possible before, the principles that govern translational behaviour and the constraints under which it operates. (Tradução minha)

reformulações (feitas pelo grupo LEAD) para serem usadas em experimentos com público alvo. Feitosa (2009) estuda legendagem para ouvintes de filmes comercializados em DVD, descrevendo legendagens comerciais e piratas através de um estudo comparado entre essas. A partir da análise manual de um corpus que consta das legendas de dois filmes, Perego (2003) investiga a existência do fenômeno da explicitação na legendagem; já Perego (2009) examina as legendas de três filmes e apresenta resultados qualitativos da pesquisa sobre a importância dos elementos não verbais e sua representação na legenda na transmissão e compreensão da mensagem escrita. Esses estudos marcam o progresso das pesquisas na interface TAV e ETBC, mas por terem um foco diferente deste trabalho, não serão aprofundados aqui. Diferentemente, os outros dois estudos de Perego (2008, 2010) são fundamentais nesta pesquisa e serão revisados na próxima subseção.

Dentre as obras citadas anteriormente, a metodologia de análise de Kalantzi (2008) é a mais relevante para esta dissertação. Baseada em Díaz-Cintas (2004), a pesquisadora ressalta a importância do estudo da legendagem por meio da análise de grandes corpora a fim de investigar as práticas atualmente adotadas na legendagem, bem como o comportamento de tradutores, para que se chegue a conclusões substanciais (KALANTZI, 2008, p.100). Entretanto, a autora ressalta que a pesquisa baseada em corpus na subárea da TAV ainda está em estágio primevo. Os corpora compilados incluem poucos produtos audiovisuais, com um número limitado de legendas e variedade insuficiente de gêneros fílmicos.

Em Kalantzi (2008) apresenta-se uma breve revisão da definição e da tipologia de corpora, tal como proposta por Laviosa (2002), que discrimina os cinco principais tipos de corpora: (i) monolíngue, bilíngue e multilíngue; (ii) paralelo e comparável; (iii) texto completo e amostragem; (iv) oral, escrito e misto; (v) sincrônico e diacrônico. Nesse ponto, um problema é identificado: o típico corpus intralingual de LSE não se encaixa em nenhuma das categorias de corpus identificadas. Um corpus comparável lida com duas coleções separadas de textos: textos originais em uma língua e textos traduzidos para essa mesma língua originados de diferentes línguas fonte. Visto que os textos – áudio e legenda - que constam no corpus de LSE estão em relação de tradução, embora no mesmo idioma, não é possível utilizar a nomenclatura

comparável para tais corpora. De igual modo, um corpus paralelo lida com textos originais em uma língua e de suas versões traduzidas em outra língua. Como num corpus de LSE a língua do original e da tradução é a mesma, a nomenclatura “paralelo” se mostra inadequada. Kalantzi (2008) propõe, portanto, que seja criada uma categoria denominada *corpora paralelos monolíngues*, de modo a acomodar convenientemente os corpora de LSE.

Outro ponto importante a respeito de corpora de LSE destacado por Kalantzi (2008) é que esses têm o potencial de serem multimídia, contendo não apenas o roteiro original de diálogos, mas também o texto fílmico. Até então, sabe-se da existência de somente um corpus desse tipo, o Forlixt 1 – *Forli Corpus of Screen Translation* – da subárea da dublagem, desenvolvido para coletar e permitir estudos de tradução fílmica, e aplicado como ferramenta pedagógica na formação de tradutores audiovisuais. A criação de outros corpora nos moldes do Forlixt 1 seria um recurso inestimável para as pesquisas em TAV (KALANTZI, 2008, p.102). Uma revisão mais detalhada do trabalho de Kalantzi (2008) será apresentada na subseção seguinte.

A seguir serão apresentados os estudos da subárea da tradução audiovisual focados em acessibilidade sensorial especificamente através da legendagem adotados neste trabalho.

1.2) ACESSIBILIDADE E LEGENDAGEM COMO MODALIDADE DE TAV

Ivarsson e Carrol (1998) é um trabalho pioneiro em legendagem, de caráter internacional, que propõe princípios gerais e metodologias de trabalho mais adequados para a prática da legendagem. O estudo trata da história e evolução da legendagem, do mercado profissional da área, dos princípios, da prática, das convenções, dos equipamentos e ferramentas, do controle de qualidade, política e economia da legendagem.

Relevante para este estudo, o aspecto de quebra de linhas da legendagem figura entre os princípios gerais tratados em Ivarsson e Carrol (1998). Algumas das recomendações dadas pelos autores estão enumeradas a seguir (IVARSSON e CARROL, 1998, p.76-78):

- Colocar o texto em uma linha sempre que possível;

- Em telas muito largas, como no cinema, é preferível formar duas linhas curtas a uma linha longa;
- Legendas de duas linhas deveriam ser reservadas para casos em que o tamanho do texto e as restrições de tempo as fazem preferíveis;
- O movimento dos olhos do final da primeira linha ao início da segunda linha leva uma fração de segundo e pode causar tensão nos músculos oculares, portanto, deve-se observar atentamente o alinhamento das linhas;
- Quando a legenda tiver mais que uma linha, a quantidade necessária de movimento dos olhos deveria ser mínima;
- O legendista deve sempre deixar o máximo de tempo para os olhos capturarem a imagem;
- As linhas devem ser divididas de forma que palavras ligadas por lógica, semântica ou gramática sejam colocadas na mesma linha sempre que possível;
- É importante harmonizar quebras de linha e blocos semânticos, principalmente para espectadores com alguma desvantagem linguística, mas até espectadores regulares desfrutem mais de legendas se as quebras de linha ocorrem em unidades semânticas razoáveis, deixando mais tempo para apreciação das imagens.

Um dos tópicos brevemente considerados em Ivarsson e Carrol (1998) e também de interesse para a presente pesquisa é a LSE. Na época essa legendagem era transmitida via *teletexto*, – informação textual transmitida como código binário no intervalo vertical em branco entre os quadros do canal e decodificadas por aparelho próprio acoplado ao televisor - e representava um reflexo das primeiras preocupações com a questão da acessibilidade sensorial. Ivarsson e Carrol (1998) apresenta e justifica algumas demandas básicas e específicas da modalidade como a representação de efeitos sonoros, identificação de falantes e menor velocidade de leitura. No entanto, os autores ressaltam que apenas um capítulo de seu trabalho não é suficiente para tratar de um assunto tão importante e que merece total dedicação.

Um estudo seminal dedicado à LSE é De Linde e Kay (1999), o qual propõe uma abordagem holística para a análise do conteúdo de legendagens

através de teorias linguísticas aliado à observação empírica do comportamento do público alvo.

Os autores diferenciam a legendagem intralingual (para surdos e ensurdecidos) da interlingual (para filmes em língua estrangeira). A legendagem interlingual, conforme afirma De Linde e Kay (1999), não é adequada para todo o público, pois a velocidade de leitura requerida do espectador é muito alta para a maioria do público SE e porque apenas os elementos linguísticos da trilha sonora são colocados na tela. Segundo De Linde e Kay (1999) a demanda de sincronismo entre os canais auditivo e visual é uma característica comum a todas as técnicas de TAV. No caso da legendagem, a coordenação de som e imagem é ainda mais complicada com a adição do componente textual, o que enfatiza a essencialidade da interação semiótica intrínseca aos textos audiovisuais.

O objetivo da legendagem intralingual é substituir o diálogo e outros traços importantes da trilha sonora por uma representação escrita posta na tela; já na interlingual busca-se uma forma de equivalência de tradução. Os autores sugerem que a transformação do diálogo em legendas e a inserção destas no produto audiovisual devem ser realizadas levando em consideração e respeitando toda a combinação de elementos visuais e auditivos que compõem um produto audiovisual. É importante sempre observar e preservar o sincronismo entre esses elementos, tendo em vista que a legendagem é o meio principal de SE terem acesso à informação veiculada por meios audiovisuais.

De Linde e Kay (1999) considera a habilidade de leitura como um dos principais fatores de compreensão de legendas. Assim, deve-se levar em conta que no caso da legendagem, a velocidade de leitura não está sob o controle do leitor; e que o espectador tem de fazer, pelo menos, duas coisas ao mesmo tempo: ler e ver as imagens. Nesse aspecto os SE ainda precisam lidar com as desvantagens de ter um nível de leitura abaixo da média, devido a uma aprendizagem/ensino da língua em uso na comunidade pouco adequada à sua condição de usuários da língua de sinais. De Linde e Kay (1999) faz uma breve revisão da literatura ressaltando estudos que comparam as habilidades de leitura entre ouvintes e SE, os quais indicam que os últimos são capazes de ler texto alfabético, porém com menor proficiência que os primeiros. São também apresentados estudos que apontam quatro possíveis estratégias de

recodificação (código articulatorio, datilologia, língua de sinais e não recodificação), sendo que os SE podem usar qualquer uma delas, contudo nenhuma delas parece ser eficiente como a recodificação fonológica usada por ouvintes.

Quanto à integração entre texto e filme, De Linde e Kay (1999) se propõe a analisar três fatores: (i) a integração entre legendas e a estrutura semiótica de um filme; (ii) a alteração que ocorre no discurso falado ao ser transformado em escrito; e (iii) as condições da legenda em relação às capacidades de leitura do público. As características textuais da legenda sugerem que a legendagem é entendida como uma forma de edição de texto; edição esta que pode afetar o estilo da linguagem e a coesão textual. A maioria das legendas é uma representação do diálogo. Sendo assim, precisam manter características de oralidade, mas ao mesmo tempo caracterizam-se por trazer, na forma escrita, a informação de modo condensado. Isso resulta em menos orações e maior densidade lexical enquanto conserva características interpessoais suficientes para manter fortes traços de discurso oral. Além de causar mudanças estilísticas, a transição entre língua escrita e falada também pode ter efeitos na semântica. A demanda de redução no tamanho do diálogo significa que omissões devem ser feitas a partir de escolhas. Nesse processo, geralmente, elementos de coesão textual (referentes, substituições e elipses, conjunções e coesão lexical) são omitidos. Tais recursos, todavia, têm um papel importante na compreensão do texto da legenda visto que explicitam relações entre entidades e eventos e a omissão deles pode levar à dificuldade de processamento e perda de significado. Os autores afirmam que é importante lembrar também que as propriedades fílmicas que compõem a narrativa visual na tela são tão importantes quanto a narrativa textual para o público SE. Manipulação de câmera e edição são recursos que governam a complexa criação de informação através de imagens mostradas de maneira a causar impressão de serem contínuas no tempo e no espaço. De Linde e Kay (1999) menciona que quando o diálogo falado é substituído pelo discurso escrito, imagens e discurso linguístico não podem mais ser processados simultaneamente, o que afeta a maneira como um filme é recebido.

De Linde e Kay (1999) justifica seu trabalho argumentando que poucos estudos tinham sido feitos a respeito de aspectos linguísticos da

legendagem quanto à reprodução de significado da língua oral na língua escrita e os efeitos da edição. O trabalho foca exclusivamente a legendagem intralingual para SE na tentativa de considerar todos os fatores relacionados a esta modalidade de tradução e ao público alvo, incluindo uma análise de propriedades linguísticas de legendas frente ao conteúdo do filme e do comportamento de leitura por parte dos surdos. De Linde e Kay (1999) apresenta as ferramentas usadas para descrever e analisar legendagens intralinguais já produzidas, bem como aquelas empregadas no estudo do comportamento de leitura e recepção do público alvo. Entretanto, os autores não esgotam as possibilidades e demonstram interesse em fomentar futuras pesquisas na área levantando questionamentos e sugerindo outras abordagens.

Quanto à legendagem como modalidade de TAV há vários estudos descritivos que tratam de aspectos técnicos e linguísticos envolvidos no legendar. Um desses estudos é a tese de Neves (2005) que trata da LSE na Europa, com maior ênfase no contexto português. Por um lado, Neves (2005) faz uma análise descritiva das práticas de legendagem adotadas na Europa que podem ser encontradas em guias de legendagem, bem como em produtos legendados. Por outro lado, Neves (2005) investiga razões e propõe mudanças, advogando por um estudo descritivo que resulte em prescrições compiladas em formato de diretrizes e formuladas também com base empírica, ou seja, a partir da identificação das necessidades e demandas dos SE portugueses, portanto, melhorando a prática da LSE em Portugal. Em relação à segmentação, a pesquisa de Neves (2005) menciona brevemente as instruções de Karamitroglou (1998), revisado mais adiante nesta subseção.

Outro trabalho que trata das características técnicas e textuais da legendagem é o de Díaz-Cintas e Remael (2007) que traz à tona a necessária discussão sobre a acessibilidade como direito humano, defendendo uma sociedade permanentemente inclusiva que deseja facilitar e garantir que surdos (e deficientes visuais) tenham direito de acesso à informação e entretenimento veiculados pelos meios audiovisuais. A obra de Díaz-Cintas e Remael (2007) configura um importante fundamento teórico, na qual os autores referem-se à legendagem como uma das áreas mais prósperas nos estudos da tradução que tem ganhado publicidade devido à multiplicação e larga difusão das produções

audiovisuais. A LSE também é considerada como novidade que está sendo adotada por pesquisadores e profissionais da área como parte integrante do campo da TAV, na medida em que as sociedades percebem a necessidade de se tornarem mais justas e inclusivas no que tange a pessoas com deficiências sensoriais. Outro tópico teórico essencial é a polissemiotividade de textos fílmicos, os quais, a fim de construir e comunicar a história ou mensagem, lançam mão não só de diálogos e falas, mas também do cenário, fotografia, iluminação, cores, figurino, gestos, expressões faciais, trilha sonora, bem como cultura e ideologia. Todos esses modos semióticos trabalham simultânea e coerentemente para criar um todo significativo. A legendagem, portanto, é mais um modo semiótico adicionado a um texto já polisemiótico, assim, para ser eficiente deveria considerar todos esses modos semióticos em ação.

Levando isso em conta, o trabalho de Díaz-Cintas e Remael (2007) conduz um estudo descritivo de aspectos fundamentais da legendagem. Essa abordagem inclui características técnicas, convenções de pontuação, questões de tradução e ainda características do texto das legendas. Nessa última categoria, trata-se do conceito de segmentação, relevante para a presente pesquisa. Díaz-Cintas e Remael (2007, p.172) define *a segmentação como a divisão do diálogo fonte, da narração etc. em seções ou segmentos – legendas – que os receptores podem entender num relance*³. De acordo com os autores, as regras de segmentação são as mesmas, seja na distribuição do texto nas duas linhas disponíveis em uma legenda – quebra de linha – ou na divisão do texto em duas ou mais legendas. Portanto, é atribuída ao legendista a responsabilidade de estruturar legendas em sentenças completas que sejam facilmente entendíveis, tenham autonomia sintática e semântica e não sejam ambíguas, de modo a garantir a coerência e coesão nas legendas. Nesse último caso, é importante considerar a limitação da memória de curto prazo e então dividir sentenças complexas em unidades menores.

No estudo apresentado aqui, serão analisados apenas os casos de quebra de linha encontrados ao longo do corpus de pesquisa. Segundo Díaz-Cintas e Remael (2007, p.172), no caso da quebra de linhas devem ser priorizados padrões sintáticos e gramaticais, ao invés de regras estéticas.

³ “[S]egmentation is the division of the ST (source text) dialogue, narration etc. into sections or segments – subtitles – that the viewers can understand at a glance.” (Tradução minha)

Linhas de legenda com formatação simétrica, ou a utilização da linha inferior apenas após o preenchimento de todos os caracteres disponíveis na linha superior, são escolhas contraindicadas. Conforme os pesquisadores, qualquer uma das linhas pode ser menor que outra, contanto que a segmentação se dê num ponto em que a informação transmitida esteja semanticamente completa, de modo que a interrupção momentânea do processamento linguístico ocorra de forma sutil e confortável. Ainda sugerem que o ideal seria que cada linha de legenda equivalesse a uma sentença completa, porém, como isso nem sempre é factível, o importante é garantir que cada legenda faça sentido por si só e ainda indique que a sentença continua na próxima legenda. A Figura 1 contém os exemplos de segmentação considerados ideais pelos autores.

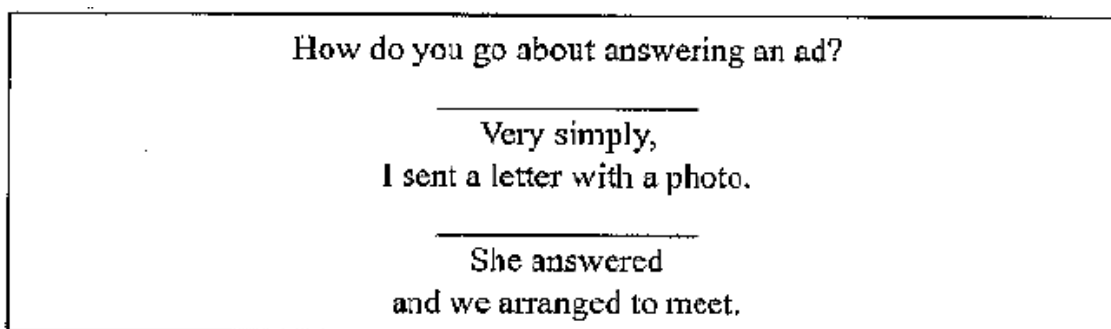


FIGURA 1: Exemplos de Segmentação
Fonte: DÍAZ-CINTAS e REMAEL, 2007

Nesse ponto, Díaz-Cintas e Remael (2007, p.174) afirma que “a segmentação sempre se beneficia de uma divisão em partes que podem funcionar independentemente, e isso é alcançado mais facilmente em orações coordenadas que subordinadas.”⁴

Citando Helen Reid (1996, p. 100), que identifica três tipos de segmentação: (i) gramatical – baseada em unidades semânticas; (ii) retórica – baseada no ritmo da fala; e (iii) visual - baseada em cortes de cena e movimentação de câmera; Díaz-Cintas e Remael (2007, p.175) apenas se atém a defender que esse último tipo de segmentação está mais relacionado a considerações espaciais, à técnica de marcação de tempo de entrada/saída de

⁴ “Segmentation always benefits from a division into parts that can function independently, and this is achieved more easily with coordinated than with subordinated clauses.” (Tradução minha)

legenda e sua relação com a narrativa visual, que a questões linguísticas. A partir disso, trata mais concretamente dos tipos gramatical e retórico.

Sobre as considerações sintático-semânticas Díaz-Cintas e Remael (2007, p.176-177) apresenta uma lista que consta das seguintes orientações para tradutores ao tomar decisões sobre quebras de linhas em legendas de textos fílmicos em inglês:

- Não hifenizar palavras.
- No caso de uma legenda consistir de duas ou mais sentenças, colocar cada sentença em uma linha.
- Se uma legenda consistir de uma sentença com duas orações subordinadas ou coordenadas, e inserir uma após a outra em uma linha for impossível, pois se excederia o número de caracteres por linha, ou resultaria numa legenda extremamente longa, deve-se usar uma linha para cada oração.
- Não separar adjetivo de substantivo ou advérbio; advérbio do verbo; artigo de substantivo; preposição dos outros constituintes do sintagma preposicional e assim por diante.
- Se a sentença contiver verbo precedido de *to* (to-infinitive), verbo frasal ou colocação (*collocation*), tentar não separá-los.
- Se a sentença contiver uma forma verbal composta, não separar o auxiliar do verbo lexical.
- Evitar separar o verbo do objeto direto ou indireto. Se uma sentença simples precisar ser dividida em duas linhas, o sujeito deveria ficar na primeira linha e o verbo mais o predicado na segunda linha.
- Se uma sentença for a resposta a uma pergunta, ou uma reação a uma afirmativa, a resposta/reação será melhor posicionada na segunda linha que na próxima legenda, a menos que isso dê a informação antes da hora.

Na segmentação retórica, conforme Díaz-Cintas e Remael (2007, p.179), há a tentativa de trazer ao texto escrito, traços prosódicos significativos da linguagem oral, tais como hesitação, pausas, jocosidade, respostas prontas dadas rapidamente, os quais conferem dinamicidade ao texto, apóiam e qualificam a mensagem do falante ao transmitir surpresa, suspense, ironia, dúvida etc., quando bem distribuídos nas linhas de legenda. Os autores apontam que as segmentações sintático-semântica e retórica usualmente

coincidem, devido à colaboração entre características linguísticas e paralinguísticas da linguagem oral, assim é provável que uma segmentação sintático-semântica sem desvios redunde em segmentação retórica de boa qualidade. A recomendação geral dada pelos autores é que nos casos em que haja necessidade de decidir quanto ao tipo de segmentação a ser usado é importante definir se o impacto retórico naquele momento é mais relevante que a semântica do enunciado, se a conclusão for de que o aspecto prosódico não contribui com a mensagem, não se deve transmiti-lo, antes é preferível garantir uma boa segmentação sintático-semântica.

Díaz-Cintas e Remael (2007) adota, como fundamento principal em suas considerações sobre segmentação, o trabalho Karamitroglou (1998), o qual por sua vez critica a abordagem descritiva, em detrimento da prescritiva, que vigorava na Europa nas pesquisas sobre legendagem. Na contramão dessa tendência, portanto, dedica-se a prescrever um conjunto de parâmetros e códigos de legendagem unificados que abranjam várias audiências europeias sem violar convenções pré-estabelecidas nos países europeus. À vista disso, Karamitroglou (1998) apresenta diretrizes para a produção e a formatação de legendas para televisão, cujo “objetivo é proporcionar máxima apreciação e compreensão do texto fílmico como um todo ao maximizar a legibilidade e a leiturabilidade (cf. p.12) da legenda inserida no texto” (Karamitroglou, 1998, sem página).⁵

As orientações enumeradas em Karamitroglou (1998) estão categorizadas em quatro grupos e lidam com parâmetros para formatações espaciais, duração de legendas, pontuação e tipo de letra e edição de texto. Dentro dessa última categoria, entre vários outros assuntos, o pesquisador trata do aspecto da segmentação, dando as seguintes instruções:

- É melhor dividir uma longa linha de legenda em duas linhas, porque assim olhos e cérebro interpretam uma legenda de duas linhas como mais volumosa e aceleram o processo de leitura.
- O texto legendado deveria aparecer segmentado no sintagma de ordem mais alta possível. Isto é, cada legenda idealmente deveria conter uma

⁵ “[...] the aim to provide maximum appreciation and comprehension of the target film as a whole by maximising the legibility and readability of the inserted subtitled text.” (Tradução minha)

sentença completa. Nos casos em que a legenda não caiba em uma única linha de legenda e precise continuar na segunda linha ou até na próxima legenda, a segmentação em cada linha deve ser arranjada de forma a coincidir com o sintagma de ordem mais alta.

Para ilustrar esse princípio, Karamitroglou (1998) se utiliza da sentença “The destruction of the city was inevitable.”⁶ apresentando um diagrama (Figura2) onde apresenta os sintagmas dessa sentença, e ainda propõe duas segmentações diferentes (Quadro 1).

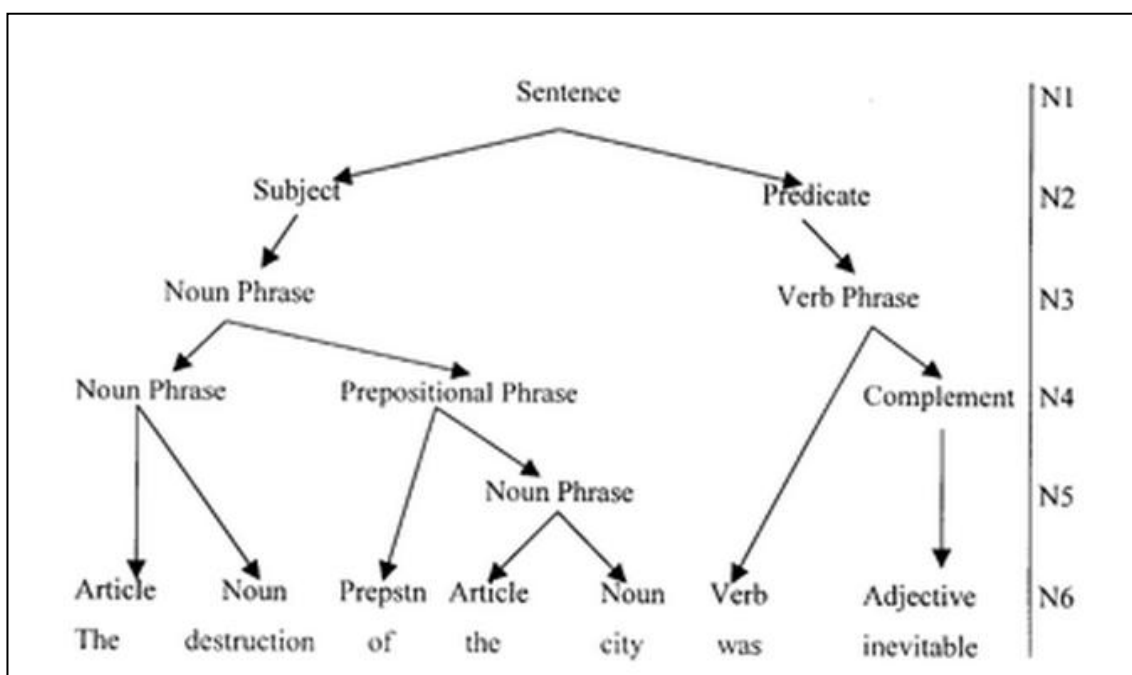


FIGURA 2: Diagrama de ordens sintagmáticas
 Fonte: KARAMITROGLOU, 1998.

QUADRO 1
 Possibilidades de Segmentação por Karamitroglou (1998)

“The destruction of the city was inevitable”	“The destruction of the city was inevitable”
Sintagma de ordem 5 (Cf. Figura2)	Sintagma de Ordem 2 (Cf. Figura2)

Dentre as possibilidades de segmentação, a proposta da direita é aquela que proporciona mais conforto de leitura, uma vez que quanto mais alta a ordem do

⁶ “A destruição da cidade foi inevitável.” (Tradução minha)

sintagma, maior é a carga semântica e mais completa a informação apresentada ao cérebro (KARAMITROGLOU, 1998). Portanto, Karamitroglou (1998) recomenda:

- As linhas superior e inferior da legenda de duas linhas devem ser proporcionalmente iguais no comprimento, visto que o espectador está mais acostumado à leitura em formato retangular – texto impresso – que triangular. Sendo assim, a segmentação da legenda deve trazer as regras de sintaxe e geometria em acordo, contudo, a sintaxe deverá predominar nas ocasiões em que uma dessas deva ser sacrificada.
- Cada enunciado falado deve corresponder a uma sentença na legenda.
- Não mais que duas sentenças são permitidas na mesma legenda. Cada sentença deve ocupar uma linha.

Apesar de criticar a sobrecarga de estudos descritivos, Karamitroglou (1998) admite utilizar resultados dessas pesquisas como fundamento para seu ponto de vista prescritivo, o qual busca fornecer uma fórmula de unificação que supere as diferentes convenções de legendagem em vigor em vários países europeus para atender as necessidades do mercado audiovisual da região como um todo.

As orientações de Karamitroglou (1998) servem de apoio também para o estudo de Perego (2008) que pretende mostrar os possíveis efeitos da segmentação do texto da legenda na leiturabilidade⁷ (PEREGO, 2008, 35). Perego (2008) se apropria desse termo já estabelecido na engenharia que diz respeito à

qualidade que possibilita o reconhecimento do conteúdo de informação de material quando é representado por caracteres alfanuméricos em agrupamentos significativos, tais como palavras, frases ou texto contínuo" (SANDERS e MCCORMICK, 1993), enquanto a legibilidade é concernente "ao atributo de caracteres alfanuméricos que os torna distinguíveis um do outro" (*ibid.*). Os dois termos são frequentemente confundidos e usados como sinônimos, embora o primeiro se refira à facilidade de leitura, conforme determinado pela organização de unidades de informação e pelo design tipográfico como um todo, enquanto o segundo se

⁷ Readability (Tradução minha).

refira à facilidade com que uma pessoa consegue identificar caracteres ou letras. (PEREGO, 2008, p. 35)⁸

A autora sustenta que a segmentação cuidadosa não é uma característica da legendagem que tem sido observada, apesar de os estudos psicolinguísticos apontarem que a leitura de grupos de palavras coerentes a cada fixação do olhar auxilia e facilita o progresso linear da leitura. Perego (2008) propõe uma análise qualitativa da distribuição do texto nas linhas de legendas e diferentes tipos de filmes, a fim de detectar os tipos de segmentação inconvenientes que podem ocorrer em níveis micro e macro da organização da linguagem.

Perego (2008) justifica-se afirmando que a literatura sobre segmentação é limitada e pouca importância – quando comparado aos outros critérios da legendagem - tem sido dada ao estudo do assunto. A discussão sobre a formatação das linhas de legenda oscila entre opiniões que defendem legendas de uma linha e as que defendem a de duas linhas; mas o parecer de Perego (2008) é que o importante é buscar o equilíbrio na edição a fim de beneficiar o espectador e que quando a legenda de duas linhas for usada, a linha superior deverá ser menor que a inferior e as duas devem ser centralizadas na tela. A linha superior menor é menos intrusiva à imagem e menos cansativa aos olhos.

Além das complexidades do processo cognitivo envolvido na leitura de texto escrito, a leitura de legendas ainda tem o complicador da pressão de tempo e de como o texto é distribuído entre as linhas, em termo de quebras sintáticas. Por isso, Perego (2008) utiliza-se das teorias de sintaxe e dos modelos de comportamento de leitura para avaliar como um influencia o outro, a partir da hipótese de que a segmentação imprevisível, incoerente, imprecisa ou implausível seria particularmente perturbadora no caso de leitura de legendas. Perego (2008) retoma o princípio de Karamitroglou (1998) de que a

⁸ [...] quality that makes possible the recognition of the information content of material when it is represented by alphanumeric characters in meaningful groupings, such as words, sentences, or continuous text” (Sanders and McCormick 1993) while legibility pertains to “the attribute of alphanumeric characters that makes it possible for each one to be identifiable from others” (*ibid.*). The two terms are often confused and used interchangeably although the former refers to the ease of reading as determined by the organization of information units and by typographic design as a whole, whereas the latter refers to the ease with which a person manages to identify characters or letters. (Tradução minha)

quebra de linha é desviante quando grupos coerentes de palavras são separados e a segmentação não acontece no sintagma de ordem mais alta.

Perego (2008) conduz uma análise descritiva e qualitativa, que começa no nível do sintagma seguido do nível da sentença, para tanto dispõe de um acervo composto por 15 filmes de diferentes gêneros e nacionalidades que tenham sido legendados para ouvintes e para SE em Inglês, Italiano e Húngaro.

No nível sintagmático a pesquisadora presume que essas são mais bem processadas quando apresentadas em blocos de acordo com sua disposição ordenada que permita que o espectador as receba com apenas uma única fixação ocular. Quanto aos sintagmas nominais (SN), Perego (2008) defende que, quando forem simples ou não forem formados de pré- ou pós-modificadores complexos, esses devem ser colocados na mesma linha. SN mais complexo com pré-modificadores é um empacotamento denso de informação que cria um conceito de unidade ao lado do núcleo do sintagma e por isso deviam ser deixados juntos - e.g. artigos + adjetivo + substantivos; título + nome próprio. Já no caso de modificadores que sucedem o núcleo, os SN são pacotes que contêm informação extra e a relação entre núcleo e modificador é mais distante, por isso podem ser facilmente processados mesmo que separados em linhas diferentes. Ainda a respeito de SN, a autora defende a observância da fixidez sintática de expressões cristalizadas, colocações e expressões idiomáticas que exigem segmentação cuidadosa para não violar os padrões costumeiros e os significados figurativos que emergem dessas. Sobre sintagmas preposicionais (SP), a autora aponta que preposições simples ou complexas podem facilmente ser movimentadas de uma linha para outra a fim de permanecerem junto de seu núcleo; inclusive nos casos em que o rephraseamento se faz necessário é melhor que se exceda o limite de caracteres por linha que separar a preposição de seu núcleo, especialmente no caso de itens ambíguos. A respeito da segmentação que envolve sintagmas verbais (SV), a separação do verbo principal de seu auxiliar pode causar diferentes níveis de ruptura. A separação do verbo principal de um auxiliar simples pode provocar uma fratura na unidade conceitual que esses elementos formam. Por outro lado, visto que os componentes de tempo verbal e modalidade são considerados unidades estruturalmente e semanticamente

independentes em um SV, a separação dessas não é tão perturbadora como a do caso anterior. Novamente, Perego (2008) aborda também o assunto sobre SV fixos e *phrasal verbs*, cuja separação pode acarretar uma reanálise quando do reconhecimento tardio – no caso da leitura de legenda - do item de desambiguação.

No nível sentencial Perego (2008) estuda casos de quebra de linha inadequadas identificadas em instâncias de coordenação, subordinação, estruturas condicionais e comparativas. Em geral, os elementos introdutores das orações coordenadas e subordinadas apareceram separados da oração que introduzem, apesar de ser naturalmente mais ligados a essa que à oração principal. No caso das estruturas condicionais e comparativas foi identificada a tendência em separar a partícula *se* e os *comparadores* da oração que introduzem, causando o rompimento da integridade da organização total da sentença.

Em sua conclusão, Perego (2008) aponta que experimentos com rastreamento ocular são fundamentais na pesquisa em legendagem a fim de investigar e entender os padrões de fixação do olhar na leitura de legendas de segmentação desviante em comparação com a leitura de legendas de segmentação sem desvios, uma vez que onde e por quanto tempo o leitor fixa o olhar parece estar relacionado à facilidade ou dificuldade de compreensão de uma determinada parte do texto. Entretanto, a autora reitera que a partir do conhecimento já alcançado sobre o assunto é possível evitar segmentações aleatórias através de simples intervenções que fariam os textos audiovisuais produtos melhores, mais úteis e acessíveis aos diferentes públicos alvo.

Com base nessa pesquisa anterior, Perego et al (2010), então, conduz um estudo para testar o processamento cognitivo da legendagem com base na relação entre distribuição de foco de atenção, reconhecimento de palavra e reconhecimento de cena. Para tanto, Perego et al (2010) levantou três hipóteses que nortearam o estudo, e cujos objetivos eram: (i) testar empiricamente a hipótese da eficácia cognitiva do processamento de legendagem analisando a relação entre reconhecimento de palavra e de cena; (ii) testar experimentalmente a influência da qualidade da segmentação no processamento cognitivo e na capacidade de reconhecimento; (iii) fornecer novas evidências empíricas sobre o processamento cognitivo de programas

legendados a partir de medidas de performance e análise de movimento ocular (PEREGO et al, 2010, p.7).

Dos três objetivos anteriormente enumerados, o segundo é o que mais interessa ao presente estudo. Para conduzir o experimento mencionado, Perego et al (2010) contrastou duas hipóteses sobre os efeitos da qualidade da segmentação da legendagem. Por um lado, a primeira hipótese previa que legendas com segmentação de má qualidade poderiam obstruir significativamente o processamento de informação, ralentar a leitura e dificultar o reconhecimento de texto dos participantes. Por outro lado, a segunda hipótese previa que os participantes não teriam maiores problemas devido à qualidade da segmentação, visto que a leitura de legendas é uma atividade já automatizada e que requer pouco esforço.

O experimento relatado em Perego et al (2010) contou com 41 informantes, estudantes de graduação e pós-graduação de uma universidade italiana, todos falantes nativos de italiano, não conhecedores de húngaro (a língua original do filme legendado) que declararam não estar acostumados a assistir programas legendados, visto que na Itália a técnica de dublagem é mais usada. O vídeo apresentado aos sujeitos foi um trecho extraído de um drama húngaro, com 15 minutos de duração e legendas em italiano. Esse trecho continha 171 linhas de legenda, sendo 90 delas segmentadas em duas linhas. Dessas 90 linhas, 28 linhas eram as legendas teste, cuja metade tinha segmentação de boa qualidade e a outra metade tinha segmentação de má qualidade. As legendas consideradas de boa segmentação não continham ruptura entre constituintes do sintagma nominal, enquanto as linhas qualificadas como mal segmentadas apresentavam quebras entre substantivo/adjetivo, substantivo/sintagma preposicional, adjetivo/substantivo e determinante/substantivo.

Resultados de testes estatísticos, conduzidos a partir dos questionários de reconhecimentos de palavras, não revelaram nenhum efeito significativo da qualidade da legendagem nessa fase do experimento. A análise do movimento ocular também não identificou diferenças significativas no número, nem na duração de fixações em legendas de alta e baixa qualidade. Assim, nenhum dos testes indicou haver diferenças de processamento ou performance relacionadas à qualidade da segmentação das linhas de legenda,

visto que os participantes aparentemente processaram legendas bem e mal segmentadas da mesma forma e com os mesmos resultados. Perego et al (2010, p.21), portanto, conclui que as preocupações psicolinguísticas com a segmentação provavelmente são exageradas. Entretanto, Perego et al (2010) ainda assim sugere que, por questões estéticas que afetam a avaliação do espectador, legendistas mantenham um padrão na segmentação de linhas de legenda. Por fim, o estudo aponta que são necessárias futuras pesquisas empíricas com participantes de outras faixas etárias, de outros níveis de escolaridade/educação, SE, bem como a utilização de outros tipos de produtos audiovisuais para averiguar a validade e a eficácia dos guias de legendagem existentes e então chegar-se a generalizações.

As obras de Karamitroglou (1998) e Perego (2008) também figuram como fundamentação teórica na pesquisa apresentada em Kalantzi (2008), a qual advoga por um estudo tripartite de grande escala que combine (i) descrição da legendagem produzida atualmente a partir de uma análise de corpus com (ii) pesquisa empírica sobre as necessidades e preferências de surdos e ensurdecidos e (iii) comparação dos dados encontrados com guias de legendagem existentes a fim de atualizá-los quando/se necessário. Entretanto, como não se propõe a ser uma pesquisa nesses moldes, a tese em Kalantzi (2008) dedica-se em maior proporção ao primeiro passo e em menor ao terceiro passo da pesquisa tripartite esboçada acima.

Kalantzi (2008) apresenta diretrizes existentes e pesquisas empíricas realizadas a respeito de seis características de legendagens: o tempo de duração; tempo de intervalo entre legendas consecutivas; número de linhas; segmentação; redução e outras estratégias usadas no ato de legendar. Para fins de comparação são apresentadas diretrizes usadas na LSE e na legendagem interlingual. Segundo Kalantzi (2008), as diretrizes são encontradas em guias e documentos sobre a prática da legendagem criados por profissionais, organizações de comunicação e pesquisadores/legendistas com o objetivo comum de garantir o mínimo de qualidade à produção de legendagem. Apesar de prescritivos, esses documentos não são considerados regras absolutas, mas sim recomendações e sugestões passíveis de questionamento e reflexão.

Para a autora, no que tange à descrição de legendagem, ainda há pouca pesquisa dedicada a medir e descrever a legendagem disponível e faltam pesquisas de larga escala que permitam que generalizações sejam feitas sobre a modalidade. A situação fica ainda mais restrita no contexto da LSE, visto que são pouquíssimas as pesquisas empíricas da área, sendo que essas, como De Linde e Kay (1999), são muito pequenas em termos de número de programas, número de legendas por programa e de participantes envolvidos. A autora cita então uma pesquisa conduzida em 2005 no Reino Unido, pela empresa *Office Communications*, que confirma que a questão da velocidade de legendas é extremamente complexa e influenciada pelo grau de surdez, idade, dependência da legendagem e nível de letramento do público (KALANTZI, 2008, p.55).

Segundo os vários guias consultados pela pesquisadora, em geral, as diretrizes de LSE indicam uma preferência pelo máximo de duas linhas por legenda, mas não se sabe de nenhuma pesquisa descritiva que tenha investigado tal aspecto da legendagem. Conforme Kalantzi (2008), guias de legendagem interlingual recomendam que sempre que possível é preferível colocar o texto da legenda em uma única linha, mas nem sempre isso é possível ou desejável; o mais importante, então, é segmentar o texto de forma a melhorar a legibilidade. Sendo que as mesmas regras de segmentação se aplicam tanto entre legendas, como dentro da mesma legenda. Cuidados na segmentação implicam em seguir considerações gramaticais e sintáticas, a fim de que palavras ligadas por semântica, lógica ou gramática estejam na mesma linha e que o conteúdo seja lido de maneira mais natural e completa possível. A autora reforça ainda que o trabalho de Perego (2008) é uma das duas únicas pesquisas que abordam o tema da segmentação em legendagem e trata de LSE (intralingual) e de legendagem interlingual. Conforme Kalantzi (2008), Perego (2008) fornece evidências teóricas e empíricas de que uma segmentação linguisticamente coerente pode beneficiar o processamento do texto.

O corpus da pesquisa relatada em Kalantzi (2008) consta de dois gêneros de programas audiovisuais (documentários e atualidades). O corpus consta apenas dos roteiros dos programas e das legendas com seus tempos de marcação. Contudo, por questões técnicas, tem a limitação de não constar

do vídeo em si, o que impossibilitou o estudo da velocidade de fala e da sincronização, bem como das informações visuais e suas relações com os tempos de marcação de legendas, com as eliminações e com a segmentação. Por esses fatores Kalantzi (2008) não se propõe a fazer um estudo em larga escala, mas sim, pretende fornecer um modelo para uma pesquisa de larga escala a partir de um corpus. A metodologia de análise do corpus se deu em três etapas:

(i) as estratégias de legendagem usadas; tempos de marcação; duração de legenda; número de caracteres por legenda; taxa de legendagem; e se limites de tempo foram respeitados; intervalo entre legendas consecutivas; número de palavras por legenda ou segmento de fala; número de palavras eliminadas por legenda; número de linhas por legenda; e erros de segmentação foram observados nos arquivos originais;

(ii) os arquivos originais de roteiro e legendas foram comparados e mesclados;

(iii) os arquivos mesclados foram manualmente anotados com as seguintes etiquetas: número de legenda; quebra de linha; erro de segmentação; parte do discurso; estratégias de legendagem usadas; duração de legenda; respeito a restrições de tempo; e legendas de três linhas. Devido ao acesso limitado ao produto audiovisual completo, ferramentas específicas, como *softwares* de legendagem foram parcialmente empregadas nos cálculos apresentados na pesquisa de Kalantzi, sendo muitos dos números encontrados resultantes de cálculos manuais.

Em um capítulo exclusivo, Kalantzi (2008) descreve e discute a segmentação das legendas estudadas, utilizando a metodologia dos ETBC. Kalantzi (2008) aponta que nos guias de legendagem a orientação sobre quebras de linha é insuficiente. Resultados quantitativos gerais da pesquisa apontam que 25% (vinte e cinco por cento) das legendas estudadas continham algum problema de segmentação. Todos os erros foram classificados de acordo com as categorias linguísticas envolvidas nos problemas. A partir dessa classificação, a pesquisadora sugere um conjunto de regras criadas primeiramente para orientar legendistas, mas que podem também ser mais bem elaboradas e usadas no desenvolvimento de uma importante ferramenta

de legendagem. Essas regras – criadas com base no sistema linguístico do inglês – serão enumeradas a seguir.

- Regras Sujeito e Predicado:

- a) Não separar sujeito de predicado se o sujeito for composto apenas de um pronome ou um sintagma nominal curto.
- b) Não separar sujeito de verbo, mas separar verbo de complemento/objeto se o sujeito for curto e o predicado for longo e os dois não caibam na mesma linha.
- c) Quebrar a linha no sintagma de ordem mais alta e colocar o sujeito e predicado em linhas separadas se os dois forem longos e não couber na mesma linha.

- Regra Verbo e Objeto/Complemento

- d) Não separar verbo de seu objeto ou complemento, a não ser no caso previsto pela regra b.

- Regras Sintagma Nominal

- e) Não separar determinante de substantivo.
- f) Não separar adjetivo do substantivo que modifica.
- g) Não separar nomes próprios, nem a titulação do nome.
- h) Não separar substantivos de uma sequência substantivo-substantivo.
- i) Não separar possessivo/genitivo do substantivo que modifica.

- Regra Sintagma Adverbial

- j) Não separar advérbio do adjetivo que modifica.

- Regras Sintagma Preposicional

- k) Não separar a preposição do resto do grupo preposicional.
- l) Não separar preposições complexas.

- Regras Sintagma Verbal

- m) Não separar advérbio do verbo que modifica.

- n) Não separar o marcador de infinitivo do infinitivo (*to verb*).
- o) Não separar formas verbais complexas.
- p) Não separar verbos frasais.

- Regras Subordinação

- q) Não separar a conjunção subordinativa do resto da oração subordinada.
- r) Não separar o pronome relativo/advérbio do resto da oração relativa
- s) No caso de haver uma oração principal e uma subordinada e as duas não caibam em uma linha, usar uma linha para cada oração.

- Regras Coordenação

- t) Não separar a conjunção coordenativa da oração que introduz
- u) No caso de haver duas orações principais e as duas não couberem na mesma linha, usar uma linha para cada oração.

- Outras Regras

- v) Não separar colocações
- w) Não hifenizar palavras

A quantidade de erros encontrados também levou Kalantzi (2008) a examinar os potenciais fatores que arrazoam a decisão de legendistas ao não criar uma quebra de linhas de legenda linguisticamente adequada, mesmo quando essa segmentação pudesse ser facilmente atingível. Os fatores identificados foram: (i) geometria, isto é, a tentativa de apresentar linhas de legenda de comprimentos iguais; (ii) linha superior mais curta para manter livre o máximo possível da imagem; (iii) preferência por legenda verbatim. Desses três, a pesquisadora concluiu que o fator que mais compete com considerações linguísticas é a geometria, aparentemente por razões puramente estéticas. Com base nisso, Kalantzi (2008) afirma que pesquisas futuras são essenciais para investigar se considerações linguísticas devem mesmo sempre ter precedência, ou se sob alguma circunstância essas podem ser postas de lado. Essas pesquisas, no entanto, devem extrapolar a abordagem descritiva e envolver legendistas e audiência, em particular SE, a fim de determinar não só as necessidades e preferências dos espectadores, mas também reconsiderar e

atualizar as diretrizes existentes que parecem divergir das atuais práticas adotadas na legendagem.

Apesar de serem resultado de uma pesquisa descritiva, as regras elaboradas em Kalantzi (2008), visto que foram criadas para constarem em guias de legendagem, assumem um aspecto prescritivo. Esse aspecto e o fato de terem sido elaboradas para o sistema linguístico do inglês impedem que as regras enumeradas acima sirvam de recurso para fundamentar a análise da gramática e da sintaxe da legendagem em escrutínio neste estudo.

Com base nos resultados encontrados em Kalantzi (2008), este trabalho investigará se o critério da geometria é também preferido na segmentação das legendas – tanto LO quanto LSE – do filme, e se esse critério também pode ser considerado o principal fator motivador de desvios linguísticos na quebra de linhas da legendagem estudada aqui.

Para fundamentar o estudo da estrutura gramatical da legendagem e reforçar o caráter descritivo desta pesquisa, foi necessário buscar uma gramática do português brasileiro. A última seção deste capítulo tratará desse tema. Antes disso, na seção seguinte, será apresentado o status da acessibilidade sensorial e da pesquisa em LSE no Brasil.

1.3) TAV - LSE NO BRASIL

No Brasil, a Portaria 310 aprovou a Norma Complementar 01/2006 determinando que até no ano 2018 todos os programas da televisão aberta sejam acessíveis a SE através de legendagem ou janela de libras – e também a deficientes visuais através da audiodescrição. Emissoras brasileiras de televisão têm disponibilizado um tipo de legendagem fechada para surdos – acessada por meio do controle remoto - que segue o modelo do sistema americano de *closed caption*. Esse sistema consiste na transcrição de quase todo discurso apresentado na televisão, transmitindo legendas que rapidamente aparecem e desaparecem na tela e raramente são sincronizadas com a imagem e a trilha sonora (diálogo).

Uma grande preocupação de pesquisadores brasileiros da área é o aspecto da velocidade de leitura de legenda, visto que atualmente a velocidade

praticada está em torno de 200 palavras por minuto. Entretanto, deve-se levar em consideração que surdos têm dificuldade de leitura, e que a legenda demanda deles habilidades de leitura avançadas em PB, a qual não é a primeira língua da maior parte dessa audiência.

Até o presente momento, não houve no Brasil pesquisa dedicada especificamente a determinar a velocidade de leitura de legenda ideal para SE. Contudo, a fim de averiguar a inadequação do status atual da LSE realizaram-se três estudos na Universidade Estadual do Ceará (UECE), cujo objetivo foi analisar como se dava a recepção dessa legendagem por parte do público alvo, para então identificar as características que gerassem insatisfação e assim propor soluções que conduzissem à formulação de parâmetros técnicos para compor um modelo brasileiro de LSE que esteja mais próximo de promover a acessibilidade, e não apenas cumprir a lei.

Como procedimentos metodológicos desses estudos (ARAÚJO, 2004, 2007, 2008; FRANCO; ARAÚJO, 2003) sujeitos SE assistiram a algumas sequências de programas variados e então responderam questionários para testar suas habilidades em combinar significados de conteúdo e imagem. Uma grande contribuição desses primeiros estudos foi a identificação de algumas preferências dos surdos que podem vir a ser definidos como parâmetros da LSE brasileira: (i) fonte amarela em fundo transparente; (ii) identificação de falantes e efeitos sonoros apresentados entres colchetes; (iii) voz vinda de fora da tela deve aparecer em itálico; (iv) português formal é o registro preferido. Apesar de não conclusivos, os dados coletados até então indicam que a LSE ainda tem um longo caminho pela frente antes de chegar a resultados mais satisfatórios para o público. Tais usuários fizeram importantes contribuições para esses primeiros estudos e ainda têm muitas contribuições substanciais a fazer. Por isso, mais pesquisas empíricas de descrição e de recepção, maior interesse acadêmico e embasamento teórico são fundamentais para o planejamento de um modelo brasileiro de LSE.

À vista disso, uma pesquisa exploratória está sendo realizada com surdos das cinco regiões do Brasil. No artigo Araújo e Nascimento (2011) são relatados os achados iniciais dessa pesquisa no nordeste do Brasil. Os experimentos da pesquisa investigam a recepção de legendas elaboradas com os parâmetros mencionados acima, e com algumas variações que envolvem o

posicionamento das legendas (colocadas sobre o falante), uso de cores para apresentar informações não verbais e três diferentes velocidades de leitura (145, 160 e 180 palavras por minuto). Os primeiros resultados indicam que as legendas de velocidade mais baixa (145 palavras por minuto) são mais adequadas para o público surdo. Entretanto, as legendas mais rápidas também obtiveram boa recepção, sugerindo que o grau de condensação e uma segmentação criteriosa facilitaram a recepção das legendas, ainda que de alta velocidade (ARAÚJO E NASCIMENTO, 2011, p.15). Em busca de resultados conclusivos, essa pesquisa ainda continua com surdos de todo o país, a fim de chegar-se a um modelo brasileiro de LSE.

A seção a seguir detalha a busca por uma gramática e justifica a escolha da gramática gerativa como fundamento para apoiar a descrição da segmentação dos sintagmas do PB.

1.4) GRAMÁTICA DOS SINTAGMAS DO PB

Por pressupor uma metodologia de trabalho que envolva coleta de dados, a gramática funcional seria a linha de pensamento ideal para fundamentar a análise de corpus proposta neste estudo. A abordagem funcionalista lida com o uso da língua. Isto é, concentra-se em estudar a competência comunicativa que permite aos seus falantes interpretar e emitir expressões numa determinada língua tendo como referência uma situação comunicativa inteira. Dessa maneira, a gramática funcionalista encontra seu objeto de estudo em instanciações da língua em uso, coletados em contextos reais de discurso, e organizados em acervos ou corpora para serem estudados. Sendo assim, mais indicada para dialogar com um estudo de corpus, a linha da gramática funcional de Neves (2003), foi a primeira tentativa na busca por uma descrição dos sintagmas do PB. Entretanto, Neves (2003) apenas menciona e define a função sintática dos itens menores que a oração, mas não oferece uma descrição da estrutura dos sintagmas que pudesse ser discutida à luz dos casos de segmentação encontrados no corpus desta pesquisa.

À vista disso e amparada pela análise sintática em diagrama de árvore adotada em Kalantzi (2008), esta pesquisa buscou fundamentação na

gramática gerativista. Essa gramática pressupõe a dicotomia chomskyana competência X desempenho, cujos estudos contam com a capacidade de intuição dos falantes e de julgamentos de gramaticalidade que usuários podem fazer acerca de uma língua. Assim, o gerativismo trata de hipóteses sobre o conhecimento que o falante tem da sua língua, na busca por explicitar o conjunto de regras subjacentes que forneceria a base para o uso real da língua. Perini (2006) foi uma das obras consultadas na tentativa de encontrar fundamentação teórica para esta pesquisa, porém o estudo descreve apenas dois tipos de sintagmas, a saber, sintagma nominal, sintagma adjetivo. E apenas menciona brevemente sem descrever, o sintagma adverbial, sobre o qual o autor afirma que ainda não há muito que dizer a respeito. Além disso, o nível de análise e a nomenclatura propostos na descrição dos outros dois sintagmas se mostraram de uma delicadeza profunda; no entanto, excessivos para as necessidades desta pesquisa, a qual requer apenas noções básicas e gerais para uma descrição da estrutura dos sintagmas do PB.

O suporte adequado foi encontrado, por fim, em Lemle (1989) que descreve cada um dos cinco sintagmas do PB (nominal, adjetivo, preposicional, adverbial e verbal), em forma de regra do sintagma que expõe a estrutura desses constituintes menores que a sentença. Afiliado à gramática gerativa, Lemle (1989), tem como objetivo fazer contribuições linguísticas teóricas ao ensino escolar da gramática a fim de reaproximar as duas áreas de trabalho. Lemle (1989) propõe uma análise gramatical que una a identificação das categorias de palavras à verificação dos agrupamentos sintagmáticos, descendo até o nível da palavra, através de diagramas em árvore. A pesquisadora começa tratando das categorias lexicais e suas dificuldades de classificação até chegar às estruturas sintáticas no nível da sentença. No processo de descrever o PB, a autora considera que a questão central da descrição da língua é como se dá o agrupamento das unidades lexicais na formação das sequências significativas, denominadas sintagmas. Portanto, é discutida a estrutura da subdivisão interna a que todos os tipos de sintagmas se submetem. Para descrever a estrutura de cada sintagma, Lemle (1998) adota a seguinte notação:

Nome ----- N

Adjetivo ----- Adj
 Determinante ----- Det
 Quantificador ----- Quant
 Verbo ----- V
 Preposição ----- Prep
 Advérbio ----- Adv
 Sentença ----- S

Além disso, os parênteses indicam que um item é facultativo e o asterisco à direita indica a ocorrência de uma cadeia ilimitada de itens da classe sinalizada. Toda essa notação será utilizada nas regras sintagmáticas apresentadas a seguir.

Para o sintagma nominal, Lemle (1989) apresenta a seguinte regra:

$RSN \rightarrow (\text{Quant}(\text{Det})) (\text{Adj})^* \text{N} (\text{Adj})^* (\text{SPrep})^* (\text{Adj})^* (\text{S})^*$

Segundo a regra o único elemento obrigatório é o núcleo nominal (N). Nessa regra, em especial, os parênteses dos símbolos Quant e Det sinalizam que a presença do primeiro requer a presença do segundo.

A parte em destaque - exemplo [1] - da linha de legenda extraída de IF ilustra a estrutura do SN.

[1] Esse homem é **uma criatura renovada no amor de Jesus...**

Para o sintagma adjetivo tem-se a seguinte regra:

$RSAdj \rightarrow (\text{Adv}) \text{Adj} (\text{Adv}) (\text{SPrep})^* (\text{S})^*$

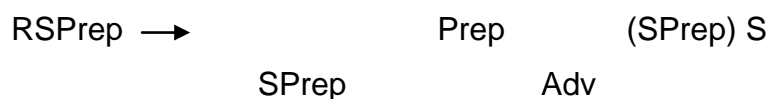
Segundo a regra o único elemento obrigatório é o adjetivo (Adj). Lemle (1989) ainda aponta que é mais comum que os advérbios à direita e à esquerda do adjetivo, geralmente sejam de intensidade, tempo ou lugar.

A parte em destaque no exemplo [2], da linha legenda extraída de IF, representa a estrutura do SAdj.

[2] Estamos todos **muito felizes com a presença de Paulo** aqui

Para o sintagma preposicional, a regra é a seguinte:

$\left(\left\{ \begin{array}{c} \text{Adv} \\ \text{SN} \end{array} \right\} \right)$



Segundo a regra, o núcleo preposicional pode ocorrer como um elemento único, representado por Prep, como exemplificado em “João dormiu fora.”, cuja preposição *fora* é tratada em Lemle (1989) como preposição intransitiva, juntamente com outras como *antes*, *depois*, *acima*, *atrás*, *além*, e *abaixo*, pretendendo assim pôr um fim na complicação de classificar tais palavras em duas classes, a dos advérbios e a das locuções prepositivas, como encontrado em gramáticas didáticas. Essa proposta implica, conseqüentemente, em tratar preposições também como transitivas, como ilustrado em “João dormiu fora de casa.” Com exemplos como “para com os colegas”, ou “para fora de casa” a descrição abrange casos de locuções prepositivas. O símbolo Adv no início da regra trata da posição de modificador, que em geral quando realizada, traz uma expressão de quantidade ou de intensidade, como em “muito sem graça” ou “quase de graça”. Já o símbolo SPrep do início, representa sintagmas preposicionais modificado por outro sintagma preposicional, como em “em ponto às três horas”.

A parte em destaque no exemplo [3], da linha de legenda retirada de IF, representa a estrutura do SPrep.

[3] Mas me converti **depois de meu encontro com o Senhor**.

Para o sintagma adverbial, Lemle (1989) não apresenta uma regra. Ao admitir a distinção entre preposições transitivas e intransitivas, pode-se verificar que sintagmas preposicionais e advérbios desempenham papéis semânticos idênticos e ocupam posições sintáticas idênticas. Assim, a autora retoma os contextos possíveis para o sintagma preposicional e mostra que advérbios podem assumir as mesmas posições no papel de modificador. Para os verbos modificados por advérbios Lemle (1989) ilustra com “trabalhe atentamente”, cujo sinônimo “trabalhe com atenção” é um verbo modificado por um sintagma preposicional. Para os adjetivos modificados por advérbios, tem-se “já morto” como sinônimo de “morto desde algum tempo”, que traz um adjetivo modificado por sintagma preposicional. Há também o sintagma

preposicional modificado por advérbio, como em “excessivamente com cuidado” que pode ser modificado por outro sintagma preposicional: “com cuidado em excesso”. Um último contexto trata dos nomes modificados por advérbios, como em “a votação agora” que também podem ser modificados por sintagmas preposicionais, como em “a votação neste momento”. Dessa forma, Lemle (1989) entende o sintagma preposicional como uma forma de expansão indefinida das possibilidades de expressão da modificação adverbial. A autora ainda aponta que o movimento inverso não é verdadeiro, uma vez que há contextos em que ocorre sintagma preposicional, mas não ocorreria advérbio sinônimo. Sendo assim, Lemle (1989) considera que a classe mais abrangente é a das preposições e que os advérbios, por sua vez, é que formariam uma subclasse das preposições.

A parte em destaque no exemplo [4], da linha de legenda retirada da LO de Irmãos de Fé, representa a estrutura do SAdv, nesse caso modificando um verbo.

[4] Prefiro ir **imediatamente à sinagoga**.

Para o sintagma verbal, a regra é a seguinte:

$$RSV \rightarrow (Adv)^* V (Adv)^* \left[\begin{array}{c} SN \\ Adj \end{array} \right] (SPrep)^* (S)^*$$

Segundo a regra, o único termo obrigatório é o núcleo verbal (V). Outro ponto importante sobre essa regra é que a escolha entre SN ou Adj é determinada pelo tipo de verbo, ou seja, verbo seguido de objeto direto ou verbo copulativo seguido de adjetivo.

A parte em destaque no exemplo [5], da linha de legenda retirada de IF, ilustra a estrutura do SV.

[5] O Paulo **já tinha até abandonado a escola**.

A descrição em Lemle (1989) dos sintagmas do PB foi então adotada como base para descrever as ocorrências de segmentação de sintagmas nas legendas do corpus e assim, dialogar com as regras apresentadas em Kalantzi (2008).

No próximo capítulo, descreve-se o corpus e a metodologia adotados nesta pesquisa.

CAPÍTULO 2

METODOLOGIA

O corpus deste estudo é constituído de LSEs e LOs de um filme brasileiro. O procedimento metodológico adotado envolve a escolha e decifração dos filme, extração das legendas LO e LSE, correção e preparação dessas em arquivos de extensão *txt* para subsequentes análises quantitativas e qualitativas, conforme detalhado nas próximas subseções.

2.1) CORPUS

O filme escolhido para esta pesquisa faz parte do ESTRA (Corpus para Estudo de Estilo em Tradução) criado no LETRA da Faculdade de Letras da UFMG, integrando um de seus subcorpora. O LETRA ainda possui uma coletânea composta por outros 35 filmes legendados e/ou audiodescritos em vários idiomas que poderiam vir a integrar o ESTRA.

Uma vez que a proposta deste trabalho é de descrever uma pequena parte da produção de legendagem em Português no Brasil, o primeiro requisito adotado na seleção de textos fílmicos foi que o filmes fosse uma produção originalmente brasileira, com áudio original em PB. Com base no estágio primevo da cinematografia brasileira em si, é possível concluir que a recente preocupação com a acessibilidade sensorial ainda esteja surtindo pouco impacto na produção de LSE no Brasil, e, portanto, as opções de filmes brasileiros sensorialmente acessíveis no momento são bastante reduzidas. Foi selecionado, então, o filme *Irmãos de Fé* (IF) de Moacyr Góes.

O filme IF conta com LO e LSE criadas por autor não identificado nos créditos da produção. A obra de Góes foi escolhida por ser o único filme brasileiro publicado comercialmente em DVD acessível até o início desta pesquisa - isto é, contém LSE e audiodescrição para cegos com audionavegação - que consta do acervo do LETRA. Além disso, o filme também foi disponibilizado em DVD com LO, permitindo, portanto, um trabalho de análise comparativa entre esses dois tipos de legenda elaborados em PB.

Devido à incipiência das pesquisas na interface ETBC e legendagem, e apesar de comporem um corpus pequeno - menos de 80 mil palavras, segundo Berber Sardinha (2004) - o texto fílmico selecionado para o estudo configuram uma amostra representativa do material audiovisual até então disponível no Brasil, visto que é uma produção completa em termos de acessibilidade sensorial e que atende às necessidades mais emergenciais das pesquisas na área. Nas subseções abaixo, detalha-se o filme que compõe o corpus deste estudo.

2.1.1) FILME IRMÃOS DE FÉ

A Figura 3 abaixo mostra a capa do DVD do filme *Irmãos de Fé*:



FIGURA 3 - Capa do DVD Irmãos de Fé

Fonte: <http://capasbr.blogspot.com.br/2008/09/irmos-de-f.html>

Trata-se de um drama com duração de 105 minutos dirigido por Moacyr Góes, produzido no Brasil, lançado e distribuído em 2004 pela *Columbia Pictures*. No elenco conta com Thiago Lacerda, Padre Marcelo Rossi, Sabrina Rosa e Micael Borges como protagonistas.

Irmãos de Fé conta a história de Paulo (Micael Borges), um adolescente rebelde e violento internado na FEBEM sob a acusação de sequestro, que acredita que sua irmã Mariana (Sabrina Rosa) o delatou. Mariana pede ajuda ao padre (Padre Marcelo Rossi) para cuidar de seu irmão já que ele recusa toda ajuda que ela lhe oferece. O padre dá uma Bíblia de

presente ao menino, que inicialmente a recusa, mas depois passa a ler com curiosidade a história recomendada pelo padre no livro de *Atos dos Apóstolos* sobre o apóstolo Paulo (Thiago Lacerda), um ávido perseguidor dos cristãos que se torna um dos pilares da igreja cristã após um encontro sobrenatural com Jesus ressurreto e assunto ao céu. O filme, portanto, conta as histórias de dois Paulos e num movimento de digressão fílmica, cenas da vida do apóstolo Paulo são mostradas à medida que o menino Paulo lê a Bíblia.

Na próxima seção, discorre-se sobre os procedimentos metodológicos adotados nesta pesquisa.

2.2) PROCEDIMENTOS DE EXTRAÇÃO E COMPILAÇÃO

Nesta seção, apresentam-se os procedimentos usados para a extração do filme, a compilação do corpus de LSE e LO e a utilização das ferramentas dos *Wordsmith Tools*® e do *Subtitle Workshop*® para levantamento de dados e sua subsequente análise.

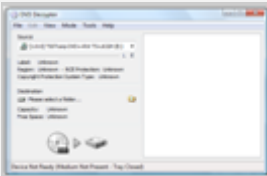

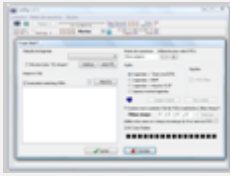

2.2.1) EXTRAÇÃO DO FILME

Após a seleção do filme o primeiro procedimento é a extração desse do DVD, a fim de que possa ser aberto no programa *Subtitle Workshop*®. Em seguida, os arquivos de legenda são obtidos com o auxílio de softwares específicos para a extração desses, conforme apresentado a seguir.

2.2.1.1) SOFTWARES UTILIZADOS

O Quadro 2 a seguir mostra os softwares utilizados na compilação do corpus da pesquisa. As funções de cada programa serão detalhadas na subseção seguinte.

QUADRO 2
Softwares utilizados na pesquisa

PROGRAMA	DVD DECRYPTER	FORMAT FACTORY	SUB RIP	SUBTITLE WORKSHOP
				
EMPRESA	Lightning UK!	Free Time©	Brain	URU Soft Subtitle
DATA	2005 - descontinuado	2008-2010	Sem informações	2001-2004
TIPO	Software gratuito	Software gratuito	Software Livre	Software gratuito
AMBIENTE	Windows	Windows	Windows	Windows
FUNÇÕES	Facilitar a remoção das restrições de cópia para copiar os conteúdos de DVDs e criar arquivos de imagens	Transcoder de mídia multifuncional, com funções de conversão, reparação e redução de arquivos de imagem, áudio e vídeo.	Extração de legendas com seus respectivos tempos de marcação como arquivos de texto.	Criação e edição de legendas e é largamente utilizado por legendistas amadores e profissionais.

Já o programa *Wordsmith Tools*© 5.0, uma coleção de ferramentas que permite a procura por padrões em textos, desenvolvido por Mike Scott, publicado em 1996 e comercializado pela Oxford University Press, foi usado com licença para um usuário. Este programa é um dos mais usados em pesquisas da subárea de ETBC (BAKER 1993, 1995, 1996). Entre as ferramentas usadas nesta pesquisa estão a *WordList*, que retorna dados quantitativos como frequência de palavras e variedade lexical pela razão type/token e *Concord*, que exhibe o contexto de palavras ou frases pesquisadas.

2.2.1.2) COMPILAÇÃO DO CORPUS

Após a escolha do filme, fez-se necessário compilar os dados extraído os arquivos de filme e de legenda. Para tanto, os programas apresentados no Quadro 2 foram baixados gratuitamente da internet. O filme foi extraído pelo programa *DVD Decrypter* e depois transformados em formato avi com o programa *Format Factory*. As legendas foram extraídas via OCR (Optical Character Recognition) com o programa *Sub Rip* e salvas em formato srt. Em seguida, os arquivos de legenda foram corrigidos manualmente no programa *Subtitle Workshop*® em comparação com as legendas dos filmes, a fim de verificar possíveis falhas de leitura de caracteres cometidas pelo programa *Sub Rip*. Erros de digitação ou de falta de edição identificados nas legendas que aparecem na tela, ao assistir-se o DVD original, foram mantidos.

Os arquivos de legenda srt, além de usados no *Subtitle Workshop*®, podem ser acessados também em formato txt, nesse formato as marcações de tempo foram retiradas, a fim de que os arquivos de legenda fossem processados pelo *Wordsmith Tools*® 5.0.

2.3) PROCEDIMENTOS DE ANÁLISE

A análise das legendas se dará em duas etapas, primeiramente uma análise quantitativa seguida de uma análise qualitativa, ambas descritas nas próximas subseções.

2.3.1) ANÁLISE QUANTITATIVA

O corpus de estudo é composto por dois arquivos: (i) legenda para ouvintes do filme *Irmãos de Fé*, nomeado LO_IF; e (ii) legenda para surdos e ensurdecidos do filme *Irmãos de Fé*, nomeado LSE_IF. Tal corpus poderia ser considerado insuficiente para uma análise de corpus, no entanto, dado o escopo da pesquisa e a não existência, até o momento, de filmes comerciais com LSE, não é possível usar um corpus de maior dimensão. Entretanto, ainda

assim, as ferramentas de corpus podem ser úteis para a análise e descrição das legendagens até então produzidas em PB.

Na próxima subseção, apresentam-se os procedimentos de anotação do corpus.

2.3.2) ANOTAÇÃO DO CORPUS

A análise da segmentação nas legendas do corpus foi feita por meio da inserção manual de etiquetas pré-estabelecidas, com base na classificação dos sintagmas do PB por Lemle (1989). Foram inseridas etiquetas antes e após cada ocorrência de segmentação identificada. Essas etiquetas são idênticas exceto pela presença de uma barra (/) na etiqueta posicionada ao final da ocorrência, para que fosse delimitado o começo e o fim de cada incidência. A inserção de etiquetas serviu ao propósito da análise quantitativa dos dados.

Para identificar-se cada uma das linhas de legenda, essas receberam a etiqueta <L>, ou seja, a primeira linha recebeu a etiqueta <L1>, a segunda <L2> e assim por diante, como exemplificado na Figura 4 abaixo.

```
LSE_IF - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda
|<NOME: Legendagem para Surdos e Ensurdecidos Irmãos de Fe>
|<AUTOR: C olumbia Pictures>
|<ANO: 2004>
|<LAN ADO E DISTRIBUIDO: Brasil>
|<IDIOMA ORIGINAL: Portugu s Brasileiro>
|
|<L1> [m sica com voz feminina] </L1>
|<L2> [estrondo] </L2>
|<L3> [Paulo, em off] <LS> Ainda que eu fale as l nguas </LS> </L3>
|<L4> dos homens e dos anjos, </L4>
|<L5> se n o tiver amor, </L5>
|<L6> ser i como o sino que ressoa <SOCN> ou como o prato que retine. </SOCN> </L6>
|<L7> Ainda que eu tenha o dom da profecia <SOCN> e saiba todos os mist rios </SOCN> </L7>
|<L8> e todo o conhecimento e tenha <SSVD1> uma f  capaz de mover montanhas, </SSVD1> </L8>
|<L9> se n o tiver amor, nada ser i. </L9>
|<L10> Ainda que eu d  aos pobres <SSVD1> tudo o que possuo, </SSVD1> </L10>
|<L11> e entregue meu corpo <SOSN> para ser queimado, </SOSN> </L11>
|<L12> se n o tiver amor, </L12>
|<L13> nada disso valer . </L13>
|<L14> [estrondo] </L14>
|<L15> [porta batendo] </L15>
|<L16> [estrondo] </L16>
|<L17> [som met lico] </L17>
|<L18> [gritaria] <LS> [som de passos apressados] </LS> </L18>
|<L19> [estrondo] </L19>
|<L20> [m sica de fundo] </L20>
|<L21> -T  com medo? <LS> -N o, p ! </LS> </L21>
|<L22> -Tu   sujeito homem, n ? <LS> -Sou, claro! </LS> </L22>
|<L23> -Ent o, j   . <LS> -Ai, cara. </LS> </L23>
|<L24> vai. </L24>
|<L25> [motor do carro] </L25>
|<L26> [m sica de a o] </L26>
|<L27> Sai, sai! </L27>
|<L28> Vambora! Vambora! Vambora! <LS> Sai! Sai! Sai do carro! </LS> </L28>
|<L29> Sai do carro! Sai, agora! </L29>
|<L30> -C lma nada! <LS> -N o fica nervoso. </LS> </L30>
|<L31> [m sica de a o cont ua] </L31>
|<L32> N s vamos fazer tudo <SOSN> que voc s quiserem. </SOSN> </L32>
|<L33> -Vai! Vai! Vai! Vai! <LS> -Vamos fazer o que quiserem. C lma. </LS> </L33>
|<L34> C la a boca! </LS> [m sica de a o cont ua] </L34>
|<L35> Tem dinheiro dentro do carro? </L35>
|<L36> Tem, n o. <LS> Tem tudo dentro da minha bolsa. </LS> </L36>
|<L37> T  tudo a  dentro. Se estiver <SSVD2> escondendo alguma coisa, vai bailar! </SSVD2> </L37>
|<L38> -N o t  escondendo nada, n o! <LS> -Vai bailar! </LS> </L38>
|<L39> Homem de f , for a! </L39>
|<L40> Esp rito </L40>
|<L41> -C nta tamb m! <LS> -Esp rito</LS> </L41>
|<L42> Esp rito Santo </L42>
|<L43> Esp rito Santo <LS> [multid o canta] </LS> </L43>
|<L44> Arrepi ! <LS> [multid o canta] </LS> </L44>
|<L45> Esp rito </L45>
|<L46> Esp rito Santo. </L46>
|<L47> Juntos, juntos! <LS> Esp rito! </LS> </L47>
|<L48> Esp rito </L48>
|<L49> Esp rito </L49>
```

FIGURA 4 – Exemplos de identifica o das linhas de legenda.

No caso da segmentação de estruturas completas, como por exemplo, duas sentenças, cada uma dessas colocada em uma linha, as etiquetas utilizadas para rastrear esse tipo de segmentação são: <LS> (início de legenda segmentada) e seu respectivo encerramento representado por </LS>. Conforme ilustrado na Figura 5 seguinte, que traz o resultado na ferramenta *Concord* de uma pesquisa por <LS>.

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	lead	lead	Sec	Sec	File	%
1	voz feminina] [estrondo] [Paulo, em off] <LS> Ainda que eu fale as línguas dos			8	023%		0	0%		0	0%		LSE_IF.txt	0%
2	[estrondo] [som metálico] [gritaria] <LS> [som de passos apressados]			97	274%		0	1%		0	1%		LSE_IF.txt	1%
3	[música de fundo] -Tá com medo? <LS> -Não, pô! -Tu é sujeito homem,			108	333%		0	1%		0	1%		LSE_IF.txt	2%
4	-Não, pô! -Tu é sujeito homem, né? <LS> -Sou, claro! -Então, já é. -Aí, cara,			115	533%		0	1%		0	1%		LSE_IF.txt	2%
5	homem, né? -Sou, claro! -Então, já é. <LS> -Aí, cara. Vai. [motor do carro]			120	733%		0	2%		0	2%		LSE_IF.txt	2%
6	Sai, sai! Vambora! Vambora! Vambora! <LS> Sai! Sai! Sai do carro! Sai do			134	1230%		0	2%		0	2%		LSE_IF.txt	2%
7	Sai do carro! Sai, agora! -Calma nada! <LS> -Não fica nervoso. [música de			146	187%		0	2%		0	2%		LSE_IF.txt	2%
8	que vocês quiserem. -Vail Vail Vail Vail <LS> -Vamos fazer o que quiserem.			164	2317%		0	2%		0	2%		LSE_IF.txt	2%
9	dinheiro dentro do carro? Tem, não. <LS> Tem tudo dentro da minha bolsa.			184	2814%		0	2%		0	2%		LSE_IF.txt	2%
10	bailar! -Não tô escondendo nada, não! <LS> -Vai bailar! Homem de fé, força!			206	3233%		0	3%		0	3%		LSE_IF.txt	3%
11	de fé, força! Espírito -Canta também! <LS> -Espírito Espírito Santo Espírito			215	3511%		0	3%		0	3%		LSE_IF.txt	3%
12	-Espírito Espírito Santo Espírito Santo <LS> [multidão canta] Arrepial			220	3537%		0	3%		0	3%		LSE_IF.txt	3%
13	Espírito Santo [multidão canta] Arrepial <LS> [multidão canta] Espírito Espírito			223	3617%		0	3%		0	3%		LSE_IF.txt	3%
14	Espírito Espírito Santo. Juntos, juntos! <LS> Espírito! Espírito Espírito Multidão			230	3830%		0	3%		0	3%		LSE_IF.txt	3%
15	Espírito Multidão -Espírito Santo. <LS> Padre: -Vem controlar. Todo o			235	4025%		0	3%		0	3%		LSE_IF.txt	3%
16	. Todo o meu ser. [música de ação] <LS> Imão! Direto pro banco! Mas eu			245	4139%		0	3%		0	3%		LSE_IF.txt	3%
17	Eu tô preocupada! Por misericórdia! <LS> Pelo amor de Deus! Meu marido			261	4620%		0	3%		0	3%		LSE_IF.txt	4%
18	Meu marido na mala. Ele é cardíaco! <LS> Ele pode ficar sufocado. Pelo			272	4920%		0	4%		0	4%		LSE_IF.txt	4%
19	, clamando: Espírito! [multidão canta] <LS> Espírito Padre: -Força! [multidão			315	5530%		0	4%		0	4%		LSE_IF.txt	4%
20	Espírito Padre: -Força! [multidão canta] <LS> Espírito [multidão e padre			320	5625%		0	4%		0	4%		LSE_IF.txt	4%
21	Espírito [multidão e padre cantam] <LS> Espírito Santo de Deus! Padre:			325	5637%		0	4%		0	4%		LSE_IF.txt	5%
22	cantam] Espírito Santo de Deus! Padre: <LS> Espírito Santo de Deus! Multidão			330	5733%		0	4%		0	4%		LSE_IF.txt	5%
23	de Deus! Multidão canta: -Espírito! <LS> Padre: -Clame! Multidão canta:			337	5933%		0	4%		0	4%		LSE_IF.txt	5%
24	: -Clame! Multidão canta: -Espírito! <LS> Padre: -Deseje-o! Padre: -O poder			342	6133%		0	4%		0	4%		LSE_IF.txt	5%
25	Padre: -Deseje-o! Padre: -O poder, vai! <LS> Multidão: -Espírito Santo de			348	6317%		0	4%		0	4%		LSE_IF.txt	5%
26	Santo de Deus! Não pare! Força! Peça! <LS> Cantando: Espírito! Derrama a			357	6733%		0	5%		0	5%		LSE_IF.txt	5%
27	Cantando: Espírito! Derrama a unção! <LS> [multidão continua cantando]			362	698%		0	5%		0	5%		LSE_IF.txt	5%
28	continua cantando] Derrama o poder, <LS> como o apóstolo São Paulo,			368	6954%		0	5%		0	5%		LSE_IF.txt	5%
29	, invocando! Espírito Santo de Deus! <LS> [multidão continua cantando]			378	7120%		0	5%		0	5%		LSE_IF.txt	5%
30	[multidão continua cantando] Força! <LS> [multidão canta] Tá tudo certinho.			382	7217%		0	5%		0	5%		LSE_IF.txt	5%
31	[multidão canta] Tá tudo certinho. <LS> O caixa eletrônico é esse aí. Eu			387	7314%		0	5%		0	5%		LSE_IF.txt	5%
32	aí. Eu já vou parar. -Calma, calma. <LS> -Pára, pára! -Calma, calma. -Eu			399	7633%		0	5%		0	5%		LSE_IF.txt	6%

FIGURA 5 - Pesquisa por <LS> em IF retornada pela ferramenta *Concord*

As etiquetas utilizadas para rastrear os casos de segmentação dentro de uma mesma sentença são: <SOSN> (início segmentação de oração subordinada normal) e seu respectivo encerramento </SOSN>; <SOSD> (início segmentação de oração subordinada com desvio) e seu respectivo encerramento </SOSD>; <SOCN> (início segmentação de oração coordenada normal e seu respectivo encerramento); e <SOCD> (início segmentação de oração coordenada com desvio e seu respectivo encerramento). Conforme ilustrado na Figura 6 seguinte, que traz o resultado da ferramenta *Concord* de uma pesquisa por <SOSN>.

N	Concordance	Set	Tag	Word #	Sem	Sem	Par	Par	lead	lead	Sec	Sec	File	%
1	meu coração. Por favor, não diga nada <SOSN> que possa nos afastar. Não,			4,821	65330%				052%				LSE_IF.txt	62%
2	, preciso lhe falar sobre sentimentos <SOSN> que guardo em meu coração.			4,811	65238%				052%				LSE_IF.txt	62%
3	. A Lei não pode ser fardo insuportável <SOSN> que impeça um homem de			4,591	62557%				059%				LSE_IF.txt	59%
4	. Pois é, então, ele me disse também <SOSN> que se meu irmão tivesse um.			5,160	69637%				057%				LSE_IF.txt	66%
5	sem que para isso tenha ao menos <SOSN> que se submeter à circuncisão			5,097	68730%				056%				LSE_IF.txt	66%
6	. Não chore. Se Jesus nos uniu assim, <SOSN> é assim que devemos			4,890	66326%				053%				LSE_IF.txt	63%
7	Nunca! E nunca se verá um grego <SOSN> admitir coisa semelhantel Vê,			4,560	62270%				059%				LSE_IF.txt	59%
8	e sermos impiedosos com aqueles <SOSN> que trairam nossos pais. Meu			4,073	55975%				053%				LSE_IF.txt	53%
9	. Pode-se dizer tudo a seu respeito, <SOSN> menos que não é um homem			3,850	53324%				050%				LSE_IF.txt	50%
10	. Ele foi capaz de abandonar tudo <SOSN> para viver a palavra do Senhor.			3,793	52754%				049%				LSE_IF.txt	50%
11	tomou conta. Mas eu não vim <SOSN> para agradar a todos. Aqueles			4,397	60236%				057%				LSE_IF.txt	57%
12	Filho. E se esse é o seu grande plano, <SOSN> que Lei poderá contra ele?			4,234	57934%				055%				LSE_IF.txt	55%
13	. Se Deus não recusa homem algum, <SOSN> quem poderá ter esse poder?			4,191	57638%				054%				LSE_IF.txt	55%
14	com o padre! Pois é exatamente nele <SOSN> que eu estava pensando.			5,325	71636%				059%				LSE_IF.txt	68%
15	Lei, eu morri para a Lei. Morri pela lei <SOSN> a fim de viver para Deus! Eu			6,697	90540%				056%				LSE_IF.txt	86%
16	pela minha boca. Eu sei. Reúna tudo <SOSN> o que conseguiu arrecadar.			6,450	86843%				053%				LSE_IF.txt	83%
17	, já que momento mais apropriado <SOSN> não poderia existir, já que			6,266	85631%				051%				LSE_IF.txt	81%
18	, mas antes nós queremos bebidas <SOSN> para celebrar este momento.			7,295	94875%				054%				LSE_IF.txt	94%
19	para falar, vocês devem estar ávidos <SOSN> para fazer dela um outro usol			7,278	94738%				054%				LSE_IF.txt	94%
20	fé sem obras é vã? Não foi pelas obras <SOSN> que o nosso pai Abraão, foi			6,932	92626%				059%				LSE_IF.txt	89%
21	vocês. Tenho andado por várias cidades <SOSN> levando suas palavras. Muitas			6,165	84537%				079%				LSE_IF.txt	79%
22	amigo. Eu prefiro amortecer o corpo <SOSN> com o que resta desse vinho.			5,569	74830%				072%				LSE_IF.txt	71%
23	, de Teodora. Apesar das bordoadas, <SOSN> fomos felizes naquela cidade.			5,457	73250%				070%				LSE_IF.txt	70%
24	. Até porque não é só o seu irmão <SOSN> que precisa de ajuda aqui. E,			5,362	72034%				059%				LSE_IF.txt	69%
25	. Vem comigo. O Senhor me enviou <SOSN> para libertá-lo. [música suave			5,958	80838%				077%				LSE_IF.txt	76%
26	perigo] Herege! Deus cobra um preço <SOSN> a quem o desrespeita. E você			5,922	80536%				076%				LSE_IF.txt	76%
27	. Quando eu vendi tudo o que possuía <SOSN> para me juntar ao povo de			5,599	75140%				072%				LSE_IF.txt	71%
28	querido Saulo, aqui estão as cartas <SOSN> que você nos pediu. É tudo o			1,593	29131%				021%				LSE_IF.txt	22%
29	. Espero que quem roubou a Bíblia <SOSN> faça um bom proveito dela.			1,561	29116%				020%				LSE_IF.txt	21%
30	, vá pro inferno! Eu não quero saber <SOSN> se ela tá viva ou se tá morta.			1,547	29038%				020%				LSE_IF.txt	21%
31	de atenção. Ele precisa de alguém <SOSN> que o ensine a ser um homem			2,270	36812%				029%				LSE_IF.txt	31%
32	à Lei, e quem me ensinou a amá-la <SOSN> foi o senhor, mestre. Quando			1,672	29934%				022%				LSE_IF.txt	23%

FIGURA 6 - Pesquisa por <SOSN> em IF retornada pela ferramenta Concord

As regras gerais para cada tipo de sintagma de Lemle (1989) serviram de base para a criação das etiquetas utilizadas no rastreamento de legendas que apresentavam estrutura sintática diferente e, portanto, desvio dessas regras. Há etiquetas específicas para cada sintagma e cada categoria de desvio em cada tipo de sintagma. Cada uma dessas será apresentada e explicada nos quadros a seguir. No Quadro 3 seguinte, apresentam-se as etiquetas para a segmentação do sintagma nominal.

QUADRO 3
Etiquetas para o Sintagma Nominal

Sintagma Nominal	Etiqueta	Etiqueta (Início da Ocorrência)	Etiqueta (Final da Ocorrência)
Sintagma Nominal sem desvio	SSN	<SSN>	</SSN>
Separação de substantivo e adjetivo.	SSND1	<SSND1>	</SSND1>
Separação de determinante e substantivo.	SSND2	<SSND2>	</SSND2>
Separação de substantivo e sintagma preposicional modificador.	SSND3	<SSND3>	</SSND3>

Uma imagem da tela da ferramenta Concord do programa *Wordsmith Tools 5.0*® apresenta os resultados da pesquisa por <SSN que retorna todos os casos e tipos de segmentação de SN como mostrado na Figura 7 a seguir.

N	Concordance	Set	Tag	Word #	Sen	Para	lead	Sec	File	%
1	Esse livro contém muitas histórias <SSND3> de coragem, de amor. Por			839	17230%	011%		011%	LSE_IF.txt	11%
2	, de amor. Por sinal, uma história <SSND3> de um homem que tem o			847	17388%	011%		011%	LSE_IF.txt	12%
3	, ele vai ser uma excelente companhia <SSND3> para você, ainda mais aqui			878	17433%	011%		011%	LSE_IF.txt	12%
4	meu irmão vai ser uma chaga aberta <SSND3> em seu corpo. Você não			1,035	20330%	013%		013%	LSE_IF.txt	14%
5	Mas há uma coisa, meu querido <SSND1> Saulo, que me atormenta.			1,617	29434%	021%		021%	LSE_IF.txt	22%
6	. Voltarei de Damasco com um grande <SSND2> número de nazarenos			1,702	30134%	022%		022%	LSE_IF.txt	23%
7	e objetiva. Devemos chamar a atenção <SSND3> das autoridades romanas,			1,812	31233%	023%		023%	LSE_IF.txt	25%
8	romanas, essa é uma guerra <SSN> entre nós, judeus. Ninguém			1,819	31230%	023%		023%	LSE_IF.txt	25%
9	ficasse e fosse testemunha da mão <SSND3> de Jesus agindo em mim.			2,167	35257%	028%		028%	LSE_IF.txt	30%
10	Testemunha? Sim, Judas. Testemunha <SSND3> de que Jesus, O Nazareno,			2,176	35520%	028%		028%	LSE_IF.txt	30%
11	a luz do Senhor, você será testemunha <SSND3> de tudo que viu e do que viveu			2,327	37036%	030%		030%	LSE_IF.txt	31%
12	Senhor. Reconheço. Reconheço Jesus, <SSN> como o filho de Deus, o			2,364	37314%	030%		030%	LSE_IF.txt	32%
13	Eu te batizo. Como uma nova criatura <SSND3> no Cristo Jesus. -você será			2,405	37538%	031%		031%	LSE_IF.txt	33%
14	. Esse homem é uma criatura <SSND1> renovada no amor de Jesus.			2,447	37956%	032%		032%	LSE_IF.txt	33%
15	. Se antes a raiva, a obstinação e a <SSND2> cegueira me levaram a um			2,642	39626%	034%		034%	LSE_IF.txt	35%
16	do que eu arredar de apenas uma <SSND2> só palavra proferida em			2,701	39931%	035%		035%	LSE_IF.txt	36%
17	, Tiago, João e os outros apóstolos <SSN> que o Senhor escolheu, para			2,772	40357%	036%		036%	LSE_IF.txt	37%
18	, temos aqui então uma evidência <SSND3> da Sua vontade, do Seu			2,972	43273%	038%		038%	LSE_IF.txt	40%
19	, homem? Ester, dê um pouco <SSND2> de água pra ele. Não precisa			2,986	43436%	038%		038%	LSE_IF.txt	40%
20	, Ester. Essa notícia sobre a conversão <SSND3> de Saulo de Tarso. -Eu não			3,000	43630%	039%		039%	LSE_IF.txt	40%
21	O Senhor tem muitas moradas, <SSN> meu amigo. Esta aqui será			3,119	45675%	040%		040%	LSE_IF.txt	42%
22	é Barnabé. Sou irmão de Pedro, Tiago <SSN> e os outros apóstolos de Cristo.			3,231	47130%	042%		042%	LSE_IF.txt	43%
23	uma missão. Eu preciso que Pedro, <SSN> Tiago e João me recebam. Nós			3,285	47630%	042%		042%	LSE_IF.txt	44%
24	muitas, muitas. Homens, mulheres, <SSN> gente de todos os lugares. Eu			3,496	49638%	045%		045%	LSE_IF.txt	46%
25	tradições. Deus não faz distinção <SSN> entre os homens. Estamos			3,680	51533%	047%		047%	LSE_IF.txt	48%
26	Um pedaço dessa carne de porco <SSN> para mim e para o meu irmão			3,814	52935%	049%		049%	LSE_IF.txt	50%
27	de almas? Então, vocês são Paulo <SSN> e Barnabé, não são? Somos o			3,887	53736%	050%		050%	LSE_IF.txt	51%
28	encontrá-los! É que recebemos notícias <SSND3> da igreja de Jerusalém e			3,905	54023%	050%		050%	LSE_IF.txt	51%
29	estar atentos à presença dos traidores <SSND3> da nossa Lei. Se alguém			4,029	55636%	052%		052%	LSE_IF.txt	53%
30	Meu nome é Paulo, e sou discípulo <SSND3> do Senhor Jesus, o Cristo,			4,129	56533%	053%		053%	LSE_IF.txt	54%
31	, uma palavra que não tem dono <SSND3> na Terra, não tem chefe. Se			4,180	57578%	054%		054%	LSE_IF.txt	54%
32	permaneceremos em oração dia e noite <SSN> pelos nossos irmãos de			4,723	63952%	051%		051%	LSE_IF.txt	61%

FIGURA 7 - Pesquisa por <SSN em IF retornada pela ferramenta Concord

No Quadro 4 a seguir, apresentam-se as etiquetas da segmentação do sintagma verbal.

QUADRO 4
Etiquetas para o Sintagma Verbal

Sintagma Verbal	Etiqueta	Etiqueta (Início da Ocorrência)	Etiqueta (Final da Ocorrência)
Sintagma sem desvio	SSV	<SSV>	</SSV>
Separação de verbo e sintagma nominal ou adjetivo (objeto e complemento).	SSVD1	<SSVD1>	</SSVD1>
Separação de verbo auxiliar e verbo principal.	SSVD2	<SSVD2>	</SSVD2>
Separação verbo e advérbio	SSVD3	<SSVD3>	</SSVD3>

A Figura 8 a seguir apresenta uma imagem da tela da ferramenta Concord do programa *WordSmith Tools 5.0*© exibindo os resultados da pesquisa por <SSV que retorna todos os casos e tipos de segmentação de SV.

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	lead	lead	Sec	Sec	File	%
1	Só então eu vou poder encaminhar <SSVD1> o processo ao judiciário.			5,195	69934%				037%				LSE_IF.txt	67%
2	. Minha filha, o juiz não vai acreditar <SSV> que um homem com essa idade			5,235	70735%				037%				LSE_IF.txt	67%
3	com essa idade, vai conseguir controlar <SSVD1> um reincidente como o teu			5,244	70774%				038%				LSE_IF.txt	67%
4	Caminho, ótimo, mas que se convertam <SSVD3> antes às nossas Leis.			5,133	69132%				036%				LSE_IF.txt	66%
5	instrução, e mesmo assim consegue <SSV> com sua palavra converter			5,047	68333%				035%				LSE_IF.txt	65%
6	sua palavra converter muitos judeus <SSV> para o Caminho do Nazareno.			5,053	68333%				035%				LSE_IF.txt	65%
7	de Herodes. Se os gentios querem <SSVD2> seguir o Caminho, ótimo,			5,125	69124%				036%				LSE_IF.txt	66%
8	. Só o seu avô, ele não vai conseguir <SSVD1> a soltura do teu irmão. Mas			5,271	71254%				038%				LSE_IF.txt	68%
9	nasci. É porque você não tem rezado <SSV> para o verdadeiro Deus. Que			5,700	76334%				073%				LSE_IF.txt	73%
10	agitada] Milagrel Milagrel Você deve ser <SSVD1> um enviado dos deuses! Eu			5,766	78130%				074%				LSE_IF.txt	74%
11	Percebam! Por que vocês insistem <SSVD1> em permanecer nesta			5,801	78736%				075%				LSE_IF.txt	74%
12	desse vinho. Você não podia imaginar <SSV> que fosse assim tão duro, não			5,579	74942%				072%				LSE_IF.txt	71%
13	para ele. Eu acho até que vai ser <SSVD1> muito natural ele recusar.			5,350	71954%				039%				LSE_IF.txt	69%
14	não sei. Quando é que você vai receber <SSVD1> a sua irmã e perdô-la? Não			5,411	72638%				070%				LSE_IF.txt	69%
15	cidade. Bordoadas estão levando <SSVD1> meus pés nessa caminhada.			5,464	73330%				070%				LSE_IF.txt	70%
16	todos o amam, que sua obra já cruza <SSVD1> as fronteiras de Jerusalém.			5,008	67973%				035%				LSE_IF.txt	65%
17	agradar a todos. Aqueles que você viu <SSV> na sinagoga não me toleram,			4,405	60320%				037%				LSE_IF.txt	57%
18	esses é inadmissível que eu converta <SSVD1> alguém que não é judeu e			4,442	60638%				037%				LSE_IF.txt	57%
19	é judeu e que esse irmão possa não <SSVD2> ser subserviente à Lei			4,453	60631%				037%				LSE_IF.txt	58%
20	Menos mal, Paulo, os que o agrediram <SSV> na sinagoga, eles são judeus			4,313	59140%				036%				LSE_IF.txt	56%
21	de ação] A quem devo agradecer <SSV> por ter salvo minha vida?			4,250	58139%				035%				LSE_IF.txt	55%
22	Jesus. Claro. Por cujo nome corre-se <SSV> até mesmo risco de vida? Falar			4,282	58630%				035%				LSE_IF.txt	56%
23	risco de vida? Falar de Jesus se tornou <SSVD1> o meu propósito, a minha			4,292	58730%				035%				LSE_IF.txt	56%
24	. Para esses judeus, o que eu deveria <SSVD2> fazer além de celebrar			4,481	61132%				038%				LSE_IF.txt	58%
25	nas ruas, as crianças não encontram <SSVD3> mais leite no seio de suas			4,703	63843%				031%				LSE_IF.txt	61%
26	. E quando voltarmos, levaremos <SSVD1> a salvação dos nossos			4,763	64330%				031%				LSE_IF.txt	61%
27	. Minha vida já não me pertence e não <SSVD3> posso dividir algo que não			4,865	65830%				033%				LSE_IF.txt	63%
28	em Jerusalém. Meu irmão, conte-nos <SSV> o que se passa na Cidade			4,679	63636%				030%				LSE_IF.txt	60%
29	Jesus? Começaria por não poder <SSVD2> comer carne de porco ou			4,490	61233%				038%				LSE_IF.txt	58%
30	. Tudo que você precisa para aceitar <SSVD1> Jesus é abrir seu coração.			4,603	62638%				039%				LSE_IF.txt	59%
31	ajuda. Diz que não há o que comer <SSV> em Jerusalém. Meu irmão,			4,674	63530%				030%				LSE_IF.txt	60%
32	o verdadeiro Messias. Vocês ainda vão <SSVD2> conhecer Jesus Cristol			5,824	79057%				075%				LSE_IF.txt	74%

FIGURA 8- Pesquisa por <SSV em IF retornada pela ferramenta Concord.

No Quadro 5 abaixo, apresentam-se as etiquetas da segmentação do sintagma adjetivo.

QUADRO 5
Etiquetas para o Sintagma Adjetivo

Sintagma Adjetivo	Etiqueta	Etiqueta (Início da Ocorrência)	Etiqueta (Final da Ocorrência)
Sintagma sem desvio	SSAdj	<SSAdj>	</SSAdj>
Separação de Adjetivo e sintagma preposicional modificador.	SSAdjD1	<SSAdjD1>	</SSAdjD1>
Separação de advérbio modificador e adjetivo.	SSAdjD2	<SSAdjD2>	</SSAdjD2>

A Figura 9 abaixo apresenta uma imagem da tela da ferramenta Concord do programa *WordSmith Tools 5.0*© exibindo os resultados da

pesquisa por <SSAdj que retorna todos os casos e tipos de segmentação de SAdj.

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	lead	lead	Sec	Sec	File	%
1	ou livre. E se isso não está claro <SSAdj> para todos, se isto não está			6,397	86323%	032%			032%				LSE_IF.txt	82%
2	, uma pessoa que seja respeitada <SSAdj> pela comunidade, quem			5,309	71339%	038%			038%				LSE_IF.txt	68%
3	você. -Não. Aqui você será mais útil <SSAdjD1> à nossa causa. Junte todo			4,791	85037%	032%			032%				LSE_IF.txt	62%
4	amor é bondoso. Não é nada invejoso, <SSAdj> arrogante, orgulhoso. Jamais			7,725,00	71%	030%			030%				LSE_IF.txt	99%
5	de viver para Deus! Eu fui crucificado <SSAdj> junto com o Cristol Eu vivo,			6,706	90630%	036%			036%				LSE_IF.txt	86%
6	claro para todos, se isto não está claro <SSAdj> para Tiago, nosso irmão,			6,404	86345%	033%			033%				LSE_IF.txt	82%
7	o que restou deles. Como fico contente <SSAdj> em encontrá-los! É que			3,899	53937%	030%			030%				LSE_IF.txt	51%
8	. Ela o transformou em algo mais <SSAdjD2> corajoso, implacável e			2,682	39834%	035%			035%				LSE_IF.txt	36%
9	em mim. Eu que julgava estar sendo fiel <SSAdj> aos desígnios do nosso Deus,			2,574	39320%	033%			033%				LSE_IF.txt	35%
10	, não. Esses meninos aí são piores <SSAdj> que muita gente grande,			799	16835%	010%			010%				LSE_IF.txt	11%
11	homens. Estamos todos muito felizes <SSAdj> com a presença de Paulo			3,687	51624%	048%			048%				LSE_IF.txt	48%
12	Se o senhor escolheu o mais feroz <SSAdjD1> de nossos inimigos, temos			2,964	43236%	038%			038%				LSE_IF.txt	40%
13	o menor dos apóstolos. Nem sou digno <SSAdjD1> de ser chamado de			2,801	40827%	036%			036%				LSE_IF.txt	38%
14	ou livre. E se isso não está claro <SSAdj> para todos, se isto não está			5,751	91423%	032%			032%				LO_IF.txt	81%
15	, uma pessoa que seja respeitada <SSAdj> pela comunidade, quem			4,740	75239%	037%			037%				LO_IF.txt	67%
16	você. -Não. Aqui você será mais útil <SSAdjD1> à nossa causa. Junte todo			4,268	68737%	031%			031%				LO_IF.txt	60%
17	amor é bondoso. Não é nada invejoso, <SSAdj> arrogante, orgulhoso. Jamais			7,011,06	71%	030%			030%				LO_IF.txt	99%
18	de viver para Deus! Eu fui crucificado <SSAdj> junto com o Cristol Eu vivo,			6,051	95830%	036%			036%				LO_IF.txt	86%
19	claro para todos, se isto não está claro <SSAdj> para Tiago, nosso irmão,			5,758	91445%	032%			032%				LO_IF.txt	81%
20	o que restou deles. Como fico contente <SSAdj> em encontrá-los! É que			3,395	57337%	048%			048%				LO_IF.txt	48%
21	. ela o transformou em algo mais <SSAdjD2> corajoso, implacável e			2,266	42730%	032%			032%				LO_IF.txt	33%
22	em mim. Eu que julgava estar sendo fiel <SSAdj> aos desígnios do nosso Deus,			2,158	42320%	031%			031%				LO_IF.txt	31%
23	, não. Esses meninos aí são piores <SSAdj> que muita gente grande,			673	17635%	010%			010%				LO_IF.txt	10%
24	homens. Estamos todos muito felizes <SSAdj> com a presença de Paulo aqui			3,198	54824%	045%			045%				LO_IF.txt	46%
25	Se o senhor escolheu o mais feroz <SSAdjD1> de nossos inimigos, temos			2,504	46336%	036%			036%				LO_IF.txt	36%
26	o menor dos apóstolos. Nem sou digno <SSAdjD1> de ser chamado de			2,362	43827%	034%			034%				LO_IF.txt	34%

FIGURA 9 - Pesquisa por <SSAdj retornada pela ferramenta Concord.

No Quadro 6 abaixo, apresentam-se as etiquetas para a segmentação do sintagma preposicional.

QUADRO 6
Etiquetas para o Sintagma Preposicional

Sintagma Preposicional	Etiqueta	Etiqueta (Início da Ocorrência)	Etiqueta (Final da Ocorrência)
Sintagma sem desvio	SSPrep	<SSPrep>	</SSPrep>
Separação de locução prepositiva.	SSPrepD1	<SSPrepD1>	</SSPrepD1>
Separação de preposição e substantivo	SSPrepD2	<SSPrepD2	</SSPrepD2

A Figura 10 abaixo apresenta uma imagem da tela da ferramenta Concord do programa *Wordsmith Tools 5.0*© exibindo os resultados da pesquisa por <SSPrep que retorna todos os casos e tipos de segmentação de SPrep.

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	lead	lead	Sec	Sec	File	%
1	Somos testemunhas do poder de <SSPrepD2> Cristo. O que mais			6,506	87836%		034%				034%		LSE_IF.txt	84%
2	. A gente podia se sentir livre em cima <SSPrepD1> de um jumentinho, não é,			5,509	73930%		071%				071%		LSE_IF.txt	70%
3	um deles, Tito. Mas me converti depois <SSPrepD1> de meu encontro com o			4,349	59645%		056%				056%		LSE_IF.txt	56%
4	parte dos ladrões, corri perigo por parte <SSPrepD1> de meus irmãos de			6,631	89738%		035%				035%		LSE_IF.txt	85%
5	beleza e todos possam viver em paz <SSPrep> com suas diferenças.			7,636	99044%		038%				038%		LSE_IF.txt	98%
6	do seu trabalho e de sua fé em nome <SSPrepD1> de Nosso Senhor Jesus			7,106	93742%		032%				032%		LSE_IF.txt	91%
7	e desertos. Este homem que junto <SSPrepD1> a nós levou a Palavra do			7,049	93517%		031%				031%		LSE_IF.txt	90%
8	. E como judeus, estamos aqui na <SSPrepD2> sinagoga para repartir			4,100	56232%		053%				053%		LSE_IF.txt	53%
9	. É difícil de acreditar, mas vim com <SSPrepD2> a missão de levá-lo à			3,308	47830%		043%				043%		LSE_IF.txt	44%
10	. Na verdade, a morada Dele é dentro <SSPrepD1> de nós, não nas pedras			3,141	45830%		040%				040%		LSE_IF.txt	42%
11	, uns drogados! Mariana, cuidado com <SSPrepD2> o que você fala. Não seja			1,339	24930%		017%				017%		LSE_IF.txt	19%
12	. E devo confessar que estar aqui com <SSPrepD2> vocês é a segunda			3,432	48937%		044%				044%		LSE_IF.txt	45%
13	pastores. Minha missão é correr atrás <SSPrepD1> de outras ovelhas. Paulo,			3,751	52137%		048%				048%		LSE_IF.txt	49%
14	de vocês. Meu caminho é para além <SSPrepD1> das fronteiras de			3,732	51930%		048%				048%		LSE_IF.txt	49%
15	. Há problema em levar a palavra de <SSPrepD2> Cristo a algum homem na			3,571	50437%		046%				046%		LSE_IF.txt	47%
16	Somos testemunhas do poder de <SSPrepD2> cristo. O que mais			5,854	92936%		033%				033%		LO_IF.txt	83%
17	. A gente podia se sentir livre em cima <SSPrepD1> de um jumentinho, não é,			4,930	77930%		070%				070%		LO_IF.txt	69%
18	um deles, Tito. Mas me converti depois <SSPrepD1> de meu encontro com o			3,833	63145%		054%				054%		LO_IF.txt	54%
19	parte dos ladrões, corri perigo por parte <SSPrepD1> de meus irmãos de			5,976	94938%		035%				035%		LO_IF.txt	85%
20	beleza E todos possam viver em paz <SSPrep> com suas diferenças. O			6,933	04336%		038%				038%		LO_IF.txt	98%
21	do seu trabalho e de sua fé, em nome <SSPrepD1> de Nosso Senhor Jesus			6,452	98933%		032%				032%		LO_IF.txt	91%
22	e desertos. Este homem que junto <SSPrepD1> a nós levou a Palavra do			6,395	98717%		031%				031%		LO_IF.txt	90%
23	. É difícil de acreditar, mas vim com <SSPrepD2> a missão de levá-lo à			2,843	50930%		040%				040%		LO_IF.txt	41%
24	. Na verdade, a morada Dele é dentro <SSPrepD1> de nós, não nas pedras			2,676	48930%		038%				038%		LO_IF.txt	39%
25	, uns drogados! Mariana, cuidado com <SSPrepD2> o que você fala. Não seja			1,084	26130%		015%				015%		LO_IF.txt	16%
26	. E devo confessar que estar aqui com <SSPrepD2> vocês é a segunda			2,946	52037%		042%				042%		LO_IF.txt	42%
27	. E como judeus, estamos aqui na <SSPrepD2> sinagoga para repartir			3,590	59732%		051%				051%		LO_IF.txt	51%
28	pastores. Minha missão é correr atrás <SSPrepD1> de outras ovelhas. Paulo,			3,262	55337%		046%				046%		LO_IF.txt	46%
29	de vocês. Meu caminho é para além <SSPrepD1> das fronteiras de			3,243	55130%		046%				046%		LO_IF.txt	46%

FIGURA 10 - Pesquisa por <SSPrep retornada pela ferramenta Concord

Os procedimentos para análise qualitativa estão detalhados na seção seguinte.

2.3.3) ANÁLISE QUALITATIVA

A análise qualitativa dos dados será feita a partir dos dados quantitativos obtidos com a segmentação, conforme etiquetada segundo explicação anterior. Conforme já identificado na literatura da área, as razões de quebra de linha podem ser (i) geometria (linha superior e inferior do mesmo tamanho); (ii) preferência por linha superior mais curta; (iii) preferência por legenda verbatim; (iv) visual (pelos cortes e movimentos da câmera); (v) gramatical (pela sintaxe); e (vi) retórica (pelo ritmo das falas e pausas). Devido

a essa variedade de critérios, pesquisas anteriores já identificaram que a segmentação não é uma característica muito objetiva de se estudar (KALANTZI, 2008; FEITOSA, 2009;) e por isso não geram resultados quantitativos mensuráveis. Aparentemente há padrões de segmentação seguidos pelos legendistas, todavia, vários critérios de quebra parecem coocorrer em cada caso, dificultando a classificação dos tipos de segmentação, mas provendo dados substanciais para um estudo qualitativo. Sendo assim, é imprescindível fazer uma análise cuidadosa de caso por caso, visto que este é um aspecto que também influencia fortemente a leitura e recepção da legendagem. À vista disso, esta pesquisa é motivada por duas perguntas:

- i) Que razão de segmentação (visual, retórica, estética, legenda verbatim) teve precedência nas legendas de duas linhas com desvio gramatical do filme *Irmãos de Fé*?
- ii) A segmentação da LSE apresenta mais desvios que a da LO devido às informações adicionais que devem ser transmitidas?

Para essa investigação, cada ocorrência de segmentação com desvio dentro da mesma legenda será descrita a partir das regras gerais de constituição de sintagmas de Lemle (1989). Todos os casos de segmentação foram etiquetados e quantificados, mas apenas na fase de análise qualitativa serão descritos e qualificados segundo categorias e funções gramaticais.

Além disso, buscou-se nos tipos de segmentação - geometria; linha superior mais curta; legenda verbatim; visual; gramatical; e retórica - uma justificativa, ou um padrão para as ocorrências de segmentação com desvio e verificou-se a possibilidade de sugerir-se uma segmentação sem desvio ou, ao menos, com desvio menor de maneira a facilitar a leitura da legenda.

O próximo capítulo será dedicado à análise quantitativa e qualitativa dos dados e posterior discussão dos resultados encontrados à luz das hipóteses formuladas e perguntas de pesquisa levantadas.

CAPÍTULO 3

RESULTADOS E DISCUSSÃO

Após a compilação, tratamento e etiquetação do corpus, passou-se à análise dos dados.

3.1) RESULTADOS GERAIS DE IF FORNECIDOS PELO WORDSMITH TOOLS E PELO SUBTITLE WORKSHOP

O tratamento do corpus pelo software *Wordsmith Tools*© 5.0 forneceu dados que apresentam um panorama do texto e dos números das legendas estudadas nesta pesquisa. A Figura 11 abaixo mostra a tela da ferramenta *WordList* com a lista das palavras mais frequentes do corpus.

N	Word	Freq	%	Texts	%_lemmas	Set
1	QUE	557	3.75	2	100.00	
2	DE	531	3.57	2	100.00	
3	O	384	2.58	2	100.00	
4	A	375	2.52	2	100.00	
5	NÃO	372	2.50	2	100.00	
6	É	286	1.92	2	100.00	
7	EU	280	1.88	2	100.00	
8	E	274	1.84	2	100.00	
9	UM	162	1.09	2	100.00	
10	VOCÊ	156	1.05	2	100.00	
11	SE	144	0.97	2	100.00	
12	MÚSICA	142	0.95	1	50.00	
13	ME	134	0.90	2	100.00	
14	DEUS	132	0.89	2	100.00	
15	PARA	132	0.89	2	100.00	
16	PAULO	121	0.81	2	100.00	
17	COM	117	0.79	2	100.00	
18	ELE	116	0.78	2	100.00	
19	MEU	110	0.74	2	100.00	
20	EM	98	0.66	2	100.00	
21	MAS	98	0.66	2	100.00	
22	DO	95	0.64	2	100.00	
23	JESUS	94	0.63	2	100.00	
24	POR	94	0.63	2	100.00	
25	OS	92	0.62	2	100.00	
26	SENHOR	90	0.61	2	100.00	
27	UMA	84	0.56	2	100.00	
28	HOMEM	73	0.49	2	100.00	
29	ESTÁ	70	0.47	2	100.00	
30	NÓS	66	0.44	2	100.00	
31	VAI	66	0.44	2	100.00	
32	MAIS	65	0.44	2	100.00	
33	ESPINHO	64	0.43	2	100.00	

FIGURA 11 - Lista de palavras mais frequentes no corpus

Conforme esperado, as palavras gramaticais e os nomes próprios estão entre as mais frequentes do corpus. Entretanto, ao contrário do que se espera, por serem textos de legendas que em geral passam por muita edição e condensação; os pronomes (principalmente de primeira pessoa) - itens considerados redundantes, portanto, dispensáveis - também apareceram como mais frequentes.

Já a Figura 12 a seguir mostra a tela da ferramenta *WordList* contendo os dados estatísticos do corpus dessa pesquisa.

	N	Overall	1	2
text file		Overall	LO_IF.txt	LSE_IF.txt
file size		140,162	65,184	74,978
tokens (running words) in text		14,870	7,072	7,798
tokens used for word list		14,864	7,069	7,795
sum of entries				
types (distinct words)		1,682	1,601	1,680
type/token ratio (TTR)		11.32	22.65	21.55
standardised TTR		40.86	41.39	40.34
standardised TTR std.dev.		54.10	51.07	52.04
standardised TTR basis		1,000	1,000	1,000
mean word length (in characters)		4.33	4.27	4.39
word length std.dev.		2.31	2.30	2.33
sentences		2,070	1,065	1,005
mean (in words)		7.18	6.64	7.76
std.dev.		6.44	5.74	7.07
paragraphs		2	1	1
mean (in words)		7,432.00	7,069.00	7,795.00
std.dev.		513.36		
headings				
mean (in words)				
std.dev.				
sections		2	1	1
mean (in words)		7,432.00	7,069.00	7,795.00
std.dev.		513.36		
numbers removed		6	3	3
stoplist tokens removed				
stoplist types removed				
1-letter words		1,361	673	688
2-letter words		2,187	1,044	1,143
3-letter words		2,801	1,379	1,422
4-letter words		1,891	933	958
5-letter words		2,577	1,235	1,342
6-letter words		1,455	696	699

FIGURA 12 - Dados estatísticos do corpus IF

Coerentemente com informações fornecidas pelo *Subtitle Workshop*®, que apontam a que a LSE de IF tem aproximadamente 300 linhas de legenda a mais que a LO de IF; os dados sobre as legendas de IF, obtidos com o *Wordsmith Tools*® 5.0, revelam que a LSE é maior que a LO também em número de *tokens* e *types*. No entanto, imprevistamente, a razão *type/token* padronizada da LSE de IF é ligeiramente menor que a da LO de IF, indicando que a frequência de palavras repetidas é maior e que o texto dessa legenda seria mais redundante. Uma das possíveis explicações para essa

característica, a priori, é a apresentação de elementos não linguísticos, e.g. identificação de falantes, efeitos sonoros, sons ambiente, etc.

A Tabela 1 abaixo contém os dados estatísticos extraídos com o *Wordsmith Tools 5.0*© do corpus de legendas.

TABELA 1
Dados quantitativos obtidos com o *Wordsmith Tools*

	TOTAIS PARA O CORPUS	LO IF	LSE IF
TAMANHO DO ARQUIVO	140346	65278	75068
TOKENS	14872	7073	7799
TYPES	1682	1601	1680
RAZÃO TYPE/TOKEN	11,32	22,64	21,55
RAZÃO TYPE/TOKEN PADRONIZADA	40,86	41,40	40,33
NÚMERO DE SENTENÇAS	2085	1081	1004
TAMANHO MÉDIO DAS SENTENÇAS	7,13	6,54	7,76

As informações sobre *types* e *tokens* da Tabela 1 acima estão rerepresentadas em valores absolutos no Gráfico 1 abaixo, a fim de facilitar a visualização.

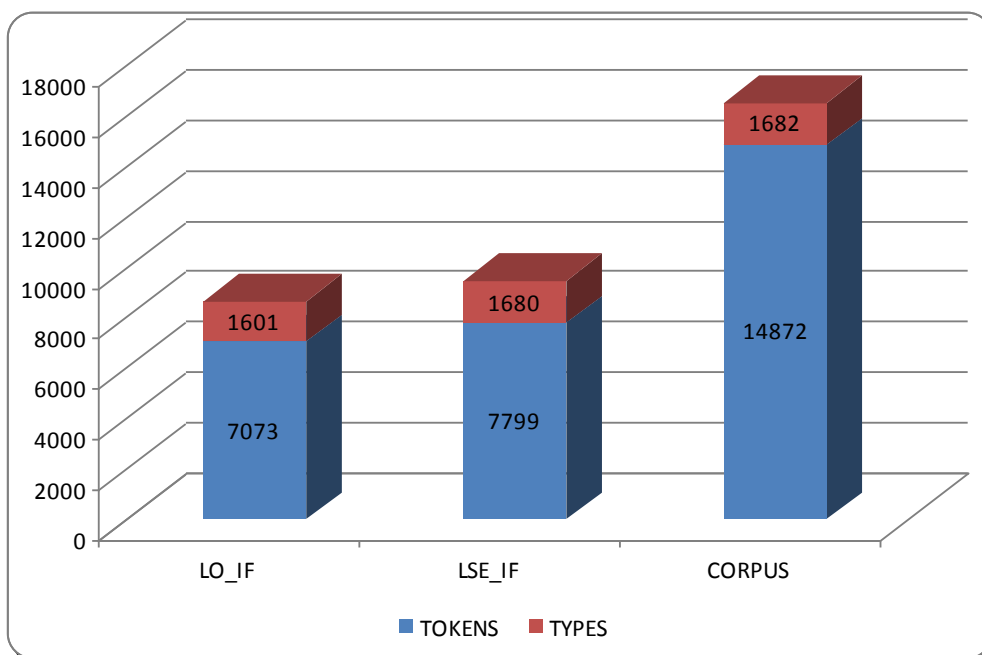


GRÁFICO 1 – Valores absolutos de *types* e *tokens* distribuídos em IF.

Como mencionado anteriormente, os números encontrados nos campos de *types* e *tokens* corroboram as informações obtidas com o *Subtitle Workshop*© 2.51 sobre os subcorpora que compõem o presente corpus. Assim como o número de *types* e *tokens* é menor para a LO de IF, também é menor o número de legendas (1107) em LO_IF, quando comparado com a LSE do mesmo filme (1301) como pode ser melhor visualizado no Gráfico 2 a seguir.

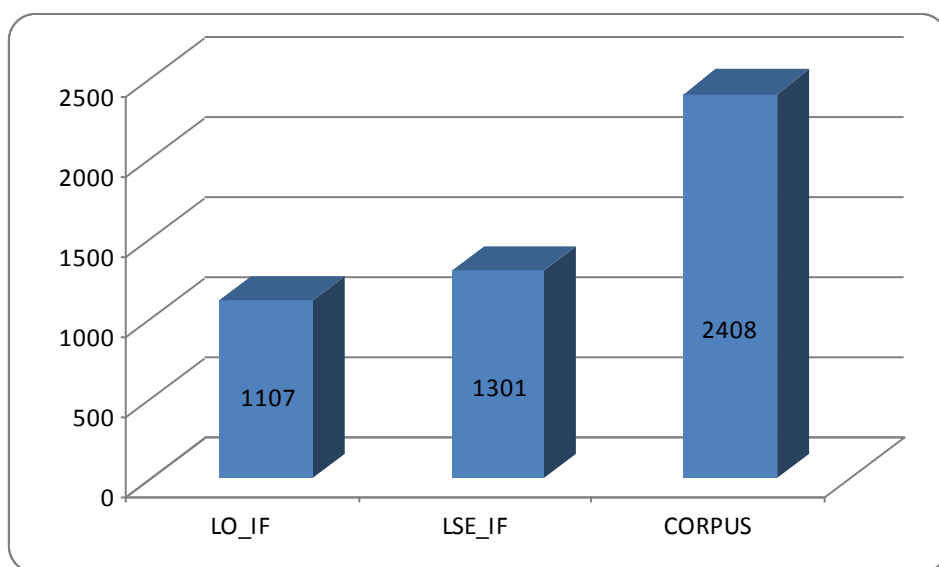


GRÁFICO 2 – Valores absolutos de linhas de legenda em IF.

Os números absolutos de *types* e *tokens* e os números absolutos de linhas de legenda revelam um mesmo padrão de distribuição dos textos no corpus, isto é, LO-IF < LSE-IF. A análise desses números de uma perspectiva percentual reforça esse padrão. Para facilitar a visualização, os valores percentuais estão apresentados nos Gráfico 3 e 4 a seguir exibindo a distribuição dos textos no corpus a partir do número de *tokens* e do número de linhas de legenda, respectivamente.

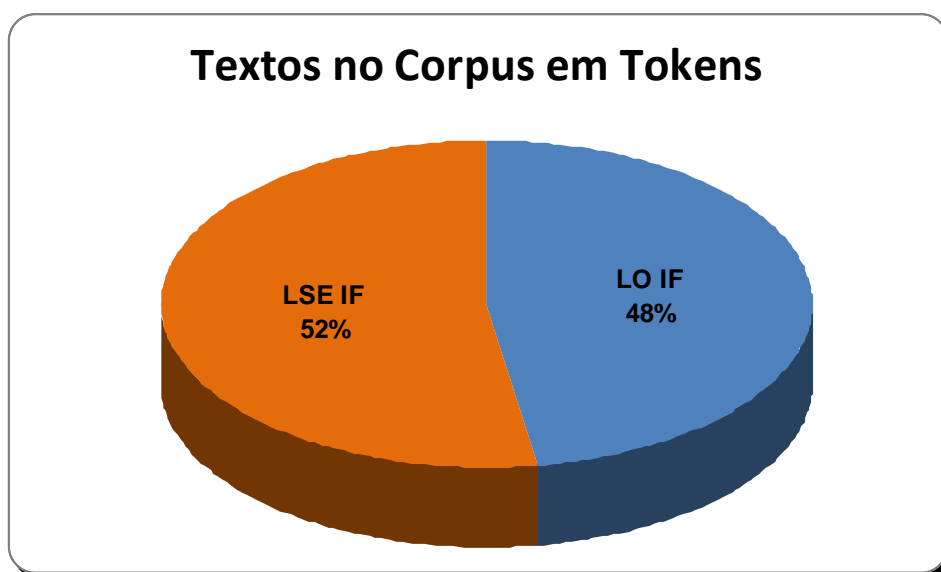


GRÁFICO 3 - Distribuição dos textos no corpus em número de *tokens*

O gráfico acima mostra que a LO de IF representa 48% dos *tokens* totais do corpus; enquanto que a LSE de IF representa 52% dos *tokens* totais. Essa diferença de quatro pontos percentuais duplica quando se trata do número de legendas nos subcorpora, como mostra o Gráfico 4 a seguir. A LO_IF está representada em 46% do total de linhas de legenda no corpus e a LSE_IF está em 54% do total de linhas de legenda desse corpus.

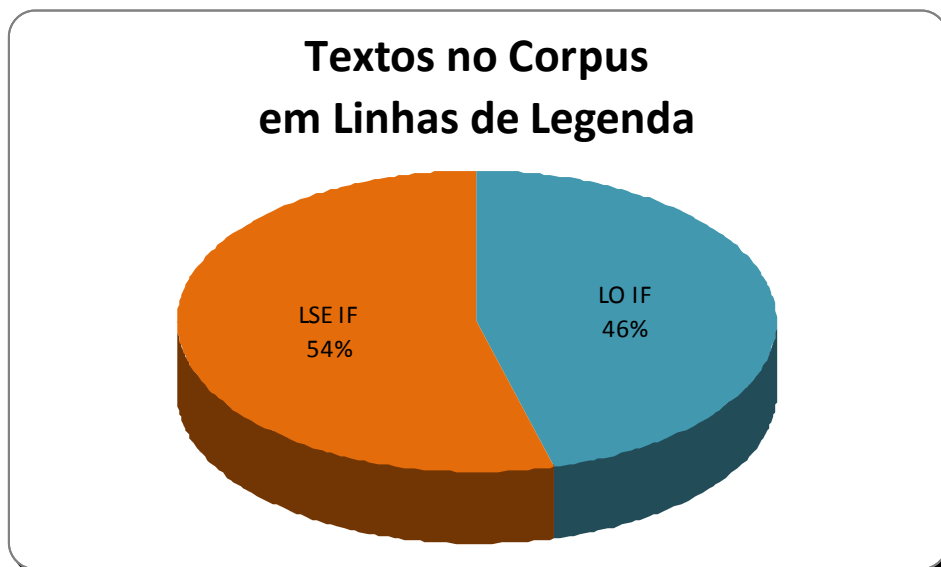


GRÁFICO 4 - Distribuição de número de linhas de legenda em IF

Esses dados alimentam uma possível análise primária a respeito de características das legendas estudadas. Ainda comparando legendas interlinguais e intralinguais, o estudo de De Linde e Kay (1999) já apontava características fundamentais da modalidade que atualmente estão presentes nas denominadas LO e LSE: a velocidade de leitura e a representação de elementos da trilha sonora.

O estudo já esclarecia que as interlinguais (hoje LOs) são legendas mais rápidas. Isso implica em textos de legenda menos reduzidos (em relação ao áudio original) comparados a textos de LSEs, os quais deveriam ser mais enxutos a fim de se ajustar à velocidade de leitura, adequada ao público SE. Díaz-Cintas e Remael (2007) trata desse aspecto em termos de palavras por minuto, isto é, quanto mais palavras por minuto, mais rápida é uma legenda; e, portanto, quanto menos palavras por minuto, mais lenta é a legenda.

Além disso, De Linde e Kay (1999) também já ressaltava a importância da LSE exibir na tela informações além daquelas contidas nos diálogos. Nesse tipo de legenda, há a necessidade de apresentar-se para o espectador SE em formato gráfico, na tela, as informações relevantes contidas na trilha sonora (identificação de falantes, sons e ruídos significativos para o conteúdo da mensagem, identificação de música).

Esses dois aspectos mencionados influenciam e estão diretamente relacionados ao tamanho de textos de legenda. Conforme citado anteriormente,

achados iniciais da pesquisa de Araújo e Nascimento (2011) revelam que o grau de redução e uma boa segmentação foram determinantes – mais que a velocidade de leitura - para a boa recepção de legendas, ou seja, ainda que rápidas, legendas contendo menos texto e melhor organizadas promovem a satisfação do público SE. Segundo Díaz-Cintas e Remael (2007), as estratégias de redução envolvem, por exemplo, simplificação de perífrases verbais; uso de generalizações; uso de sinônimos ou palavras equivalentes mais curtos; transformação de discurso direto em indireto; uso de sentenças simples ao invés de complexas; mescla de frases/sentenças; e omissões. Uma consequência imediata da aplicação dessas estratégias na legendagem é a diminuição da quantidade de palavras no texto da legenda. Como sugere Araújo e Nascimento (2011), portanto, é importante que a LSE seja mais reduzida que a LO e consequentemente, contenha menos palavras. Os primeiros dados apresentados do caso sob escrutínio neste trabalho, entretanto, revelam que a LSE de IF contém mais palavras (*tokens*) que a LO de IF. São 726 palavras a mais em LSE_IF em comparação com LO_IF. Pode-se argumentar que, devido ao tipo de legendagem – uma LSE –, essa diferença no número de palavras não esteja relacionada ao grau de redução de texto, mas sim ao acréscimo de informações adicionais característico das LSEs e, logo, a LSE_IF contém mais *tokens* que a LO_IF devido a esses acréscimos. No entanto, com o uso da ferramenta *Information and Errors* do programa *Subtitle Workshop*® é possível visualizar e remover todas as linhas de legenda marcadas como “Hearing Impaired”; as linhas de LSE_IF que contém apenas informações adicionais (não acompanhadas de diálogos) foram, então, eliminadas para a contabilização do número de palavras (*tokens*) que seriam também excluídas: 538 palavras foram eliminadas, as quais estavam distribuídas em 191 linhas de legenda (quase o total da diferença entre LO_IF e LSE_IF que é de 194 linhas de legenda). Em seguida, outras 81 linhas de legenda que continham diálogo mais informação extralinguística tiveram suas informações adicionais removidas para a aferição do total de palavras presentes em LSE_IF. O número resultante é de 7081 *tokens* ao todo na LSE_IF; são ainda 08 *tokens* a mais que a LO_IF. Esses dados sugerem, assim, que o texto da LSE_IF, ao contrário do esperado, foi menos reduzido que o da LO_IF.

Além disso, o *Subtitle Workshop*® revela também que, na maioria dos casos, mesmo com a adição de elementos extras à linha de legenda, o tempo de duração da legenda na tela permanece o mesmo na LO_IF e na LSE_IF, como mostrado no exemplo da Figura 13 a seguir.

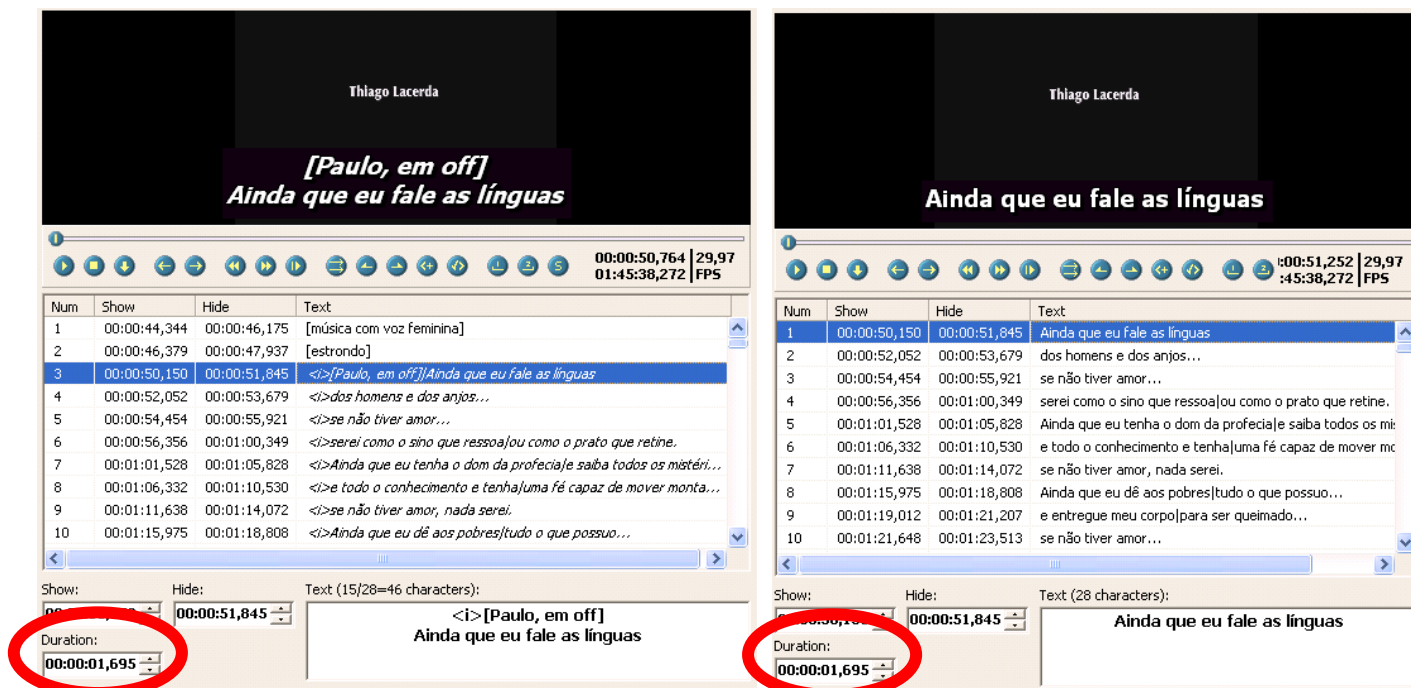


FIGURA 13 – Mesmo tempo de duração de legenda em LSE_IF e em LO_IF

A figura traz uma imagem parcial de duas telas do *Subtitle Workshop*®: do lado esquerdo há uma linha da LSE_IF com 46 caracteres e do lado direito apresenta-se a mesma linha da LO_IF com 28 caracteres. Os círculos vermelhos localizados na parte inferior da imagem destacam que o tempo de duração das duas linhas é igual, apesar da diferença na quantidade de caracteres.

Conforme as tabelas de equivalência entre segundos/frames e caracteres elaboradas em Díaz-Cintas e Remael (2007), essa linha de legenda em LO_IF tem uma velocidade de leitura de 160 palavras por minuto, enquanto a linha de legenda em LSE_IF tem uma velocidade de leitura de 180 palavras por minuto. Essa ilustração é apenas uma entre as outras várias ocorrências da mesma prática. São 81 linhas de legenda que contêm diálogo mais informações adicionais; em 18 desses casos (aproximadamente 22%) o tempo de duração da legenda foi alterado na LSE_IF para acomodar o aumento do

número de caracteres; porém nos outros 63 casos não houve alteração no tempo.

Todavia, essa é apenas uma análise panorâmica desses dados. Os números apresentados acima fornecidos pelo *Wordsmith Tools*® e pelo *Subtitle Workshop*® renderiam uma análise minuciosa de parâmetros técnicos e textuais da legendagem estudada que foge ao escopo desta pesquisa. Entretanto, essa primeira análise já sugere que a LO e a LSE de IF receberam praticamente o mesmo tratamento. Isto quer dizer que, com exceção da inserção de elementos extralinguísticos, o legendista não observou outros aspectos - tais como redução de texto e velocidade de legenda - que caracterizam e distinguem as duas modalidades de legendagem.

Outra informação relevante fornecida pelo *Wordsmith Tools*® é a razão *type/token*. Segundo Baker (2000), a razão *type/token* mede a variedade lexical de um texto, indicando a proporção de palavras diversificadas em relação ao número total de palavras de um texto. Sendo assim, quanto maior a razão *type/token*, maior a variedade lexical, portanto menos repetições serão encontradas nesse texto. Entretanto, Baker (2000) adverte que quanto maior o texto, maiores são as chances de uso de vocábulos repetidos, e assim reforça que a razão *type/token* padronizada, cujo cálculo leva em conta o tamanho do texto, é mais confiável.

Corroborando a preocupação de Baker (2000) quanto à influência do tamanho do corpus no cálculo da razão *type/token*, os números desta pesquisa apontam que o maior texto do corpus, LSE_IF, é menos variado lexicalmente (razão *type/token* de 21,55) que o menor texto, LO_IF (razão *type/token* de 22,64). Eliminando-se, porém, a variável do tamanho do texto, o padrão se repete no cálculo da razão *type/token* padronizada, em que a variedade lexical em LSE_IF (40,33) é ainda ligeiramente menor que em LO_IF (41,40). Para facilitar a visualização, o Gráfico 5 a seguir reapresenta os valores da Tabela 1 reproduzida anteriormente, com relação à comparação da razão *type/token* padronizada verificada para os subcorpora analisados e para o corpus como um todo.

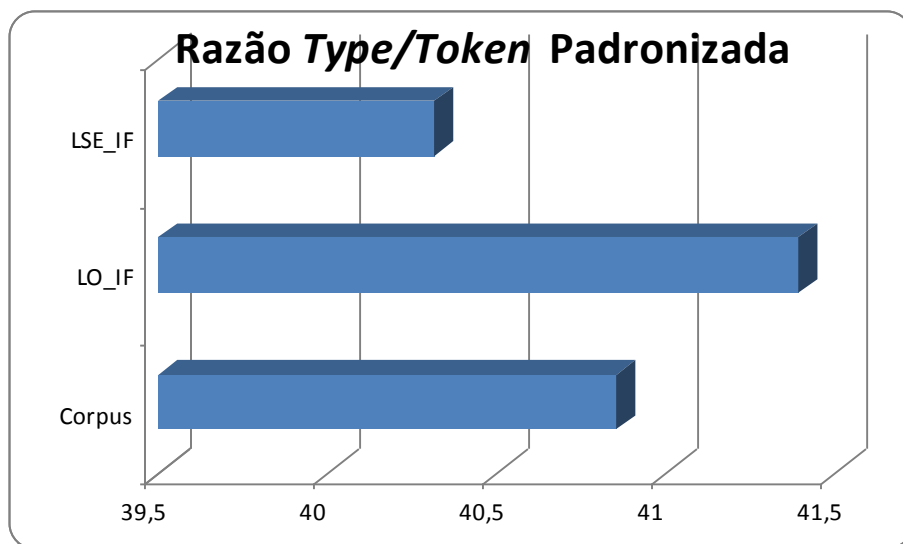


GRÁFICO 5 - Razão *type/token* na LO_IF, na LSE_IF e em IF

Esses dados sugerem que a frequência de palavras repetidas é maior na LSE de IF e que o texto dessa legenda seria mais redundante. Uma possível explicação para essa característica é a necessidade de apresentação de elementos extralinguísticos, e.g. identificação de falantes, efeitos sonoros, sons ambiente, etc. Por exemplo, só a palavra “música” que não aparece nenhuma vez no texto de LO_IF, ocorre 142 vezes no texto de LSE_IF, todas elas para indicar ao espectador SE a informação contida na trilha sonora do filme. A Figura 14 abaixo traz uma lista de frequência de palavras da LSE_IF comparada à frequência dessas mesmas palavras em LO_IF - gerada pela ferramenta *Wordlist* do programa *Wordsmith Tools*® - que mostra as palavras repetidas na LSE causadoras da diferença de variedade lexical entre os subcorpora. As cinco palavras (5 *types*) que constam da lista aparecem entre as 75 palavras mais frequentes na LSE e correspondem a um total de 269 *tokens*, os quais representam aproximadamente 3,5% do total de *tokens* em LSE_IF.

N	Key word	Freq.	%	freq. in LO_IF.lst	%	Keyness	P_lemma	Set
1	MÚSICA	142	1.82	0		184.56	0.0000000000	
2	CONTINUA	63	0.81	0		81.57	0.0000000000	
3	ESTRONDO	23	0.29	0		29.72	0.000000469	
4	MULTIDÃO	22	0.28	0		28.43	0.000000942	
5	SUSPENSE	19	0.24	0		24.55	0.000007210	

FIGURA 14 – Lista de palavras comparada LSE_IF e LO_IF

Um dado inesperado revelado pelo *Wordsmith Tools*© é a diferença no número de sentenças. O Gráfico 6 a seguir facilita a visualização dos números apresentados anteriormente na Tabela 1.

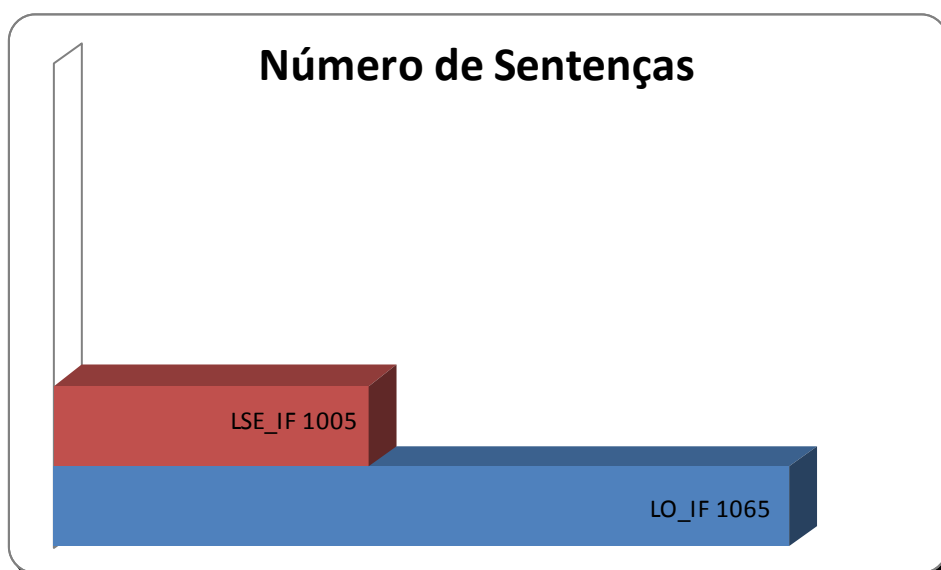


GRÁFICO 6 – Número de sentenças nos subcorpora

Essa diferença de 65 sentenças a mais no texto de LO_IF não era esperada, principalmente, devido aos números de linhas de legenda em cada subcorpus. Como mostrado anteriormente, LSE_IF possui 194 linhas de legenda a mais que LO_IF. A partir dessa diferença, seria mais plausível pressupor que o número de sentenças na LSE fosse maior que na LO. Os dados quanto ao tamanho médio das sentenças, aparentemente, ajudam a esclarecer a quantidade de sentenças. O abaixo reapresenta os valores da Tabela 1 reproduzida acima quanto ao tamanho médio das sentenças em cada texto do corpus.

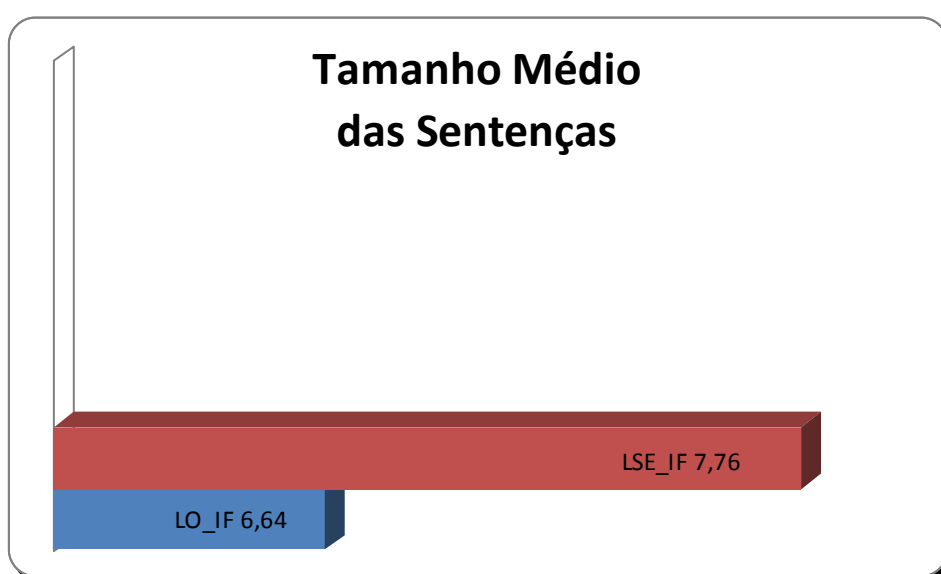


GRÁFICO 7 – Tamanho médio das sentenças nos subcorpora

A contagem do *Wordsmith Tools*© revela que a LO_IF contém mais sentenças, porém mais curtas, enquanto na LSE_IF as sentenças são maiores e mais espaçadas. Isso se dá porque a ferramenta do *Wordsmith Tools*© inclui as informações adicionais como parte das sentenças, o que amplia a quantidade de palavras nas sentenças de LSE_IF, fazendo-as maiores e em menor número que na LO_IF. Esses valores sugerem que há diferenças não apenas nos critérios de pontuação, como também de segmentação das legendas. Informações fornecidas pelo *Subtitle Workshop*© 2.51 evidenciam isso, pois revelam que a LO_IF é formada por 501 legendas de uma linha e 606 legendas de duas linhas, enquanto a LSE de IF é composta por 608 de uma linha e 693 de duas linhas.

Os números de legendas mencionados anteriormente estão organizados na Tabela 2 a seguir que apresenta os dados da segmentação nos dois textos do corpus em valores absolutos.

TABELA 2
Distribuição de tipos de linha de legenda em números absolutos

	1 LINHA	2 LINHAS	TOTAL DE LEGENDAS
LO – IF	504	603	1107
LSE – IF	613	688	1301
CORPUS	1117	1291	2408

Para uma melhor visualização dos dados, o GRÁF. 8 a seguir, reinterpreta a distribuição, em números absolutos, dos tipos de linha de legenda nos subcorpora, conforme exposto na Tabela 2 acima.

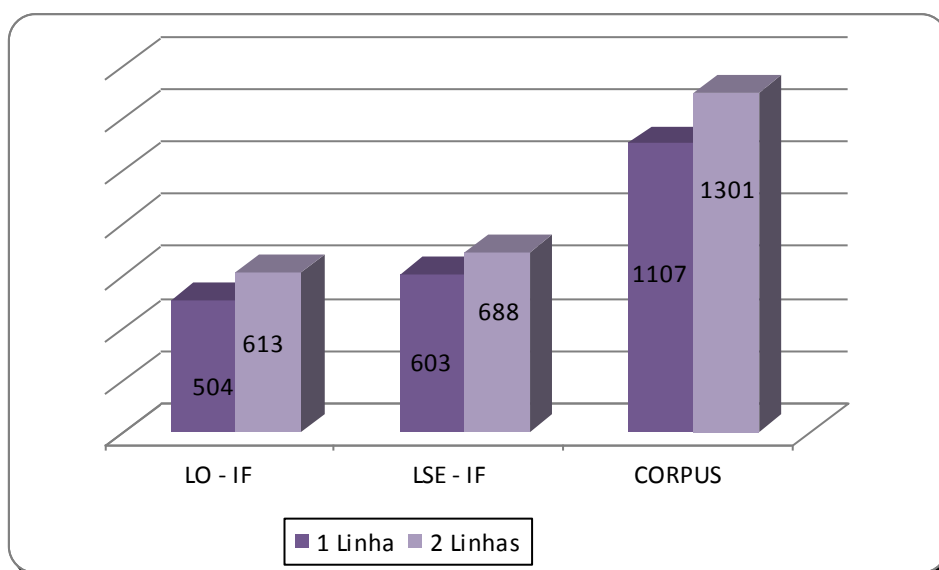


GRÁFICO 8 - Legendas de 1 e 2 linhas em números absolutos

As diferenças observadas na análise dos números absolutos ficaram também refletidas nos valores percentuais. A partir desses números fornecidos pelo *Subtitle Workshop*®, os gráficos a seguir permitem visualizar, em valor percentual, a distribuição da segmentação em uma e duas linhas de legenda para cada subcorpus e para o total do corpus.

Legendas de 1 e 2 Linhas em LO - IF

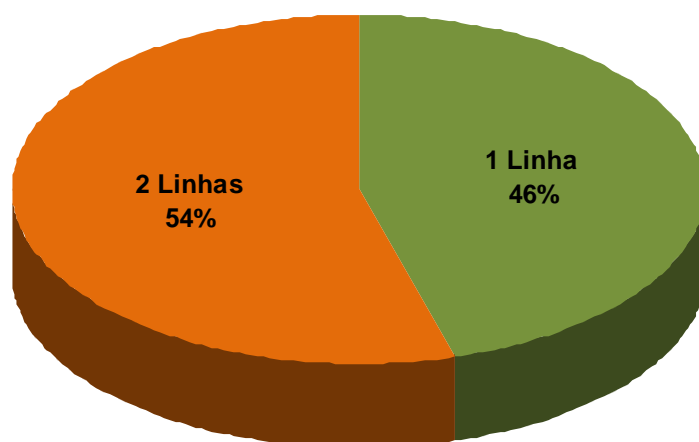


GRÁFICO 9 - Percentual de legendas de 1 e 2 linhas em LO_IF

Segundo o Gráfico 9 (acima) e o Gráfico 10 (abaixo) os dois tipos de legenda – LO e LSE – estão predominantemente divididos em duas linhas. Contudo, nota-se que a preferência por legendas de uma linha em LSE_IF (47%) foi ligeiramente maior que em LO_IF (46%), consequentemente, a preferência por legendas de duas linhas foi menor na LSE (53%) que na LO (54%).

Legendas de 1 e 2 Linhas em LSE - IF

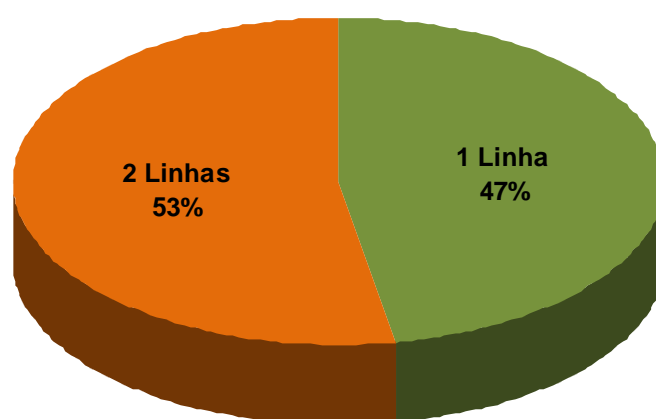


GRÁFICO 10 - Percentual de legendas de 1 e 2 linhas em LSE_IF

O padrão encontrado entre os subcorpora repete-se nos números gerais do corpus. O Gráfico 11 abaixo reafirma a predominância de legendas de duas linhas no corpus como um todo.

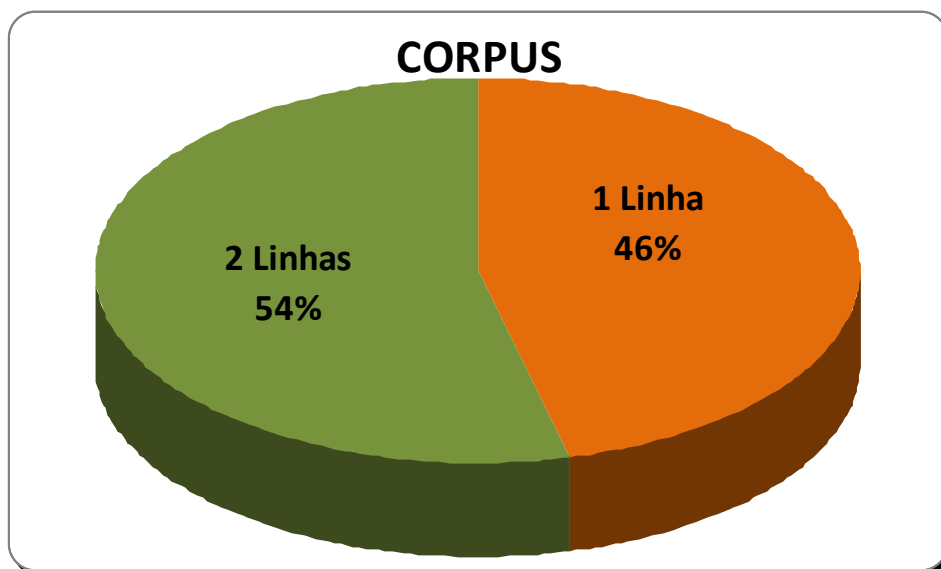


GRÁFICO 11 - Percentual de legendas de 1 e 2 linhas no corpus

Mais uma vez, essas diferenças podem ser atribuídas à presença das linhas de legenda que constam de informações adicionais na LSE_IF, as quais tendem a ser curtas e a ocuparem apenas uma linha de legenda, explicando, portanto, o maior percentual de legendas de uma linha na LSE.

O aspecto da segmentação, mais especificamente, da quebra de linhas no corpus IF é o foco principal desta pesquisa e será investigado na próxima subseção.

3.2) ANÁLISE DOS DADOS DA SEGMENTAÇÃO

A ferramenta *Concord* do *Wordsmith Tools*© 5.0 fornece informações ainda mais detalhadas sobre as características da segmentação – do tipo quebra de linha – das legendas do corpus IF apresentadas na próxima subseção. Conforme mencionado na seção de metodologia, todas as legendas divididas em duas linhas foram etiquetadas. A etiquetagem incidiu sobre segmentações de sentenças, orações e constituintes menores. Entretanto, para

a etiquetagem das quebras de linha seguiram-se critérios sintático-gramaticais mais detalhados, baseados em Lemle (1989) para rastrear a segmentação de itens menores que a oração, a saber, os sintagmas constituintes de oração. Dessa forma, para a segmentação dos cinco sintagmas – nominal, adjetivo, adverbial, verbal e preposicional – estabeleceram-se etiquetas que marcassem quando uma segmentação seguia os padrões gramaticais e outras etiquetas para quando a segmentação era diferente da estrutura gramatical prevista em Lemle (1989) (ver capítulo de metodologia).

A Figura 15 a seguir apresenta a imagem (parcial) da ferramenta *Concord*, na qual são mostrados os resultados da busca por ocorrências de legendas segmentadas (LS) na LSE de IF.

N	Concordance	Word #	ben	ben	Para	Para	lead	lead	Sec	Sec	File	%
1	Estou pronto, senhor. [cabras balindo] <LS> [confusão de vozes] Um pedaço	3,805	52926%	049%	049%						LSE_IF.txt	50%
2	a escola. [estrondo metálico] [passos] <LS> [estrondo] [estrondo metálico]	712	15037%	0 9%	0 9%						LSE_IF.txt	10%
3	recomeça] [Mariana soluça mais forte] <LS> [estrondo] [música triste]	823	17148%	011%	011%						LSE_IF.txt	11%
4	. É melhor. Calma! [música contínua] <LS> [estrondo] Um momento. Por favor	774	16430%	010%	010%						LSE_IF.txt	11%
5	vai subindo] Vamos escrever mais. <LS> [estrondo] -Você gostou da	7,532	97657%	037%	037%						LSE_IF.txt	97%
6	Pedrol Pedrol [som de chibatadas] <LS> [gritos de Pedro] [mais	4,927	67012%	034%	034%						LSE_IF.txt	64%
7	de perigo] [som das chibatadas] <LS> [gritos de Pedro] Tiago: Ester!	4,952	67035%	034%	034%						LSE_IF.txt	64%
8	-Espírito Espírito Santo Espírito Santo <LS> [multidão canta] Arrepial	220	3537%	0 3%	0 3%						LSE_IF.txt	3%
9	Espírito Santo [multidão canta] Arrepial <LS> [multidão canta] Espírito Espírito	223	3617%	0 3%	0 3%						LSE_IF.txt	3%
10	[multidão contínua cantando] Força! <LS> [multidão canta] Tá tudo certinho.	382	7217%	0 5%	0 5%						LSE_IF.txt	5%
11	Cantando: Espírito! Derrama a unção! <LS> [multidão contínua cantando]	362	69 8%	0 5%	0 5%						LSE_IF.txt	5%
12	, invocado! Espírito Santo de Deus! <LS> [multidão contínua cantando]	378	7120%	0 5%	0 5%						LSE_IF.txt	5%
13	irmão! Meu irmão! Não! Não! Estevão! <LS> [multidão contínua gritando]	934	18214%	012%	012%						LSE_IF.txt	13%
14	ação] Estamos pertol! Estamos pertol! <LS> [música contínua] [galopar dos	1,943	324 2%	025%	025%						LSE_IF.txt	26%
15	. [música de suspense] [passos] <LS> [música contínua] [galopar dos	1,718	30231%	022%	022%						LSE_IF.txt	23%
16	Mulher em off: Assassino! Assassino! <LS> [música contínua] [música	2,097	344 7%	027%	027%						LSE_IF.txt	29%
17	de suspense] [relinchar dos cavalos] <LS> [música contínua] [música vai	1,729	30257%	022%	022%						LSE_IF.txt	24%
18	[música recomeça] [guitarra e confusão] <LS> [música contínua] [música vai	2,847	41657%	037%	037%						LSE_IF.txt	38%
19	. E você não tardará a pagar o seu. <LS> [música contínua] [sons de soco]	5,934	80633%	076%	076%						LSE_IF.txt	76%
20	suspense] [ruído de madeira contínua] <LS> [música contínua] Escuta aqui	1,208	23335%	016%	016%						LSE_IF.txt	17%
21	de ação contínua] Criança: Pail! [gritos] <LS> [música contínua] Você ainda vai	1,170	22517%	015%	015%						LSE_IF.txt	16%
22	[gritos] Socorro! Meus filhos, não. <LS> [música de ação contínua]	1,163	22440%	015%	015%						LSE_IF.txt	16%
23	poderosa. Ele está ao nosso lado. <LS> [música de ação] [estrondo]	1,148	22150%	015%	015%						LSE_IF.txt	16%
24	contra ele? Infame! [confusão e gritos] <LS> [música de ação] A quem devo	4,243	58125%	035%	035%						LSE_IF.txt	55%
25	de Damasco. [galopar dos cavalos] <LS> [música de ação] Estamos pertol!	1,936	32230%	025%	025%						LSE_IF.txt	26%
26	formação religiosa. [batidas na porta] <LS> [música de fundo] Ele está lhe	2,284	36846%	029%	029%						LSE_IF.txt	31%
27	. Queres conhecê-Lo? Jesus? Quero. <LS> [música de fundo] Levanta.	5,724	77033%	074%	074%						LSE_IF.txt	73%
28	de Deus! Não! Ail! [guitarra e confusão] <LS> [música de mistério] Ah! [música	996	20043%	013%	013%						LSE_IF.txt	14%
29	[música de perigo] [mais chibatadas] <LS> [música de perigo] [gritos de	4,937	67041%	034%	034%						LSE_IF.txt	64%
30	[gritos de Pedro] [mais chibatadas] <LS> [música de perigo] [mais	4,932	67026%	034%	034%						LSE_IF.txt	64%
31	[música contínua] [galopar dos cavalos] <LS> [música de suspense] [relinchar	1,723	30243%	022%	022%						LSE_IF.txt	23%
32	Ô, mãe! Manhê! [ruído de madeira] <LS> [música de suspense] [ruído de	1,201	23324%	015%	015%						LSE_IF.txt	17%

FIGURA 15 – Pesquisa por <LS> em LSE_IF na ferramenta Concord

Conforme esperado, a maioria dos casos de segmentação na LSE de IF incidiria em linhas de legenda onde há informações além das verbais apenas. No entanto, não se esperava que esse tipo de linha de legenda receberia etiqueta do tipo <LS>, a qual sinaliza uma segmentação que segue os padrões sintático-gramatical (veja próxima subseção). Antes pelo contrário,

supunha-se que pela carga de informação, essas legendas seriam focos de segmentações com estrutura sintático-gramatical diferente do padrão. Assim, destaca-se a importância de uma análise qualitativa que revele onde realmente estão os casos de segmentação, tanto da LSE quanto da LO, cuja gramática e sintaxe sejam desviantes do padrão.

Através da ferramenta Concord do Wordsmith Tools© 5.0 foi possível rastrear e agrupar todas as categorias de segmentações a partir das etiquetas atribuídas. Nesta subseção serão apresentados os detalhes dos dados da segmentação do corpus estudado.

3.2.1) SEGMENTAÇÃO <LS>

Como mencionado previamente, essa etiqueta marca as ocorrências de quebras de linha que incidem sobre duas estruturas completas, sendo cada uma delas colocada em uma linha. Por exemplo, duas partes de um diálogo, ou informações adicionais e texto de legenda, ou informações extralingüísticas e informações extralingüísticas. Como ilustrado, respectivamente, pelo exemplo [6] extraído da LO_IF, e pelos exemplos [7] e [8] extraídos da LSE_IF, a seguir:

[6] <L1077> - Você gostou da história, hein? <LS> - É, gostei! </LS> </L1077>

[7] <L3> [Paulo, em off] <LS> Ainda que eu fale as línguas </LS> </L3>

[8] <L18> [gritaria] <LS> [som de passos apressados] </LS> </L18>

Das 2408 linhas de legenda do corpus, 1291 são legendas de duas linhas. Dessas 1291 linhas, 618 linhas, ou seja, aproximadamente 47% das legendas de duas linhas do corpus estão nesse grupo de segmentação. O Gráfico 12 traz esses números em valores percentuais, para melhor visualização.

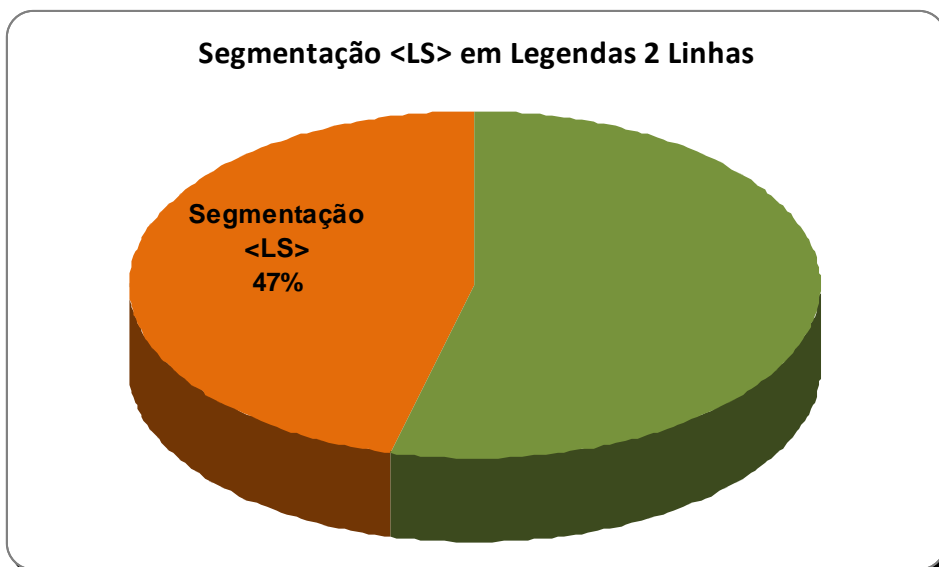


GRÁFICO 12 – Percentual de <LS> nas legendas de 2 linhas em IF

A segmentação do tipo <LS> está distribuída no corpus conforme apresentado na Tabela 3 abaixo em números absolutos.

TABELA 3
Distribuição da Segmentação <LS> em IF em números absolutos

	LO_IF	LSE_IF	CORPUS
LEGENDAS 2 LINHAS	606	693	1291
QUANTIDADE <LS>	267	351	618

Os dados da tabela mostram que a LSE_IF (351) contém mais legendas <LS> que a LO_IF (267). Os números proporcionais confirmam essa diferença e revelam que 47% das legendas de duas linhas em LSE_IF são do tipo <LS>; enquanto que 43% das legendas de duas linhas em LO_IF são do tipo <LS>. Gráfico 13 a seguir, para facilitar a visualização, apresenta os

valores percentuais da distribuição da segmentação <LS> nos subcorpora de IF em relação ao corpus total.

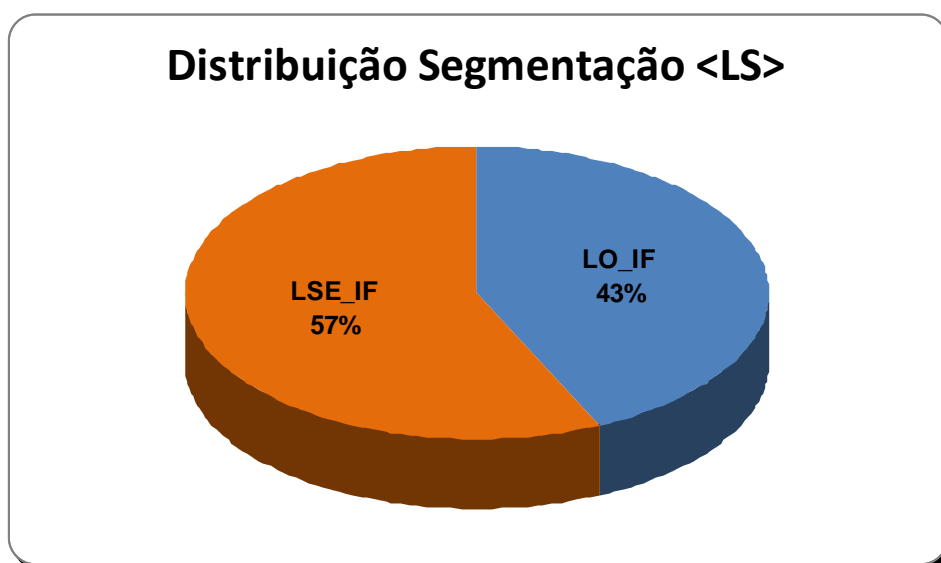


GRÁFICO 13 – Percentuais da distribuição da Segmentação <LS> em IF

A LSE_IF contém 57% das legendas <LS> de IF e a LO_IF contém 43% das <LS> e IF, segundo o gráfico acima. O resultado da pesquisa feita com ferramenta *Concord* para rastrear todas as etiquetas <LS> presentes no corpus IF revelou que, com exceção das linhas de legenda exclusivas da LSE_IF que contêm informações adicionais, todas as outras ocorrências dessa etiqueta ocorrem nas mesmas sentenças nos dois subcorpora.

Em se tratando da modalidade de legendagem, seja LO ou LSE, de acordo com Karamitroglou (1998) e Díaz-Cintas e Remael (2007) a linha que contém uma sentença clara, sem ambiguidades, que transmite uma informação completa, e ainda respeita critérios técnicos de legibilidade é considerada uma legenda ideal. Ainda segundo esses trabalhos, essas características, em geral, são encontradas em legendas de uma linha. Portanto, no corpus estudado, as legendas de uma linha (1109) somadas às legendas de segmentação <LS> (618) compõem uma figura de 71% do total do corpus IF, sendo esse o percentual aproximado de legendas ideais no corpus. Entretanto, para confirmar esse percentual seria necessário averiguar se as legendas de uma linha deste corpus realmente cumprem as exigências técnicas de legibilidade que as fariam qualificadas como ideais, o que foge ao escopo

desta pesquisa. Dessa forma, a *priori*, a maior parte do corpus da pesquisa é composta por legendas ideais.

3.2.2) SEGMENTAÇÃO <SOSN>/<SOSD> E <SOCN>/<SOCD>

Conforme explicado anteriormente, essas etiquetas incidem sobre orações subordinadas e coordenadas divididas em duas linhas de legenda. Os rótulos <SOSN> e <SOCN> marcam as ocorrências em que a segmentação obedece a uma divisão segundo padrões da gramática do PB para orações subordinadas e coordenadas, respectivamente; e que, conseqüentemente, leva em conta o princípio da segmentação no nóculo mais alto (KARAMITROGLOU, 1998); já os rótulos <SOSD> e <SOCD> marcam as ocorrências em que a segmentação desvia-se de uma divisão prevista nos padrões da gramática do PB para subordinações e coordenações, respectivamente; e que, conseqüentemente, viola o princípio da segmentação no nóculo mais alto (KARAMITROGLOU, 1998). Os exemplos [9], [10], [11] e [12] abaixo retirados da LSE_IF ilustram, os quatro tipos de ocorrência.

[9] <L1131> Morri pela lei <SOSN> a fim de viver para Deus! </SOSN>
</L1131>

[10] <L708> Mas não se preocupem que <SOSD> eu vou levá-los a casa de irmãos. </SOSD> </L708>

[11] <L785> Não poderia trabalhar aos sábados <SOCN> e deveria jejuar, </SOCN> </L785>

[12] <L1158> Mostra-me a sua fé sem obras, e eu <SOCD> lhe mostrarei a fé por minhas obras. </SOCD> </L1158>

No exemplo [9] há um período composto por uma oração principal colocada na primeira linha de legenda, e por uma oração subordinada adverbial final, introduzida por “a fim de” colocada inteiramente na segunda linha de legenda junto com a oração que introduz. Por sua vez, no exemplo [10] há uma oração subordinada substantiva adverbial causal. Sabe-se que uma adverbial

causal normalmente seria introduzida por “porque”, entretanto, nesse caso, há uma mistura de registro informal do português falado com algum socioleto – variação linguística atribuída à classe social – que provoca a redução da conjunção introdutória para “que”, a qual está na primeira linha de legenda separada da oração que introduz. No exemplo [11] há um período composto por coordenação, sendo que uma das orações está primeira linha de legenda e a oração coordenada sindética aditiva na segunda linha. Já no exemplo [12] há também um período composto por coordenação, porém a segmentação incidiu dentro da segunda oração, deixando uma parte da oração aditiva na primeira linha de legenda e separada de seus demais constituintes.

Das 2048 linhas de legenda do corpus, 1291 são legendas de duas linhas. Dessas 1291 legendas, 109 são legendas marcadas com <SOSN> e 57 são legendas marcadas com <SOSD>. Essa categoria de segmentação, portanto, representa aproximadamente 12% das legendas de duas linhas do corpus, sendo 8% <SOSN> e 4% <SOSD>. Quanto às coordenações, 50 legendas são marcadas com <SOCN> e 6 legendas são marcadas com <SOCD>. Essa categoria de segmentação, portanto, representa aproximadamente 4,2% das legendas de duas linhas do corpus, sendo 3,8% <SOCN> e 0,4% <SOCD>.

A distribuição dessas categorias de segmentação em IF está apresentada na Tabela 4 a seguir em números absolutos.

TABELA 4

Distribuição das Segmentações de Orações em IF em números absolutos

	LO_IF	LSE_IF	CORPUS
QUANTIDADE <SOSN>	55	54	109
QUANTIDADE <SOSD>	29	28	57
TOTAL <SOS*>	84	82	166
QUANTIDADE <SOCN>	25	25	50
QUANTIDADE <SOCD>	03	03	06
TOTAL <SOC*>	28	28	56

Os números apresentados na tabela acima mostram pequenas diferenças de segmentação de orações entre os subcorpora de IF. Isto é, LSE_IF e LO_IF possuem as mesmas quantidades de <SOCN> e <SOCD>. Apenas nas etiquetas <SOSN> e <SOSD> encontra-se uma ocorrência a mais de cada em LO_IF. Essa mínima diferença nos números absolutos não reflete nos números percentuais. Sendo assim, o percentual de distribuição de cada uma dessas etiquetas por subcorpus é de 50%. Segundo Díaz-Cintas e Remael (2007), na legendagem – seja LO ou LSE – o uso da coordenação é mais prático que o uso da subordinação, pois a quebra de linhas é favorecida quando as partes podem funcionar independentemente. Entretanto, com base na tabela acima nota-se que no corpus estudado, a quantidade de períodos compostos por subordinação é 46% maior que os períodos compostos por coordenação. A maior recorrência de orações subordinadas acarreta, portanto, um maior número de desvios nessas estruturas. Os desvios encontrados nas subordinações representam 54% das ocorrências em relação ao número de estruturas subordinadas normais; enquanto os desvios encontrados nas coordenações representam aproximadamente 10% das ocorrências em relação ao número de estruturas coordenadas normais. O uso das estruturas

subordinadas na legendagem, aparentemente, deve-se à preferência do legendista por legendas verbatim, ou seja, a transcrição literal do áudio com aplicação mínima ou nenhuma de estratégias de edição – cuja prática tornaria factível a transformação de subordinações em coordenações. Essa transformação, segundo Díaz-Cintas e Remael (2007), é a maneira mais fácil de favorecer a leitura de um texto segmentado. A coordenação permite que cada linha de legenda seja autossuficiente em termos sintáticos e semânticos, reforçando a leitabilidade e diminuindo o esforço de processamento do texto, o que é essencial para a apreciação do produto audiovisual como um todo, especialmente no caso da audiência SE.

As etiquetas de subordinação marcam 12% do total de legendas de duas linhas em IF e as de coordenação marcam 4% do total de legendas de duas linhas em IF. As orações subordinadas ocupam aproximadamente 13% das legendas de duas linhas em LO_IF, distribuídos como pode ser visualizado no Gráfico 14 abaixo.

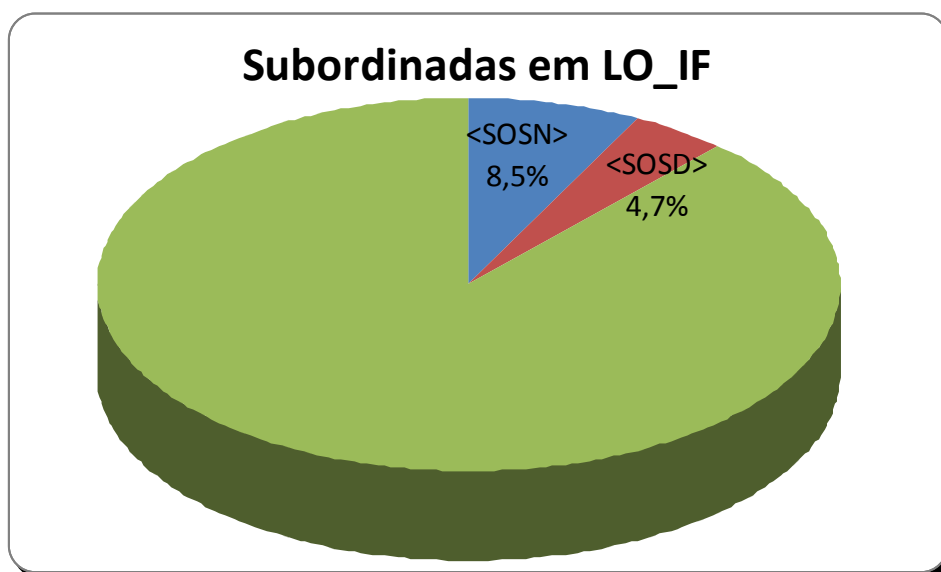


GRÁFICO 14 – Subordinadas nas legendas de 2 linhas em LO_IF

Em LSE_IF, as subordinadas representam aproximadamente 11% das linhas de duas legendas, distribuídos como mostrado no Gráfico 15 a seguir.

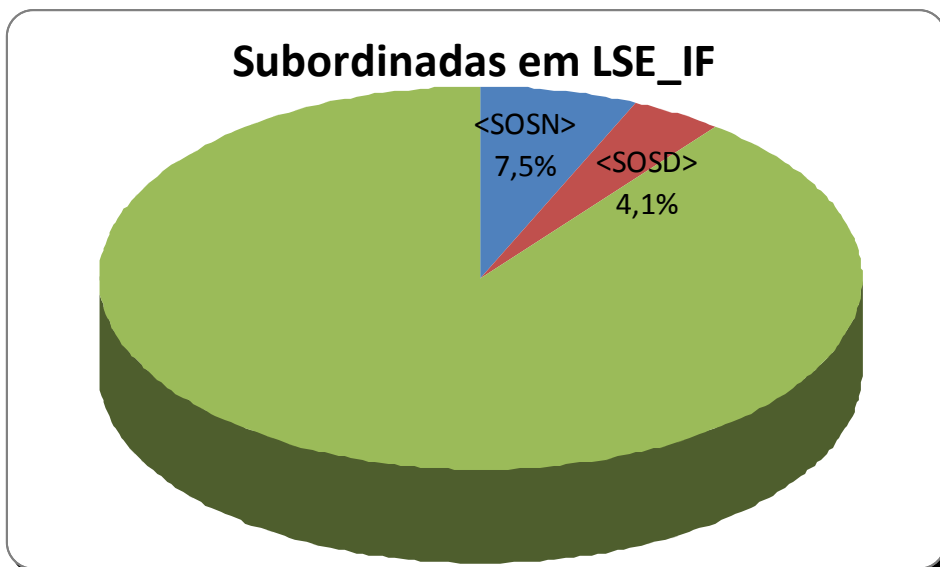


GRÁFICO 15 – Subordinadas em legendas de duas linhas LSE_IF

O Gráfico 16 abaixo revela que as orações coordenadas, ocupam aproximadamente 5% em LO_IF.

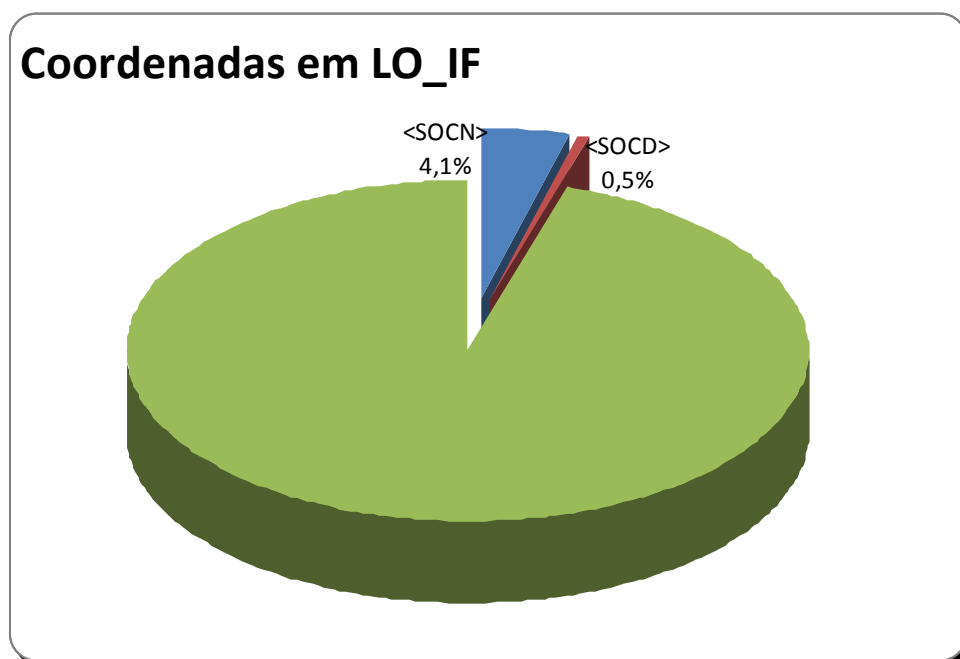


GRÁFICO16 - Coordenadas nas legendas de duas linhas LO_IF

Na LSE_IF o percentual de coordenadas é de 4% distribuídos como visualizado no Gráfico 17 a seguir.

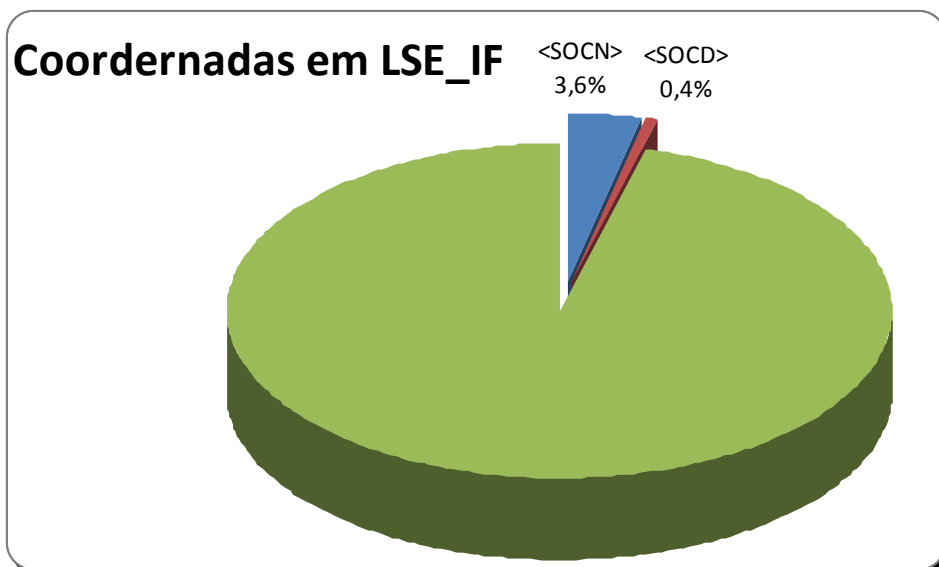


GRÁFICO 17 - Coordenadas nas legendas de duas linhas LSE_IF

Vale ressaltar que do total de 222 etiquetas atribuídas às estruturas de subordinação e coordenação 159 correspondem à quantidade de ocorrências marcadas como dentro da forma padrão segundo a gramática do PB, as quais estariam de acordo com a recomendação de Díaz-Cintas e Remael (2007) de que cada oração, em caso de período composto, deve ser colocada em uma linha de legenda. Isto significa que em aproximadamente 71% das ocorrências de períodos compostos em IF o legendista deu precedência a considerações linguísticas para segmentá-los.

Quanto à segmentação de orações subordinadas, como revelam os números absolutos da tabela anterior há apenas duas ocorrências diferentes, uma marcada <SOSN> e outra <SOSD> em LO_IF. Com exceção desses casos, o rastreamento das etiquetas revela que as ocorrências das segmentações de subordinações e coordenações incidem nas mesmas sentenças nos dois subcorpora. Como verificado nos gráficos apresentados anteriormente, a variação dos percentuais de distribuição das etiquetas, dentro dos subcorpora deve-se, portanto, exclusivamente à diferença de tamanho entre os textos.

O caso diferente de <SOSN> apresenta-se exclusivamente em LO_IF, como mostrado abaixo:

cada um dos argumentos
apresentados.

O caso anterior será discutido mais adiante na subseção 3.2 que trata de separação entre determinante/quantificador e substantivo <SSND2>.

O caso diferente de <SOSD> apresenta-se exclusivamente em LO_IF, como mostrado abaixo:

(A)
A morte é **o que**
você merece, herege!

Enquanto em LSE_IF, essa mesma sentença aparece segmentada como se segue:

(B)
Canalha! A morte
é o que você merece, herege!

A sentença em (A) foi etiquetada com desvio, pois rompe a estrutura da oração subordinada completiva nominal destacada em negrito acima. Em (B) a legenda recebeu etiqueta <LS>, pois na primeira linha está o sujeito e na segunda linha o predicado; configurando uma segmentação de legenda no nódulo mais alto (KARAMITROGLOU, 1998). Aparentemente a decisão diferente e (A) tomada pelo legendista deve-se ao elemento extra (“Canalha!”) contido em (B) e ausente em (A). A segmentação das duas sentenças foi praticada de maneira a manter a linha superior maior que a inferior. Esse caso reforça como critérios estéticos, muitas vezes, têm precedência sobre considerações linguísticas na prática da legendagem nesse filme.

Dois exemplos, um de cada tipo, das ocorrências que figuram entre os 29% dos casos de segmentação de orações com desvio da forma padrão do PB estão apresentados a seguir; para os quais são sugeridas novas segmentações com o objetivo de melhorar a legibilidade da linha de legenda. Para melhor visualização, as etiquetas foram retiradas e as legendas foram centralizadas para simular a posição dessas na tela. A linha de legenda abaixo traz um período composto de oração principal seguida de uma subordinada completiva nominal, marcada em negrito:

Nós já tivemos notícias **do que**
ocorreu no caminho de Damasco.

No entanto, a segmentação desse período não coincide com os limites entre as duas orações, uma vez que parte da subordinada está na primeira linha junto

com a oração principal. Nesse caso, aparentemente, o critério da geometria – linha superior e inferior simétricas - parece predominar, porém um pequeno ajuste renderia uma legenda melhor estruturada em termos linguísticos e ainda esteticamente apropriada, cuja linha superior seria menor que a inferior, favorecendo a visualização da imagem; como sugerido a seguir:

Nós já tivemos notícias
do que ocorreu no caminho de Damasco.

No exemplo apresentado a seguir, a legenda é um período composto por duas orações, sendo uma delas uma oração coordenada sindética aditiva, marcada em negrito:

Mostra-me a sua fé sem obras, **e eu
lhe mostrarei a fé por minhas obras.**

A segmentação aqui também não coincide com os limites dos períodos. A conjunção que introduz a oração coordenada e o sujeito da oração estão na primeira linha de legenda enquanto os outros elementos da oração estão na segunda linha de legenda. Mais uma vez parece que a geometria teve precedência sobre o aspecto linguístico. Nesse caso, com algumas intervenções é possível reescrever uma legenda linguisticamente mais apropriada e ainda sim esteticamente apropriada, como sugerido abaixo:

Mostra-me a sua fé sem obras,
e lhe mostrarei a fé por minhas obras.

Após uma análise manual das ocorrências de segmentação de período composto com desvio, identificou-se que em IF há 20 casos de legendas segmentadas por geometria, 21 casos em que se preferiu uma linha superior menor que a inferior e por fim, 22 casos semelhantes ao último citado acima, em que não é possível determinar o critério utilizado.

A seguir serão estudados os casos de quebra de linha que ocorrem dentro dos constituintes oracionais, começando pela segmentação de sintagmas nominais.

3.2.3) SEGMENTAÇÃO <SSN> E <SSND*>

Essa etiquetagem marca ocorrências em que a segmentação da linha de legenda incide em algum ponto entre elementos constituintes de um

sintagma nominal. A etiqueta <SSN> marca ocorrências em que a segmentação é feita entre os blocos constituintes, segundo Lemle (1989); já as etiquetas <SSVD*> marcam ocorrências que ocorrem dentro dos blocos constituintes, causando alguma ruptura semântico-estrutural que resulta em sequências não previstas em Lemle (1989), sendo que foram identificados três tipos diferentes de desvios. Os exemplos [13] e [14] a seguir, retirados das legendas de IF, ilustram um caso de <SSN> e um de <SSND*> respectivamente.

[13] <L475> Sou irmão de Pedro, Tiago <SSN> e os outros apóstolos de Cristo.
</SSN> </L475>

[14] <L442> Ester, dê um pouco <SSND2> de água pra ele. </SSND2>
</L442>

Das 2048 linhas de legenda do corpus, 1291 são legendas de duas linhas. Dessas 1291 legendas, 54 estão marcadas com <SSN> e 92 estão marcadas com <SSND*> de três tipos. Essa categoria de segmentação, portanto, representa aproximadamente 11% das legendas de duas linhas do corpus, sendo 4% <SSN> e 7% <SSND*>. A distribuição dessa categoria de segmentação no corpus está apresentada na Tabela 5 a seguir em números absolutos.

TABELA 5

Distribuição da Segmentação <SSN> e <SSND*> em IF em números absolutos

	LO_IF	LSE_IF	CORPUS
<SSN>	27	27	54
<SSND*>	46	46	92
TOTAL	73	73	146
CATEGORIAS DE <SSND>			
<SSND1>	04	04	08
<SSND2>	09	10	19
<SSND3>	33	32	65

A tabela acima inicialmente revela que os dois subcorpora são idênticos quanto aos números da segmentação de sintagmas nominais, tanto nas ocorrências dentro do padrão, quanto nas ocorrências marcadas como desvio. Sendo assim, cada um dos subcorpora detém 50% do total de ocorrências dessa categoria de etiquetagem em IF. Contudo, é importante ressaltar que há mais casos de segmentações com desvios (92) que casos de segmentações dentro do padrão (54). Ao averiguar-se as subcategorias de desvio, são encontradas duas ocorrências divergentes nas subcategorias <SSND2> e <SSND3>. Na LSE_IF há uma ocorrência de <SSND2> a mais que na LO_IF; enquanto que na LO_IF há uma ocorrência a mais de <SSND3> que na LSE_IF. Mais adiante, essas duas ocorrências serão estudadas.

O Gráfico 18 a seguir reinterpreta a distribuição das subcategorias de desvios de segmentações do sintagma nominal em LO_IF, em valores percentuais.

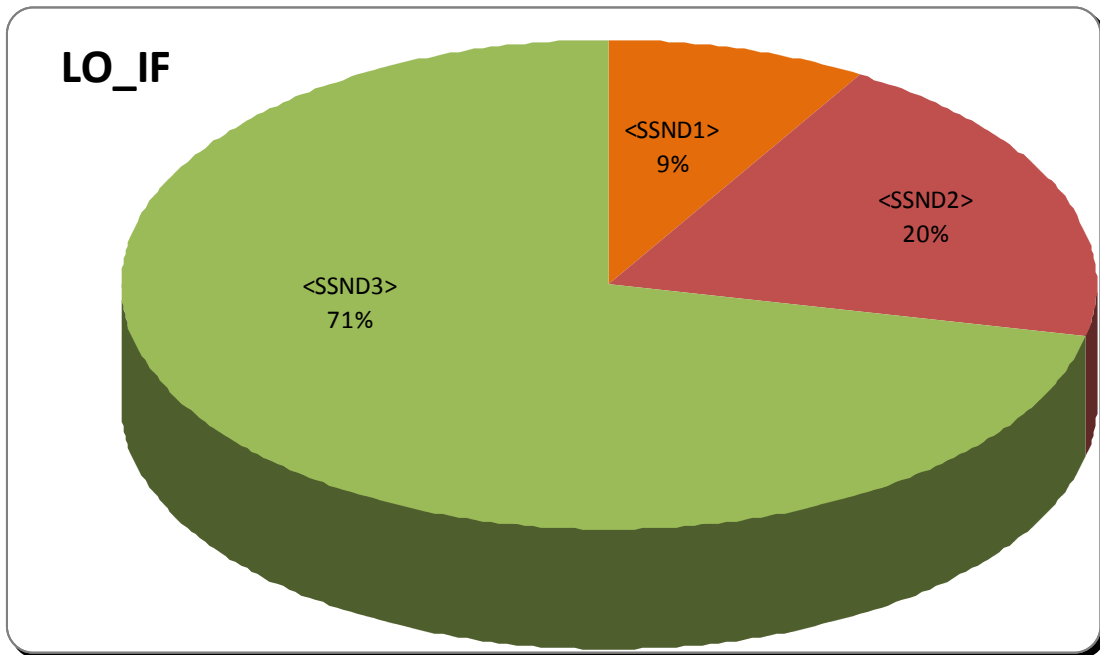


GRÁFICO18 - Distribuição das <SSND*> em LO_IF em valores percentuais

O Gráfico 19 abaixo apresenta a distribuição das categorias de desvios de segmentação do sintagma nominal em LSE_IF, em valores percentuais.

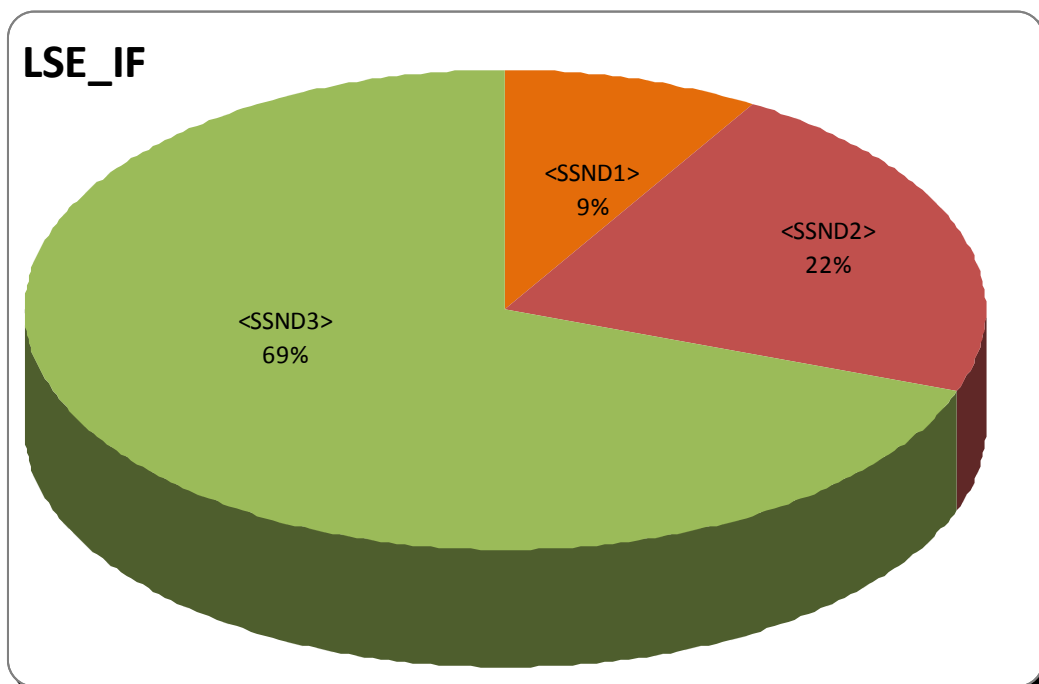


GRÁFICO19 - Distribuição das <SSND*> em LSE_IF

Os gráficos acima mostram que mais da metade – 71% em LO_IF e 69% em LSE_IF – dos desvios de segmentação de sintagmas nominais ocorrem na separação de um constituinte subsequente, tal como o sintagma preposicional modificador, do sintagma nominal (<SSND3>). Em seguida, a separação de determinante ou quantificador, constituintes do sintagma nominal, apresenta uma incidência de 20% em LO_IF e 22% em LSE_IF (<SSND2>). A separação do sintagma nominal de um constituinte adjetivo aparece como terceira causa de desvios de segmentação, com um percentual de 9% em cada subcorpus (<SSND1>).

A seguir as três categorias de segmentação <SSND> estão descritas e estudadas detalhadamente, a fim de esclarecer onde ocorrem e investigar possíveis razões para as rupturas na estrutura do sintagma nominal ao longo dos textos do corpus.

3.2.3.1) Subcategoria <SSND1>

Em cada um dos subcorpora há 04 casos dessa etiquetagem, os quais ocorrem nas mesmas sentenças nas duas legendagens do corpus. Como mencionado anteriormente, essa etiqueta marca os casos em que o constituinte adjetivo foi separado do sintagma nominal pela quebra da linha de legenda. Retomando a regra geral de segundo Lemle (1989) para sintagma nominal:

$$\text{RSN} \rightarrow (\text{Quant}(\text{Det})) (\text{Adj})^* \text{N} (\text{Adj})^* (\text{SPrep})^* (\text{Adj})^* (\text{S})^*$$

Foram marcadas as quebras de linha que incidiram assim:

$$(\text{Adj})^* \text{<SSND1> N ou N <SSND1> (\text{Adj})^*$$

Como são apenas quatro os casos distintos ocorridos no corpus, todos estão apresentados no Quadro 7 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 7
Ocorrências de <SSND1> em IF

C1	Mas há uma coisa, meu querido Saulo, que me atormenta.
C2	Esse homem é uma criatura renovada no amor de Jesus...
C3	nosso santo que hoje dirige a igreja mãe, na terra de Jerusalém.
C4	embora seja um pequeno membro do corpo,

Segundo o quadro acima, nas legendas em C1 e C4 o adjetivo fica na primeira linha de legenda separado do núcleo do sintagma nominal que está na segunda linha de legenda. Já nos outros casos – C2 e C3 – o núcleo do sintagma fica na primeira linha de legenda e o adjetivo na segunda linha de legenda.

A formatação das legendas é um indicativo da razão pela qual a segmentação se dá em determinado ponto do texto. No caso C2 observa-se que o comprimento das linhas inferior e superior é bastante simétrico, indicando a preferência por legendas geométricas. O outro tipo de formatação que pode determinar a quebra de linhas é a preferência por uma linha superior mais curta, favorecendo, assim, a visualização da imagem. Todavia, nos outros quatro casos mostrados acima, as legendas são formatadas exatamente ao contrário desse critério, ou seja, C1, C3 e C4 têm a linha superior mais longa que a linha inferior.

Considerando a recomendação de literatura da área (KARAMITROGLOU, 1998; DÍAZ-CINTAS E REMAEL, 2007; KALANTZI, 2008) de que a segmentação deve considerar aspectos linguísticos e estéticos, sendo que considerações linguísticas deveriam ter precedência; e observando aspectos técnicos da legendagem, os exemplos apresentados anteriormente foram reformulados e possíveis soluções estão apresentadas no Quadro 8 abaixo. Em alguns casos, como C3 e C4, uma nova segmentação é o suficiente para melhorar o aspecto linguístico da linha de legenda; em outros casos, maiores intervenções são necessárias.

QUADRO 8
Ocorrências <SSND1> reformuladas

C1R	Mas há uma coisa, querido Saulo, que me atormenta.
C2R	É uma criatura renovada no amor de Jesus...
C3R	nosso santo que hoje dirige a igreja mãe, em Jerusalém.
C4R	embora seja um pequeno membro do corpo,

Em C1R uma edição, a omissão do pronome “meu” – necessária para obedecer ao limite de caracteres por linha - e a ressegmentação desfazem não só a quebra do sintagma nominal, como também a ambiguidade que essa acarretava em C1 no trecho “Saulo, que me atormenta”. O fato de a linha superior da legenda superior ter ficado maior que a inferior, não é o ideal devido à visualização da imagem, mas ao verificar-se no vídeo observa-se que essa reformulação, além dos benefícios linguísticos já citados, gera uma legenda de segmentação retoricamente apropriada, isto é, que segue o ritmo da fala.

Já em C2R a reformulação tomou por base o contexto visual da cena em que a legenda aparece. A observação da imagem - que mostra o falante gesticulando e apontando para o homem ao qual se refere - e da legenda imediatamente anterior à C2 - que diz “Quem está aqui ao meu lado não é mais Saulo, nosso perseguidor.” - permitiu a omissão do SN “Esse homem”, possibilitando, assim, a reintegração do SN segmentado.

3.2.3.2) SUBCATEGORIA <SSND2>

Em LO_IF há 09 casos e em LSE_IF há 10 casos dessa etiquetagem, sendo que com exceção desse 01 caso de diferença, todas as ocorrências incidem nas mesmas sentenças nas duas legendagens do corpus. Como mencionado anteriormente, essa etiqueta marca os casos em que um quantificador ou determinante foi separado do sintagma nominal pela quebra da linha de legenda. Retomando a regra geral de segundo Lemle (1989) para sintagma nominal:

RSN → (Quant(Det)) (Adj)* N (Adj)* (SPrep)* (Adj)* (S)*

Foram marcadas as quebras de linha que incidiram assim:

(Quant (Det)) <SND2> N

Como são apenas dez os casos distintos ocorridos no corpus, todos estão apresentados no Quadro 9 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 9
Ocorrências de <SSND2>

C5	Herodes mandou matar todos os soldados que o guardavam.
C6	E Tito, traga-me a metade do vinho que conseguir beber.
C7	do que eu arredar de apenas uma só palavra proferida em nome Dele
C8	Eu quis trazê-los aqui, à minha casa, depois de toda a discussão
C9	Porque se é pela Lei que vem a justiça, então Jesus morreu em vão.
C10	Se antes a raiva, a obstinação e a cegueira me levaram a um caminho
C11	Ester, dê um pouco de água pra ele.
C12	Voltarei de Damasco com um grande número de nazarenos acorrentados.
C13	Eu acho que é muita responsabilidade para ele.
C14	cada um dos argumentos apresentados.

Nas ocorrências C5, C6, C7, C11, C12, C13 e C14 apresentadas no quadro anterior, a quebra de linha incide entre um quantificador e outros constituintes do sintagma nominal. Já nas outras ocorrências (C8, C9 e C10) há um determinante na primeira linha – artigo ou pronome – separado de outros constituintes do sintagma nominal que está na segunda linha.

A formatação das linhas de legenda dessa categoria indica que em IF dezesseis delas foram segmentadas de forma a manter a geometria das linhas; em uma delas manteve-se a linha superior mais curta; e em duas delas não se pode determinar que um critério estético tenha motivado a

segmentação. A fim de dar precedência às considerações linguísticas e assim reunir os constituintes do sintagma nominal em algumas dessas ocorrências, pequenas edições e ressegmentação seriam suficientes e resultariam em legendas tais como exibidas no Quadro 10 abaixo.

QUADRO 10
Ocorrências <SSND2> reformuladas

C5R	Herodes mandou matar todos os soldados que o guardavam.
C6R	E Tito, traga-me metade do vinho que beber.
C7R	que eu arredar de uma palavra proferida em nome Dele
C8R	Quis trazê-los à minha casa, depois de toda discussão
C9R	Porque se pela Lei vem a justiça, então Jesus morreu em vão.
C10R	Se antes a raiva, a obstinação e a cegueira me levaram a um caminho
C11R	Ester, dê-lhe um pouco de água.
C12R	Voltarei de Damasco com muitos nazarenos acorrentados.
C13R	Acho que é muita responsabilidade.
C14R	Cada argumento apresentado

A respeito das reformulações propostas no quadro acima, há algumas considerações a serem feitas. A ressegmentação praticada em C7R não apenas ajusta a estrutura do SN, como também permite que na primeira linha de legenda haja uma sentença completa e na segunda linha seja colocada a oração subordinada adjetiva reduzida de particípio. Em C10 há um SN complexo com vários SNs coordenados (a raiva, a obstinação e a cegueira), a reformulação em C10R, então, reuniu o núcleo ao seu determinante (a cegueira) e ainda posicionou a conjunção coordenativa “e” junto ao último elemento da coordenação na segunda linha de legenda.

Na ocorrência C13R, as intervenções foram maiores, visto que a ressegmentação não seria suficiente, pois geraria uma das seguintes possibilidades:

(A)

Eu acho que é
muita responsabilidade para ele.

(B)

Eu acho que é muita responsabilidade
para ele.

A ressegmentação em (A) resultaria em uma legenda com desvio de segmentação do sintagma verbal, em que o verbo estaria separado de seu complemento. A ressegmentação em (B) resultaria em linhas de legenda esteticamente desfavoráveis, especialmente à apreciação da imagem, além de a linha superior ultrapassar o limite recomendável de caracteres. Portanto, a primeira intervenção foi a omissão do pronome “Eu”, item considerado dispensável na legendagem (NEVES, 2005) em línguas *pro-drop* como o PB. A seguir os constituintes do sintagma nominal foram ajuntados com a ressegmentação da legenda. Finalmente, com base no contexto do vídeo foi possível omitir o sintagma preposicional “para ele” ao verificar-se que se faz referência a esse “ele” nas legendas anteriores e subsequentes a essa; e assim, foi possível transformar C13 em uma legenda de uma linha, cuja velocidade de leitura reduziu de 180 palavras por minuto para 145 palavras por minuto (DÍAZ-CINTAS E REMAEL, 2007).

Os casos C5 a C13 aparecem nos dois subcorpora de IF, porém C14 é uma ocorrência exclusiva de LSE_IF. Na LO_IF a legenda em C14 aparece segmentada como apresentado no Quadro 11 abaixo.

QUADRO 11
Sentença de C14 em LO_IF

cada um dos argumentos apresentados.

Em LO_IF essa linha de legenda recebeu etiqueta <SOSN>, como mostrado anteriormente na subseção 2. Essa etiquetagem deve-se ao fato de que o trecho “cada um dos argumentos” é na verdade objeto direto na sentença “eu acompanhei” que se iniciou há três legendas anteriores. Sendo assim, o verbo “apresentados” é parte do sintagma nominal “cada um dos argumentos

apresentados” cuja quebra de linha incidu entre o núcleo do sintagma e sua subsequente oração subordinada adjetiva reduzida de participípio.

Nenhuma particularidade dos dois tipos de legenda – LO e LSE – parece explicar a diferença de segmentação adotada pelo legendista. Aparentemente, na LSE_IF o legendista favoreceu o critério estético da linha superior mais curta; enquanto que na LO_IF o legendista favoreceu as considerações linguísticas. Para haver uma conciliação entre esses dois aspectos, uma pequena edição renderia a linha de legenda proposta em C14R, a qual inclusive estaria mais confortavelmente acomodada dentro do tempo de duração designado, deixando de ter uma velocidade de leitura de 180 palavras por minuto, para ter 160 palavras por minuto.

Após as intervenções, as ocorrências C10R a C14R transformaram-se em linhas de legenda cuja segmentação leva em conta a sintaxe do PB e, ainda assim, permanecem dentro dos parâmetros técnicos inerentes à legendagem.

3.2.3.3) SUBCATEGORIA <SSND3>

Em LO_IF há 33 casos e em LSE_IF há 32 casos dessa etiquetação, sendo que com exceção desse 01 caso de diferença, todas as ocorrências incidem nas mesmas sentenças nas duas legendagens do corpus. Como mencionado anteriormente, essa etiqueta marca os casos em que um sintagma preposicional constituinte do SN e que complementa o N foi separado do sintagma nominal pela quebra da linha de legenda. Retomando a regra geral segundo Lemle (1989) para sintagma nominal:

$$\text{RSN} \rightarrow (\text{Quant}(\text{Det})) (\text{Adj})^* \text{N} (\text{Adj})^* (\text{SPrep})^* (\text{Adj})^* (\text{S})^*$$

Foram marcadas as quebras de linha que incidiram assim:

$$\text{N} \text{ <SSND3> } (\text{SPrep})^*$$

Como são muitos os casos desta ocorrência no corpus, alguns foram selecionados e estão apresentados no Quadro 12 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 12
Exemplos de ocorrências de <SSND3> em IF

C15	Eu sou um homem como vocês, de carne...
C16	Amanhã, seu corpo vai sentir saudade das dores de hoje, meu caro.
C17	Confrontar Tiago na igreja de Jerusalém? Isso não lhe aflige?
C18 (Cf.C51)	Há problema em levar a palavra de Cristo a algum homem na Terra?

Para melhor visualização do tipo de desvio marcado nessas legendas, a árvore sintática da ocorrência em C15 está apresentada a seguir. De acordo com Karamitroglou (1998) é preferível que a segmentação coincida com o mais alto nó sintagmático, a fim de conter um maior agrupamento de carga semântica e transmitir uma informação completa ao espectador. A Figura 16 a seguir apresenta a análise sintática de C15 em diagrama de árvore, com seus respectivos nós sintáticos sinalizados à direita da imagem.

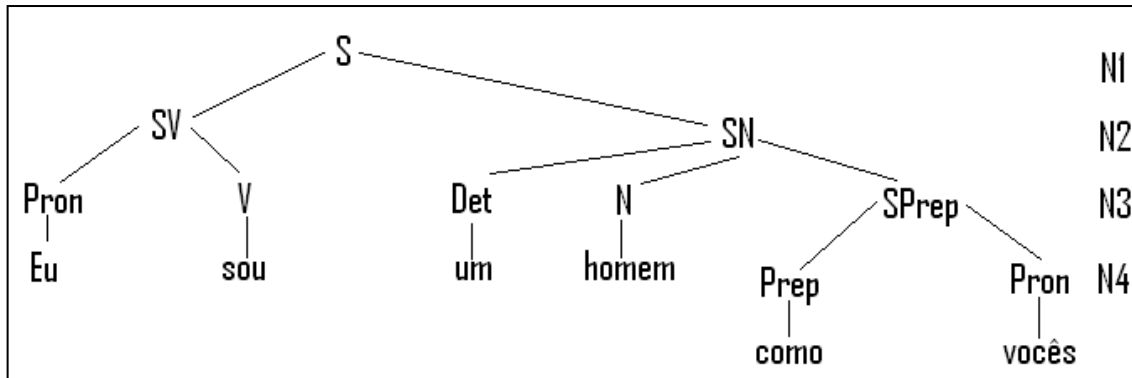


FIGURA 16 – Diagrama em árvore da ocorrência C15

Como pode ser visualizado na árvore anterior, a segmentação de C15 ocorreu em N3, ocasionando a quebra do sintagma nominal em N2. É provável que essa quebra tenha sido ocasionada por uma segmentação por geometria. Ao examinar-se essa ocorrência no vídeo, foi possível editá-la e ressegmentá-la, criando uma legenda de uma linha e uma nova sequência de linhas de legendas, como mostrado no Quadro 13 a seguir.

QUADRO 13
Reformulação de C15

C15R	<L977>	Sou um homem como vocês	</L977>
	<L978>	de carne	</L978>

Além de reconsiderar a estrutura sintática de C15, essa nova segmentação também valoriza a retórica (o ritmo do diálogo na cena) e o aspecto visual do filme, uma vez que há um corte na cena com movimentação de câmera entre as linhas <L977> e <L978> que favorece a transição de uma linha legenda para outra. Assim como mostrado na árvore sintática de C15, as outras ocorrências exemplificadas (C16 a C19) apresentam o mesmo tipo de quebra do sintagma nominal, o núcleo do sintagma nominal separado de um sintagma preposicional que o complementa. Há uma diferença no caso C16, no qual, na verdade, há dois sintagmas preposicionais depois do substantivo, sendo o segundo SPrep (de hoje) um complemento do primeiro SPrep (das dores). Assim, observa-se que o legendista preferiu segmentar entre SN e SPrep, a segmentar entre os SPreps. O Quadro 14 abaixo apresenta uma proposta de ressegmentação para C16.

QUADRO 14
Reformulação de C16

C16R	Amanhã, seu corpo sentirá saudades das dores de hoje, meu caro.
------	--

A nova legenda proposta ajusta a estrutura do SN reunindo o núcleo ao SPrep que o modifica. No entanto, a relação semântica, nesse caso, entre “sentir” e “saudades” é mais forte que a relação estrutural do N com o SPrep. Isto é, o verbo “sentir” sozinho pode ter vários sentidos, mas “sentir saudades” tem um significado conjunto, aparentemente, inseparável. Assim, a segmentação proposta pelo legendista parece mais linguisticamente adequada que a reformulação sugerida aqui. Esse tipo de caso aponta a necessidade de investigar-se com o público alvo os efeitos da segmentação no processamento da informação transmitida por essas colocações do PB.

Quanto à formatação das legendas marcadas com esse desvio, foram identificados em IF 14 casos de segmentação por geometria, 17 casos de segmentação por linha superior mais curta e 34 casos de segmentação cuja motivação lingüística ou estética não pôde ser determinada. Há uma ocorrência (C18) apresentada no quadro anterior, que foi marcada em LO_IF, mas não aparece em LSE_IF, cuja segmentação aparentemente foi definida pela geometria das linhas. Essa linha de legenda em LSE_IF aparece segmentada como se segue:

Há problema em levar a palavra **de**
Cristo a algum homem na Terra?

Enquanto, em LO_IF essa legenda teve sua quebra de linha rompendo a estrutura do sintagma nominal, em LSE_IF a ocorrência acima foi etiquetada com desvio na segmentação de um sintagma preposicional constituinte do SN (<SSPrepD3>). A árvore sintática da sentença em questão está apresentada na Figura 17 abaixo.

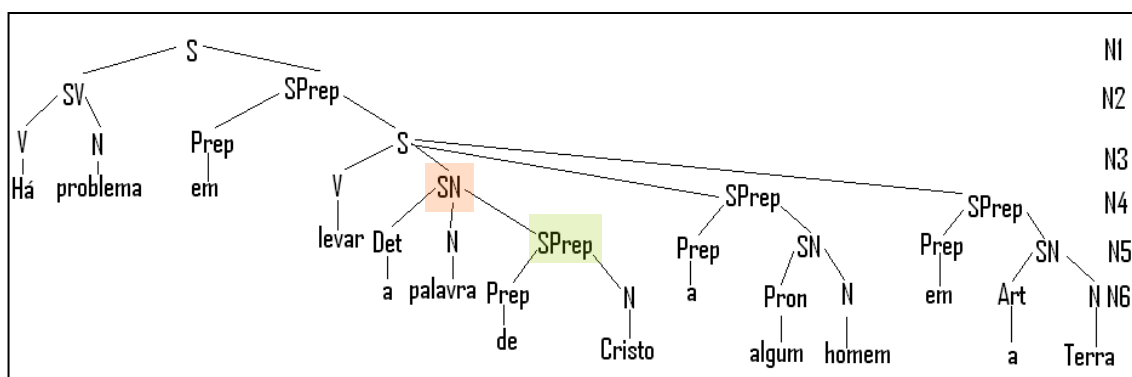


FIGURA 17 – Árvore sintática da sentença em C19

Uma análise do diagrama acima revela que a segmentação da sentença em LO_IF deu-se no nóculo sintático N5 rompendo a estrutura do SN em destaque; já a segmentação em LSE_IF deu-se no nóculo sintático N6 rompendo a estrutura do SPrep em destaque. A quebra de linha em N6, além da ruptura no sintagma preposicional, gera uma legenda cuja motivação estética não pode ser determinada, visto que não estaria geométrica, nem com a linha superior mais curta; antes pelo contrário, a linha superior ficaria ligeiramente mais longa que a inferior. Por sua vez, a segmentação em N5, apesar de quebrar a estrutura do sintagma nominal, gera uma legenda, pelo

menos, esteticamente motivada pela geometria. Ademais, partindo do princípio da segmentação no nóculo mais alto (KARAMITROGLOU, 1998) pode-se considerar essa última segmentação como preferível - pois concilia consideração linguística e critério estético - visto que uma versão de segmentação sem desvios nessa sentença parece ser improvável de atingir-se. Por fim, essa análise juntamente com um exame do filme também esclarece que não há nenhuma peculiaridade dos tipos de legendagem – LO e LSE – que parece motivar o legendista a tomar duas decisões diferentes em relação à segmentação da mesma sentença em cada um das modalidades.

3.2.4) SEGMENTAÇÃO <SSV> E <SSVD*>

Essa etiquetagem marca ocorrências em que a segmentação da linha de legenda incide em algum ponto entre elementos constituintes de um sintagma verbal. A etiqueta <SSV> marca ocorrências em que a segmentação é feita entre os blocos constituintes, segundo Lemle (1989); já as etiquetas <SSVD*> marcam ocorrências de segmentação que incidem dentro dos blocos constituintes, causando alguma ruptura semântico-estrutural que resulta em sequências não previstas em Lemle (1989), sendo que foram identificados três tipos diferentes de desvios. Os exemplos [15] e [16] a seguir, retirados da LSE de IF, ilustram um caso de <SSV> e um de <SSVD*> respectivamente.

[15] <L760> Paulo, os que o agrediram <SSV> na sinagoga... </SSV> </L760>

[16] <L878> Se os gentios querem <SSVD3> seguir o Caminho, ótimo... </SSVD3> </L878>

Das 2048 linhas de legenda do corpus, 1291 são legendas de duas linhas. Dessas 1291 legendas, 89 estão marcadas com <SSV> e 161 estão marcadas com <SSVD*> de três tipos diferentes. Essa categoria de segmentação, portanto, representa aproximadamente 19% das legendas de duas linhas do corpus, sendo 6% <SSV> e 13% <SSVD*>. A distribuição em números absolutos dessa categoria de segmentação no corpus está

apresentada na Tabela 6 e em valores percentuais nos Gráficos 19 e 20 a seguir, que mostram os percentuais por subcategorias <SSV> e <SSVD> respectivamente.

TABELA 6
Distribuição da Segmentação de Sintagmas Verbais
em IF em números absolutos

	LO_IF	LSE_IF	CORPUS
<SSV>	44	45	89
<SSVD*>	80	81	161
TOTAL	124	126	250
CATEGORIAS DE <SSVD>			
<SSVD1>	51	52	103
<SSVD2>	21	21	42
<SSVD3>	08	08	16

A tabela acima inicialmente revela que há diferenças, porém mínimas, entre os dois subcorpora quanto aos números da segmentação de sintagmas verbais, tanto nas ocorrências dentro do padrão, quanto nas ocorrências marcadas como desvio. Ainda assim, cada um dos subcorpora detém 50% do total de ocorrências dessa categoria de etiquetagem em IF. Contudo, é importante ressaltar que há mais casos de segmentações com desvios (161) que casos de segmentações dentro do padrão (89). Ao averiguarem-se as subcategorias de desvio, foi encontrada uma ocorrência divergente nos subcorpora na subcategoria <SSVD1>. Na LSE_IF há uma ocorrência de <SSVD1> a mais que na LO_IF; e há também uma ocorrência marcada como padrão em LSE_IF, a qual não aparece em LO_IF, pois incide em uma linha de legenda exclusiva para informação extralinguística.

O Gráfico 20 abaixo reapresenta a distribuição das subcategorias de desvios de segmentações do sintagma verbal em LO_IF, em valores percentuais.

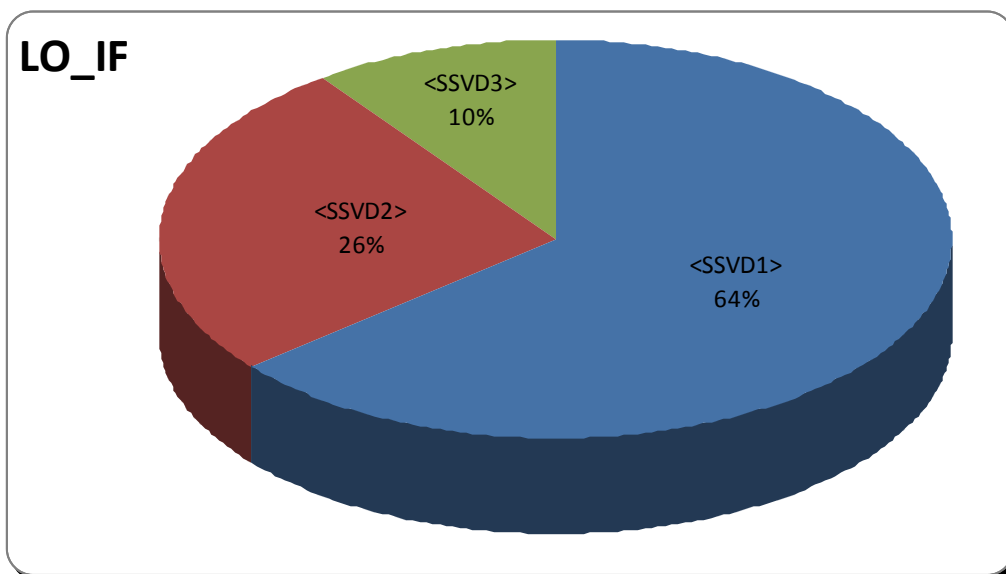


GRÁFICO 20 – Distribuição das <SSVD*> em LO_IF em valores percentuais

O Gráfico 21 abaixo reapresenta a distribuição das subcategorias de desvios de segmentações do sintagma verbal em LSE_IF, em valores percentuais.

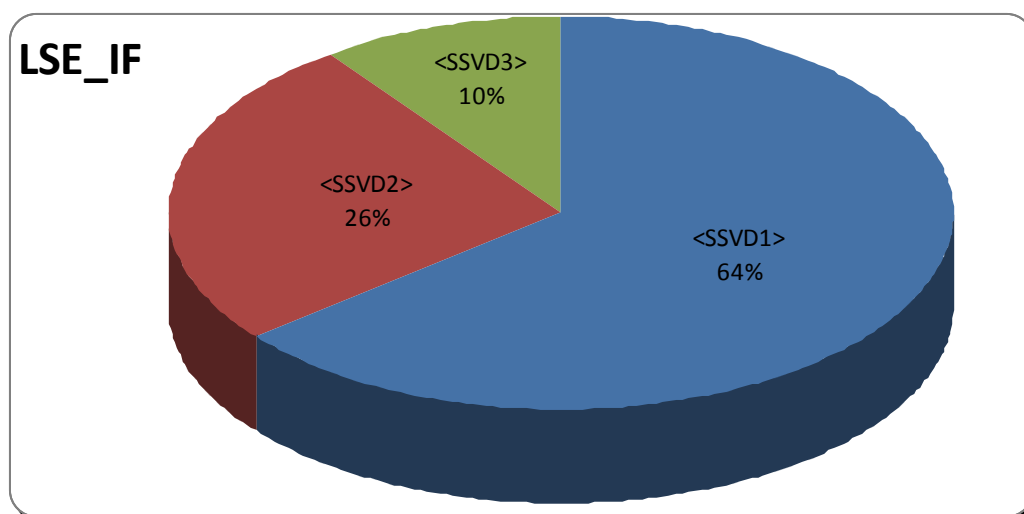


GRÁFICO 21 – Distribuição das <SSVD*> no corpus em valores percentuais

Os gráficos acima mostram que mais da metade em cada subcorpus (64%) dos desvios de segmentação de sintagmas verbais ocorrem na separação entre sintagma verbal e um constituinte objeto ou complemento (<SSVD1>). A separação entre verbos em sintagmas formados por complexos verbais aparece como segunda causa de desvios de segmentação, com um

percentual de 26% em cada subcorpus (<SSVD2>). Por fim, a segmentação entre verbo e advérbio representa um percentual de 10% em cada subcorpus (<SSVD3>).

A seguir as três categorias de segmentação <SSVD> estão descritas e estudadas detalhadamente, a fim de esclarecer onde ocorrem e investigar possíveis razões para as rupturas na estrutura do sintagma verbal ao longo dos textos do corpus.

3.2.4.1) SUBCATEGORIA <SSVD1>

Em LO_IF há 51 casos e em LSE_IF há 52 casos da etiquetagem <SSVD1>. Com exceção desse 01 caso de diferença, todas as ocorrências incidem nas mesmas sentenças nas duas legendagens do corpus. Como já mencionado, essa etiqueta marca os casos em que o objeto ou complemento foi separado de seu sintagma verbal pela quebra da linha de legenda. Retomando a regra geral segundo Lemle (1989) para o sintagma verbal:

$$RSV \rightarrow (Adv)^* V (Adv)^* \begin{array}{c} \boxed{SN} \\ Adj \end{array} (SPrep)^* (S)^*$$

Foram marcadas as quebras de linha que incidiram assim:

$$V <SSVD1> \begin{array}{c} \boxed{SN} \\ Adj \end{array}$$

Como são muitos os casos distintos ocorridos no corpus, alguns exemplos foram selecionados e estão apresentados no Quadro 15 a seguir. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 15
Exemplos de ocorrências de <SSVD1> em IF

C19	que sua obra já cruza as fronteiras de Jerusalém...
C20	Só o seu avô, ele não vai conseguir a soltura do teu irmão.
C21	Ainda que eu dê aos pobres tudo o que possuo,
C22	e tem no lugar da língua uma cobra!
C23	As coisas andam muito perigosas por aqui.

Nas ocorrências C19 e C20 apresentadas no quadro anterior, a quebra de linha incide entre o verbo e seu objeto direto. Em C21 o verbo está junto de seu objeto indireto, porém separado de seu objeto direto. Já a ocorrência em C22 é aquela que só aparece em LSE_IF, cujo verbo e objeto direto estão separados pela segmentação e intercalados por um sintagma preposicionado. E em C23 o verbo aparece separado de seu complemento.

A linha de legenda em C20 é mais um dos casos em que uma versão sem desvios parece ser difícil de atingir. A Figura 18 a seguir apresenta o diagrama em árvore para a sentença em C20.

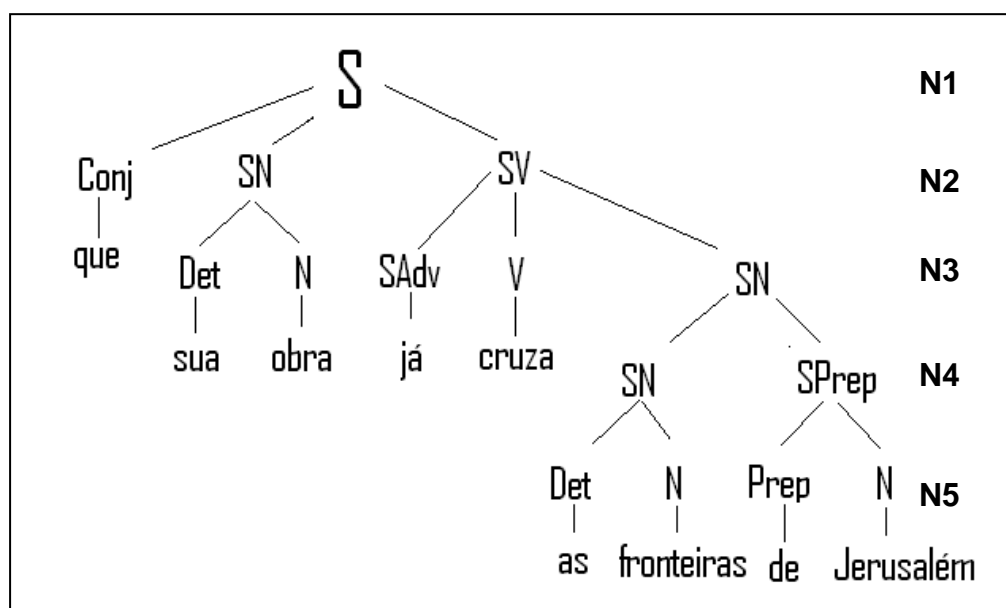


FIGURA 18 – Diagrama em árvore da sentença em C19

Segundo a regra em Lemle (1989), a quebra de linha adotada pelo legendista rompe a estrutura do sintagma verbal ao separar o verbo de um sintagma nominal subsequente – o objeto direto - o que pode ser confirmado pela análise do diagrama acima, o qual mostra que essa quebra ocorre no nódulo sintático N3. A junção de verbo e sintagma nominal geraria a linha de legenda mostrada no Quadro 16 abaixo.

QUADRO 16
Ressegmentação de C19

C19R	que sua obra já cruza as fronteiras de Jerusalém...
------	---

Entretanto, um exame do diagrama acima revela que nesse caso a quebra se daria em um nódulo sintático ainda mais baixo (N4) rompendo a estrutura do SN objeto. Além disso, a ressegmentação sugerida em C19R cria uma legenda cuja linha superior passa a ser mais longa que a inferior o que é desfavorável à apreciação da imagem. Seria possível, também, a segmentação apresentada a seguir no Quadro 17.

QUADRO 17
Segunda Ressegmentação de C19

C19R	que sua obra já cruza as fronteiras de Jerusalém...
------	---

Essa ressegmentação ocorre em um nódulo mais alto (N2) com uma legenda de linha superior mais curta e uma linha inferior que por pouco ultrapassa o limite de caracteres. Esse tipo de ocorrência levou Kalantzi (2008) a elaborar a recomendação de que é preferível separar o verbo do objeto a separar o sujeito do seu verbo – como feito originalmente pelo legendista de IF. Ao contrário do que esta pesquisa investiga, a recomendação de Kalantzi (2008) propõe que, em casos desse tipo, a segmentação ocorra dentro da estrutura sintagmática do SV e que constituintes da ordem da oração - sujeito e verbo – sejam preservados juntos. Esse caso é um exemplo típico de impasse com o qual o legendista depara-se e cuja opinião do público alvo seria determinante na tomada de decisão.

A ocorrência em C20 é um dos casos cuja ressegmentação e algumas pequenas edições reuniriam o sintagma rompido. Primeiramente, essa linha de legenda pode ser dividida em duas, gerando uma nova sequência tal como no Quadro 18 a seguir.

QUADRO 18
Reformulação de C20

C20R	<L898>	Só o seu avô	</L898>
	<L899>	ele não vai conseguir a soltura do teu irmão.	</L899>

Em segundo lugar a linha <L899> pode ser editada para caber em uma linha, como mostrado abaixo no Quadro 19 abaixo.

QUADRO 19
Segunda Reformulação de C20

C20R	<L898>	Só o seu avô	</L898>
	<L899>	não vai conseguir a soltura do Paulo.	</L899>

Essa reformulação de C20, além de eliminar a ruptura no SV, gera legendas mais lentas, mais adequadas esteticamente, que seguem o ritmo do falante na tela do vídeo, e ainda, a eliminação do “ele” retira o traço de oralidade atendendo o público SE brasileiro que já indicou em experimentos conduzidos no Brasil (ARAÚJO, 2004, 2007, 2008; FRANCO; ARAÚJO, 2003) preferência por um registro mais formal do PB na legendagem.

O caso em C21 tem uma estrutura de SV um pouco diferente daquela prevista em Lemle (1989). Conforme a regra geral para SVs o único elemento obrigatório no SV é o verbo. Após o verbo, e dependendo do tipo desse, pode-se selecionar um objeto (SN) para verbos de ação, ou um Adjetivo para verbos copulativos, o qual pode ser ou não seguido de um sintagma preposicional. Por sua vez, o SV da ocorrência C21 é uma estrutura que possui verbo, SPrep e SN como pode ser visualizado na árvore da Figura 19 abaixo.

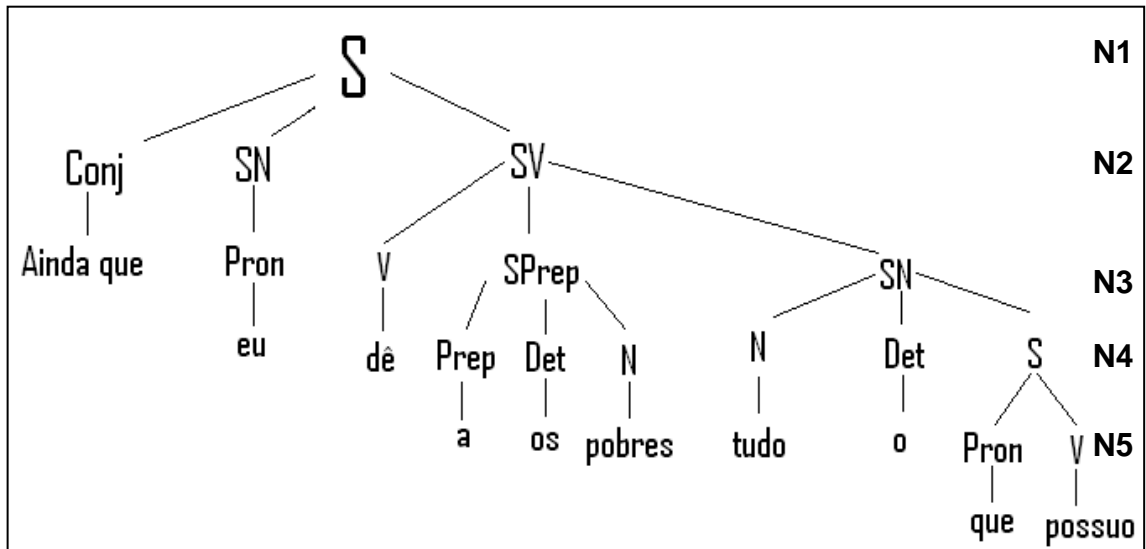


FIGURA 19 – Diagrama em árvore da sentença em C22

No entanto, como mostrado no diagrama acima, a ordem dos elementos do SV em C21 desafia a regra geral de Lemle (1989). O verbo em C21 vem seguido por um SPrep. Entretanto, a regra em Lemle (1989) prevê que haja um verbo de ação seguido de um SN (objeto direto) ou um verbo copulativo seguido de um adjetivo (predicativo do sujeito), mas a maneira como a regra foi postulada – a ordem dos constituintes e a possibilidade de escolha oferecida (SN ou Adjetivo) - não parece prever que um verbo de ação venha seguido de um SPrep como seu objeto indireto. O fato é que o caso em C21 apresenta um verbo seguido de um SPrep que é seu objeto indireto introduzido pela preposição “a”, a qual deve ser colocada com o verbo quando há objeto indireto. Sendo assim, com base na regra geral, C21 foi marcado *a priori* como desvio, pois o verbo está separado do SN subsequente. Para reparar essa ruptura seria necessário inverter a ordem em que os elementos aparecem na sentença, operação cujo resultado seria como exposto no Quadro 20 seguinte.

QUADRO 20
Reformulação tentativa de C21

C21R	Ainda que eu dê tudo o que possuo aos pobres
------	---

Essa reformulação em C21R, todavia, não atende as necessidades da modalidade, pois gera uma linha de legenda superior ainda mais longa que em C21 e ainda separa a preposição “a” do verbo ao qual está ligada. Portanto, apesar de desafiar a descrição de Lemle (1989) do SV em PB, nesse caso a proposta do legendista de IF mostrou-se como a opção mais adequada linguisticamente e também mais esteticamente apropriada para a modalidade.

A ocorrência C22, exclusiva da LSE_IF, também é uma estrutura diferente da proposta em Lemle (1989). Nesse caso o verbo também é seguido por um SPrep, contudo, esse SPrep não é um dos possíveis objetos do verbo, mas sim um adjunto adverbial, que está inserido entre o verbo e seu objeto SN. Em LO_IF essa linha de legenda não recebe etiquetagem, pois se apresenta como legenda de uma linha, como mostrado no Quadro 21 abaixo.

QUADRO 21
Sentença de C22 em LO_IF

C22	e tem no lugar da língua uma cobra!
-----	-------------------------------------

Ao examinar-se essa ocorrência no vídeo, depois de eliminada a possibilidade de alguma especificidade do tipo de legendagem ou qualquer motivação estética serem a razão da linha em LSE_IF ter sido quebrada, é possível inferir que a segmentação é uma tentativa de retratar a retórica do discurso. O falante na tela faz uma pausa ao falar “e tem no lugar da língua” e então completa “uma cobra!”. Realmente o filme revela que o ritmo da fala nessa cena é o que mais fortemente dita o perfil dessa linha de legenda – tendo precedência até sobre considerações linguísticas. No entanto, o elemento de suspense, como já é convenção da área, ficaria mais bem acomodado sendo transmitido em duas linhas de legenda, como ilustrado no Quadro 22 a seguir.

QUADRO 22
Reformulação de C22

C22R	<L724>	e tem no lugar da língua	</L724>
	<L725>	uma cobra!	</L725>

A ocorrência em C23 apresenta uma ruptura entre um verbo copulativo e o predicativo do sujeito. Segundo a regra geral em Lemle (1989), a lacuna após o verbo pode ser preenchida por um SN ou um “Adj” (adjetivo), porém em C23 o complemento do verbo é desempenhado por um sintagma adjetivo. Uma pequena edição e a ressegmentação poderiam reunir verbo e predicativo em uma única linha de legenda, como apresentado no Quadro 23 abaixo.

QUADRO 23
Reformulação de C24

C23	As coisas andam muito perigosas aqui.
-----	---------------------------------------

O exame da formatação das ocorrências dessa etiqueta mostrou que das 103 ocorrências de SV com esse desvio em IF, 26 foram formatadas por geometria, 36 foram formatadas com a linha superior mais curta e 41 casos não se pode determinar o critério estético adotado que pudesse preceder às considerações linguísticas na segmentação de linha.

3.2.4.2) SUBCATEGORIA <SSVD2>

Em cada um dos subcorpora há 21 casos dessa etiquetagem, sendo que todas as ocorrências incidem nas mesmas sentenças nas duas legendagens do corpus. Essa etiqueta marca os casos em que verbos que formam um complexo verbal foram separados um do outro pela quebra da linha de legenda. Retomando a regra geral de segundo Lemle (1989) para sintagma verbal:

$$RSV \rightarrow (Adv)^* V (Adv)^* \left[\begin{array}{c} SN \\ Adj \end{array} \right] (SPrep)^* (S)^*$$

É possível ver que a regra não especifica esse tipo de complexo verbal. Sendo assim, foram marcadas as ocorrências em que o verbo principal foi separado de um auxiliar que o acompanhe.

Visto que são muitos os casos distintos ocorridos no corpus, três exemplos foram selecionados e estão apresentados no Quadro 24 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 24
Exemplos de ocorrências de <SSVD2> em IF

C24	Vocês ainda vão conhecer Jesus Cristo!
C25	Os caras do tráfico ficam dando dinheiro pros meninos.
C26	Nós precisamos sair dessa cidade, Tito

Nas três ocorrências apresentadas acima, os verbos do complexo verbal destacados em negrito estão separados um do outro pela segmentação. O diagrama em árvore da sentença em C24 na Figura 20 abaixo ilustra a ruptura que ocorre nesses casos.

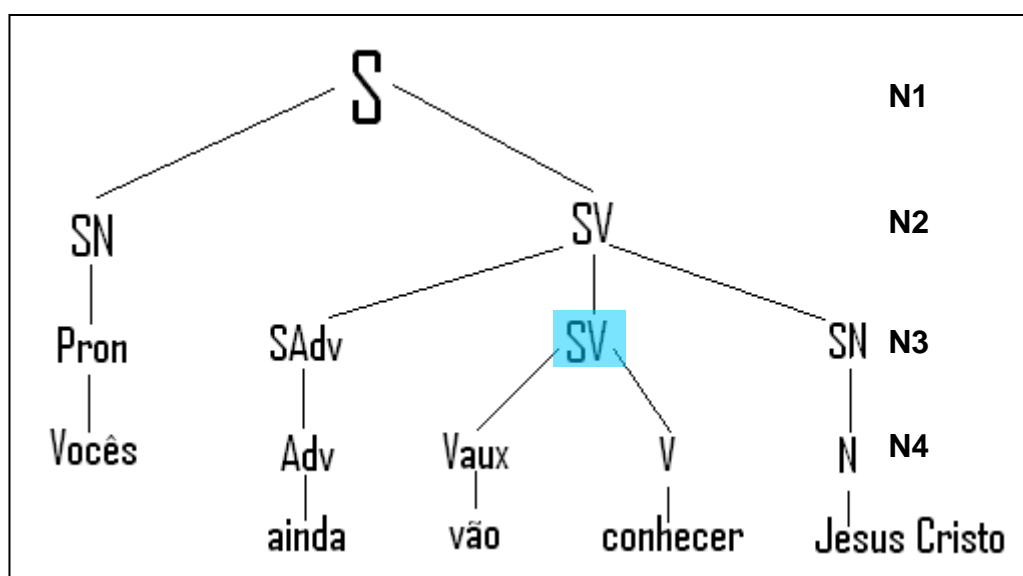


FIGURA 20 – Diagrama em árvore da sentença em C24

Assim como nas sentenças em C25 e C26, em C24 o elemento obrigatório do SV é formado por um verbo auxiliar e um verbo principal localizados no nóculo sintático N3, onde ocorre a quebra de linha que rompe a estrutura do SV realçado na figura. Essa segmentação parece ter sido feita para que a linha superior ficasse mais curta que a inferior. Conforme o diagrama acima, no caso

de C24 a única segmentação que não resultaria em ruptura no SV deveria incidir no nóculo N2. No entanto, essa ressegmentação renderia uma legenda superior muito curta e uma inferior que ultrapassaria o limite de caracteres. Uma possibilidade para C24 seria uma pequena edição - a simplificação da estrutura verbal e a omissão do nome “Cristo”, o qual foi mencionado na legenda anterior – que produziria uma legenda de uma linha, como mostrado no Quadro 25 abaixo.

QUADRO 25
Reformulação de C24

C24R	Vocês ainda conhecerão Jesus!
------	-------------------------------

A reformulação de C25 poderia também ser bem semelhante à do caso anterior, mas envolvendo a simplificação do sintagma nominal como pode ser visualizado no Quadro 26 a seguir.

QUADRO 26
Reformulação de C25

C25R	Traficantes ficam dando dinheiro para os meninos.
------	---

Na legenda sugerida acima, as linhas superior e inferior mantêm-se geométricas, a quebra de linha agora se dá entre verbo e objeto constituintes do SV – a legenda sugerida inclusive poderia ser etiquetada como <SSVD1> - e não dentro do complexo verbal. Esse é mais um dos casos em que não parece ser possível elaborar uma versão sem desvios de segmentação e que a opinião do público alvo informaria a decisão do legendista.

Para reparar a ruptura do SV em C26, a omissão do pronome “Nós” permite a ressegmentação que gera uma legenda de uma linha, como apresentado no Quadro 27 a seguir.

QUADRO 27
Reformulação de C26

C26R	Precisamos sair dessa cidade, Tito.
------	-------------------------------------

Quanto às possíveis formatações que podem ter arrazoado essas ocorrências de segmentação, dos 42 casos marcados com esse desvio, em 14 não é possível determinar o critério estético adotado; há 15 casos que foram segmentados com a linha superior mais curta; e há 13 casos que foram segmentados por geometria precedendo o aspecto linguístico.

Há um caso divergente nos dois subcorpora que é um dos casos cuja segmentação não pode ser esclarecida pelo aspecto estético em LSE_IF, e está apresentado no Quadro 28 abaixo.

QUADRO 28
Legenda <L275> em LSE_IF

C27	Mariana: Padre, eu preciso dizer uma coisa.
-----	---

Na LO_IF a geometria prevalece sobre o aspecto lingüístico nessa mesma sentença. Como mostrado no Quadro 29 abaixo.

QUADRO 29
Legenda <L216> em LO_IF

C27	Padre, eu preciso dizer uma coisa.
-----	------------------------------------

Nessas ocorrências apresentadas anteriormente, o ponto da sentença onde ocorre a segmentação é o mesmo nas duas legendas. No caso da LO_IF a estrutura do SV foi desconsiderada, mas parece haver uma motivação estética para a segmentação. Já no caso da LSE_IF, tanto o aspecto linguístico, quanto estético foram desconsiderados; e, ainda, observa-se que mesmo com a presença da identificação do falante no início da linha de legenda, o legendista optou por não reconsiderar a segmentação. Tanto na LSE_IF quanto na LO_IF a legenda pode facilmente ser reestruturada, omitindo o pronome “eu” e ressegmentando. Para a LSE_IF sugere-se a legenda como apresentado a seguir:

[Mariana] Padre,
preciso dizer uma coisa.

E para a LO_IF sugere-se a legenda como apresentado a seguir:

Padre, preciso dizer uma coisa.

Ambas as sugestões apresentadas anteriormente geram legendas linguisticamente e esteticamente motivadas.

3.2.4.3) SUBCATEGORIA <SSVD3>

Em cada um dos subcorpora há 08 casos dessa etiquetação, sendo que todas as ocorrências incidem nas mesmas sentenças nos dois textos do corpus. Essa etiqueta marca os casos em que o verbo foi separado de um advérbio precedente ou subsequente pela quebra da linha de legenda. Retomando a regra geral segundo Lemle (1989) para o sintagma verbal:

$$RSV \rightarrow (Adv)^* V (Adv)^* \left[\begin{array}{c} SN \\ Adj \end{array} \right] (SPrep)^* (S)^*$$

Foram marcadas as quebras de linha que incidiram assim:

(Adv)* <SSVD3> V <SSVD3> (Adv)*

Como são apenas 08 casos distintos ocorridos no corpus, todas as ocorrências estão apresentadas no Quadro 30 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 30
Exemplos de ocorrências de <SSVD3> em IF

C28	Não quero deixá-lo sozinho, Saulo,
C29	Quem está aqui ao meu lado não é mais Saulo, nosso perseguidor
C30	essa Escritura se realizou plenamente em mim.
C31	Um homem que tanto perseguiu aos nossos irmãos,
C32	as crianças não encontram mais leite no seio de suas mães,
C33	Minha vida já não me pertence e não posso dividir algo que não possuo
C34	mas que se convertam antes às nossas Leis.
C35	Chegada, porém, a fé, não estamos mais sob o julgo do antigo pedagogo.

Nas ocorrências C28, C30, C32, C34 e C35 apresentadas no quadro anterior, a quebra de linha incide entre o verbo e um advérbio subsequente. Em C31 e C33 o verbo está separado de um advérbio precedente. O caso mais interessante nessa subcategoria está em C29, no qual o verbo está entre advérbios – como previsto em Lemle (1989) - sendo que a segmentação ocorreu entre o verbo e o advérbio subsequente, como pode ser visto no diagrama em árvore na Figura 21 abaixo que ilustra esse tipo de segmentação.

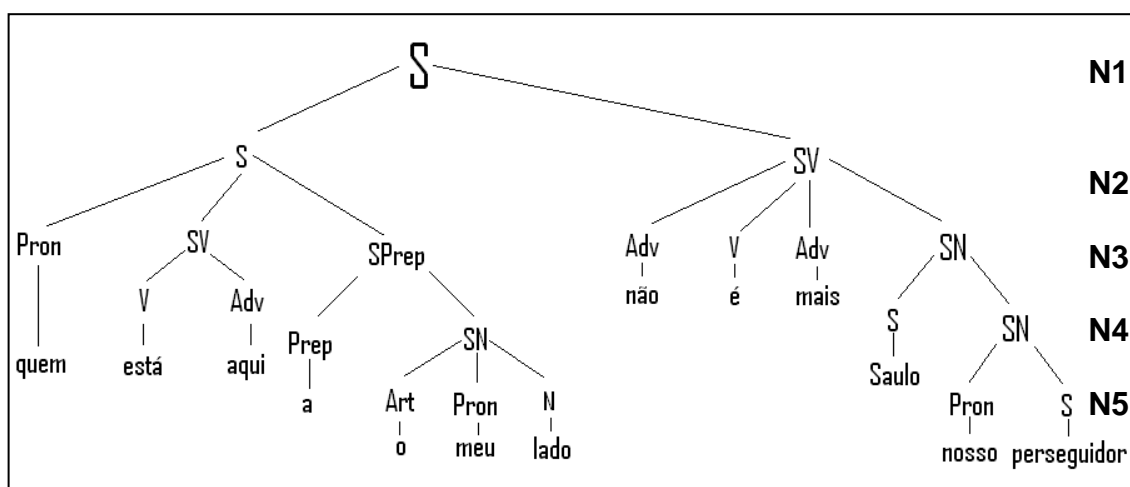


FIGURA 21 – Diagrama em árvore da sentença em C29

A segmentação proposta para a sentença em C29 divide o nóculo sintático N3 em duas partes: na primeira linha da legenda ficam o primeiro advérbio e o verbo, na segunda linha da legenda ficam o segundo advérbio junto do SN complemento que está realizado no nóculo N4. A análise dessa legenda no vídeo revelou que só uma ressegmentação não seria suficiente para adequá-la à modalidade. A sentença é muito longa para o tempo de duração marcado, imprimindo, assim, uma velocidade de leitura acima de 180 palavras por minuto. No entanto, algumas intervenções bastariam para reparar a quebra sintagmática e adaptá-la aos padrões da legendagem, como mostrado no Quadro 31 a seguir.

QUADRO 31
Reformulação de C29

C29	Este não é mais Saulo, o perseguidor.
-----	---------------------------------------

A edição feita nessa legenda foi baseada na observação da imagem. Na cena, o personagem falante se refere a Saulo também com seus gestos, dispensando assim os elementos dêiticos (aqui ao meu lado) redundantes na legenda. Essa análise foi fundamental para a reestruturação linguística e técnica da legenda.

Dos 16 casos de <SSVD3> em IF, em 08 ocorrências não se pode determinar o aspecto estético que poderia ter motivado a segmentação; em 04 ocorrências há preferência pela linha superior mais curta; e em 04 ocorrências a geometria parece ter sido o aspecto que prevaleceu sobre as considerações linguísticas na segmentação.

3.2.5) SEGMENTAÇÃO <SSADJ> E <SSADJD*>

Essa etiquetagem marca ocorrências em que a segmentação da linha de legenda incide em algum ponto entre elementos constituintes de um sintagma adjetivo. A etiqueta <SSAdj> marca ocorrências em que a segmentação é feita entre os blocos constituintes, segundo Lemle (1989); já as etiquetas <SSAdjD*> marcam ocorrências que ocorrem dentro dos blocos constituintes, causando alguma ruptura semântico-estrutural que resulta em sequências não previstas em Lemle (1989), sendo que foram identificados dois tipos diferentes de desvios. Os exemplos [17] e [18] a seguir, retirados das LSE de IF, ilustram um caso de <SSAdj> e um de <SSAdjD*> respectivamente.

[17] <L1102> Não é nada invejoso, <SSADJ> arrogante, orgulhoso. </SSAdj>
</L1102>

[18] <L683> Aqui você será mais útil <SSAdjD1> à nossa causa. </SSAdjD1>
</L683>

Das 2048 linhas de legenda do corpus, 1291 são legendas de duas linhas. Dessas 1291 legendas, 18 estão marcadas com <SSAdj> e 10 estão marcadas com <SSAdjD*> de dois tipos diferentes. Essa categoria de segmentação, portanto, representa aproximadamente 2% das legendas de

duas linhas do corpus, sendo 1,3% <SSAdj> e 0,7% <SSAdjD*>. A distribuição dessa categoria de segmentação no corpus está apresentada na Tabela 7 abaixo em números absolutos

TABELA 7
Distribuição da Segmentação de Sintagmas Adjetivos em IF
em números absolutos

	LO_IF	LSE_IF	CORPUS
<SSADJ>	09	09	18
<SSADJD*>	04	04	08
TOTAL	13	13	26
CATEGORIAS DE <SSADJD>			
<SSADJD1>	03	03	06
<SSADJD2>	01	01	02

A tabela acima inicialmente revela que os dois subcorpora são idênticos quanto aos números da segmentação de sintagmas adjetivos, tanto nas ocorrências dentro do padrão, quanto nas ocorrências marcadas como desvio. Sendo assim, cada um dos subcorpora detém 50% do total de ocorrências dessa categoria de etiquetagem em IF. Contudo, é importante ressaltar que há mais casos de segmentações dentro do padrão (18) que casos de segmentações com desvios (08). A averiguação dos desvios revela que todas as ocorrências são comuns aos subcorpora, isto é, todos os casos de segmentação de sintagmas adjetivos estão nas mesmas sentenças nos dois textos do corpus. O Gráfico 22 a seguir reinterpreta a distribuição das subcategorias de desvios de segmentações do sintagma adjetivo em LO_IF, em valores percentuais.

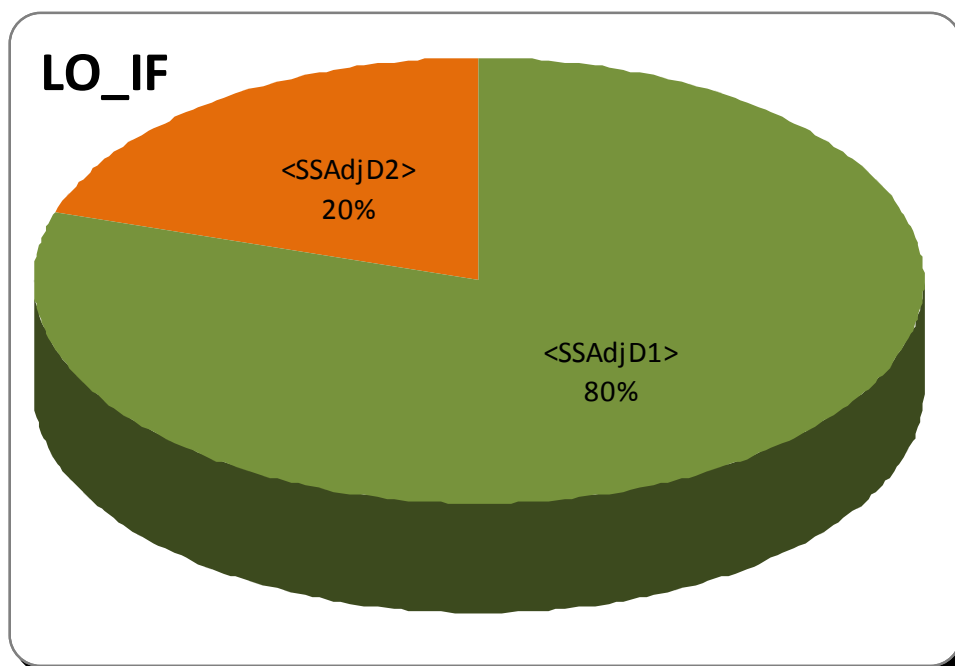


GRÁFICO 22 – Distribuição de <SSAdjD*> em LO_IF em valores percentuais

O Gráfico 23 abaixo reapresenta a distribuição das subcategorias de desvios de segmentações do sintagma adjetivo em LSE_IF, em valores percentuais.

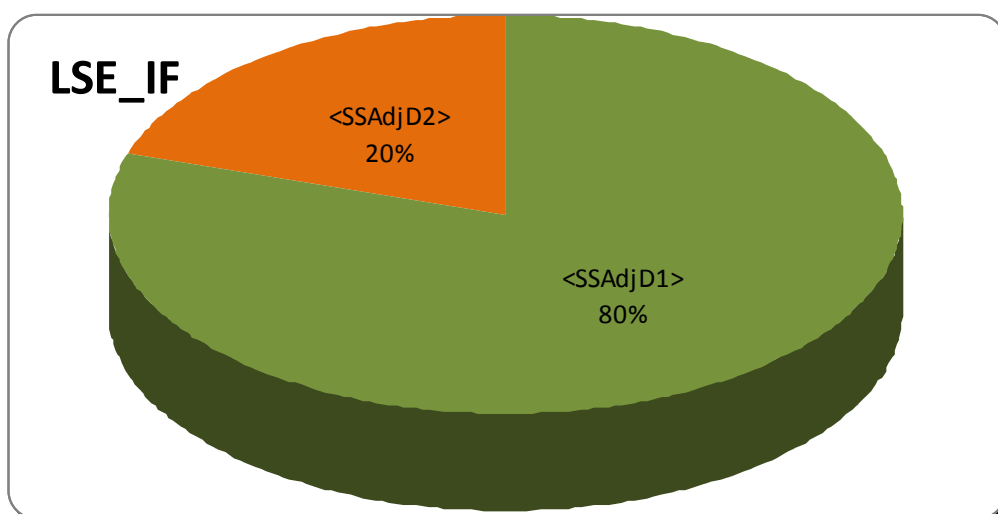


GRÁFICO 23 – Distribuição de <SSAdjD*> em LSE_IF em valores percentuais

Os gráficos acima mostram que a grande maioria – 80% em cada subcorpus – dos desvios de segmentação de sintagmas adjetivos ocorre na separação no

sintagma do adjetivo de um sintagma preposicional constituinte que o complementa (<SSAdjD1>). A separação de adjetivos de advérbios constituintes do SAdj aparece como a outra causa de desvios de segmentação, com um percentual de 17% nos dois subcorpora (<SSAdjD2>).

A seguir as duas categorias de segmentação <SSAdjD> estão descritas e estudadas detalhadamente, a fim de esclarecer onde ocorrem e investigar possíveis razões para as rupturas na estrutura do sintagma adjetivo ao longo dos textos do corpus.

3.2.5.1) SUBCATEGORIA <SSVADJD1>

Em cada um dos subcorpora há 03 casos dessa etiquetagem, sendo que todas as ocorrências incidem nas mesmas sentenças nos dois textos do corpus. Essa etiqueta marca os casos em que o adjetivo núcleo do sintagma foi separado de um sintagma preposicional que o complementa pela quebra da linha de legenda. Retomando a regra geral segundo Lemle (1989) para o sintagma adjetivo:

RSAdj → (Adv) Adj (Adv) (SPrep)* (S)*

Foram marcadas as quebras de linha que incidiram assim:

Adj <SSAdjD3> (SPrep)*

Como são apenas 03 casos distintos ocorridos no corpus, todas as ocorrências estão apresentadas no Quadro 32 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 32
Ocorrências de <SSAdjD1> em IF

C36	Aqui você será mais útil à nossa causa.
C37	Se o senhor escolheu o mais feroz de nossos inimigos,
C38	Nem sou digno de ser chamado de apóstolo

Em todas as ocorrências exibidas no quadro acima o adjetivo núcleo do SAdj está na primeira linha e o sintagma preposicional que o segue está na segunda linha de legenda. A árvore sintática na Figura 22 a seguir exemplifica o tipo de quebra ocorrida nesses casos.

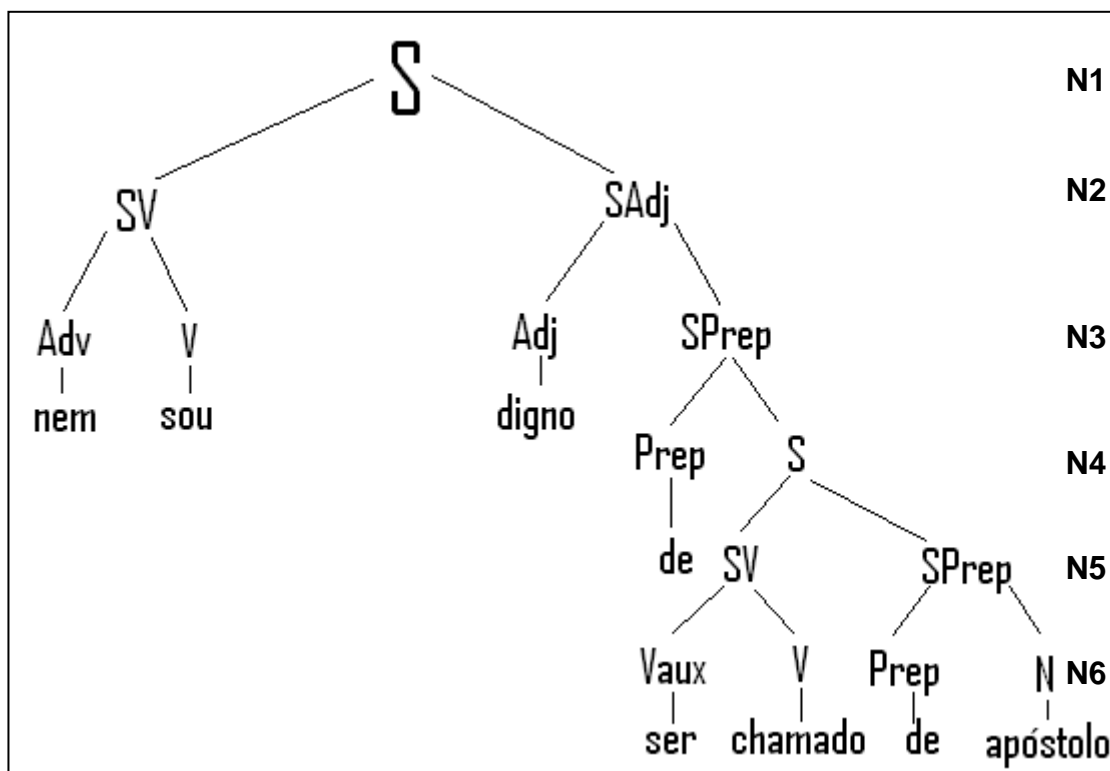


FIGURA 22 – Diagrama em árvore da sentença em C38

Com base no diagrama acima, observa-se que a segmentação na legenda em C38 ocorre no nóculo N3. Nesse nóculo há um adjetivo seguido de um SPrep. Esse SPrep, por sua vez, é uma estrutura complexa formada por uma preposição núcleo que introduz uma sentença não finita. Partindo do princípio da segmentação no nóculo sintático mais alto (KARAMITROGLOU, 1998) a única opção para a segmentação dessa sentença sem quebra dentro do SAdj, seria se essa incidisse em N2 como mostrado no Quadro 33 a seguir.

QUADRO 33
Ressegmentação de C38

C38	Nem sou digno de ser chamado de apóstolo
-----	---

No entanto, essa segmentação resultaria em um desvio no SV, separando o verbo de seu complemento; e geraria uma longa linha de legenda. Aparentemente, esse é um dos casos para o qual não seria possível elaborar uma legenda sem desvios. O mesmo acontece nas outras duas ocorrências, mesmo quando é possível reduzir a legenda com edições mais interventivas.

Em relação ao critério estético que pode ter embasado a decisão do legendista, dos 08 casos em IF, em 02 deles a opção foi pela linha superior mais curta, e em 06 deles não se pode determinar um aspecto estético que tenha determinado a segmentação da legenda.

3.2.5.2) SUBCATEGORIA <SSVADJD2>

Em cada um dos subcorpora há 01 caso dessa etiquetagem, sendo que a ocorrência incide nas mesmas sentenças nas duas legendagens do corpus. Essa etiqueta marca os casos em que o adjetivo foi separado de um advérbio que o modifica pela quebra da linha de legenda. Retomando a regra geral segundo Lemle (1989) para o sintagma adjetivo:

RSAdj → (Adv) Adj (Adv) (SPrep)* (S)*

Foram marcadas as quebras de linha que incidiram assim:

Adv <SSAdjD2> (Adj)*

O caso ocorrido no corpus está apresentado no Quadro 34 abaixo. Para melhor visualização de como aparecem na tela, a ocorrência foi centralizada e teve suas etiquetas retiradas.

QUADRO 34
Ocorrência de <SSAdjD2> em IF

C39	Ela o transformou em algo mais corajoso, implacável e inexorável.
-----	---

Em C39 a quebra de linha mantém na primeira linha de legenda o advérbio que modifica todo o SAdj composto por coordenação que está na segunda linha de legenda. Aparentemente, a motivação dessa segmentação foi a preferência por

linhas de tamanhos iguais, a geometria. Entretanto, essa opção rompe a estrutura do SAdj mostrada na árvore sintática da Figura 23 abaixo.

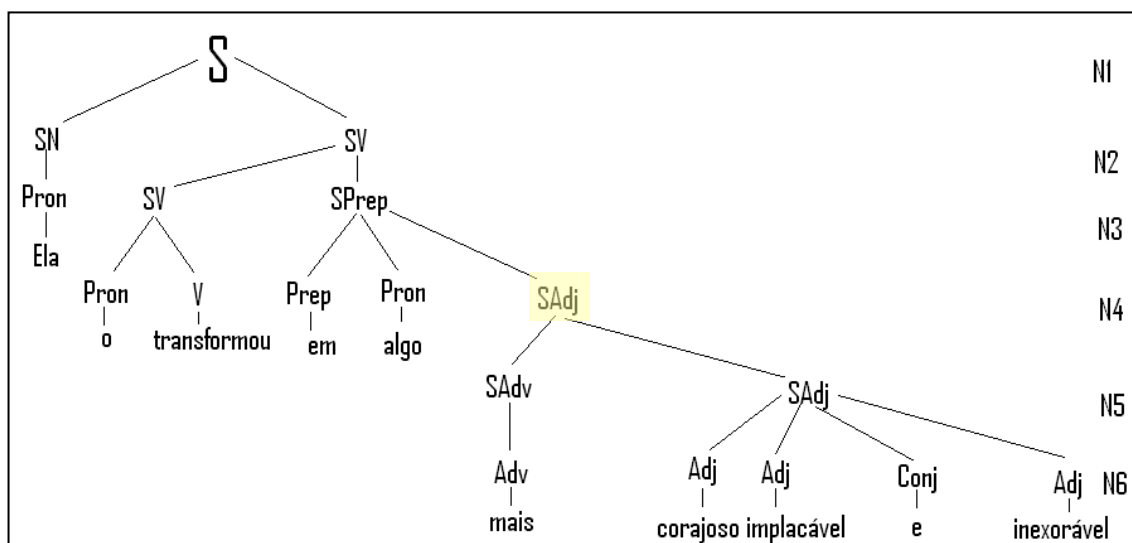


FIGURA 23 – Diagrama em árvore da sentença de C40

A segmentação de C39 ocorre no nódulo N5, rompendo a estrutura do SAdj em destaque realizado no nódulo N4. Entretanto, uma simples ressegmentação que transfira o advérbio “mais” para a linha inferior reintegraria a estrutura do SAdj, reunindo o advérbio aos três adjetivos que modifica. O Quadro 35 abaixo apresenta o resultado da reformulação sugerida.

QUADRO 35
Reformulação de C39

C39R	Ela o transformou em algo mais corajoso, implacável e inexorável.
------	--

Essa reformulação cria também uma legenda esteticamente apropriada, com uma linha superior mais curta que a inferior. Entretanto, para haver reestruturação linguística nesse caso a linha inferior precisou ultrapassar o limite de caracteres padrão adotado na legendagem, sem, contudo prejudicar a legibilidade, pois a legenda continua em uma velocidade de 145 palavras por minuto.

3.2.6) SEGMENTAÇÃO <SSPREP> E <SSPREPD*>

Essa etiquetagem marca ocorrências em que a segmentação da linha de legenda incide em algum ponto entre elementos constituintes de um sintagma preposicional. A etiqueta <SSPrep> marca ocorrências em que a segmentação é feita entre os blocos constituintes, segundo Lemle (1989); já as etiquetas <SSPrepD*> marcam ocorrências que ocorrem dentro dos blocos constituintes, causando alguma ruptura semântico-estrutural que resulta em sequências não previstas em Lemle (1989), sendo que foram identificados dois tipos diferentes de desvios. O exemplo [19] a seguir, retirado da LSE_IF, ilustra um caso de <SSPrepD*>.

[19] <L1097> Somos testemunhas do poder de <SSPrepD3> Cristo. O que mais podemos temer? </SSPrepD3> </L1097>

Das 2048 linhas de legenda do corpus, 1291 são legendas de duas linhas. Dessas 1291 legendas, 02 estão marcadas com <SSPrep> e 29 estão marcadas com <SSPrepD*> de dois tipos diferentes. Essa categoria de segmentação, portanto, representa aproximadamente 2,5% das legendas de duas linhas do corpus, sendo 0% <SSPrep> e 2,5% <SSPrepD*>. A distribuição dessa categoria de segmentação no corpus está apresentada na Tabela 8 a seguir em números absolutos.

TABELA 8
Distribuição da Segmentação de Sintagmas Preposicionais em IF
em números absolutos

	LO_IF	LSE_IF	CORPUS
<SSPREP>	01	01	02
<SSPREPD*>	13	14	29
TOTAL	13	14	27
CATEGORIAS DE <SSPREPD>			
<SSPREPD1>	08	08	16
<SSPREPD2>	05	06	11

A tabela acima inicialmente revela que cada um dos dois subcorpora possui 01 caso de segmentação de sintagmas preposicionais dentro do padrão. Sendo assim, cada um dos subcorpora detém 50% do total <SSPrep> em IF. Esse caso aparece na sentença que se segue:

E todos possam viver em paz
com suas diferenças.

Nesse caso, pode haver uma ambiguidade de interpretação. Uma possível leitura – que motivaria uma etiquetagem de SPrep segmentado com desvio - é de que há um SPrep complexo, composto de dois SPreps, um dependente do outro, sendo que o segundo modifica o primeiro (“em paz” com o quê? “com suas diferenças”). A outra possibilidade é a análise que arrazoou a etiquetagem <SSPrep>, a qual identifica dois adjuntos adverbiais que modificam o verbo “viver” (viver em paz e viver com suas diferenças). A etiqueta atribuída a esse caso, então, apenas foi definida após análise da ocorrência no vídeo. No filme, o falante diz “viver em paz” e após uma pausa no discurso completa “com suas diferenças”. A duração da pausa permitiria até que a linha inferior que contém o trecho “com suas diferenças” se transformasse em uma nova linha de legenda de uma só linha, ao invés de ter-se uma legenda de duas linhas.

Em IF há mais casos de segmentações de sintagmas preposicionais com desvios (27) que casos de segmentação sem desvios. Porém, há uma

pequena diferença quanto às ocorrências marcadas como desvio. A averiguação dos desvios revela que a diferença se encontra na subcategoria <SSPrepD2>. Em LSE_IF há uma ocorrência desse tipo a mais que em LO_IF. Sendo assim, LO_IF detém 48% do total dessas ocorrências, e LSE_IF detém 52% dessas em IF.

O Gráfico 24 a seguir reapresenta a distribuição das subcategorias de desvios de segmentações do sintagma preposicional em LO_IF, em valores percentuais.

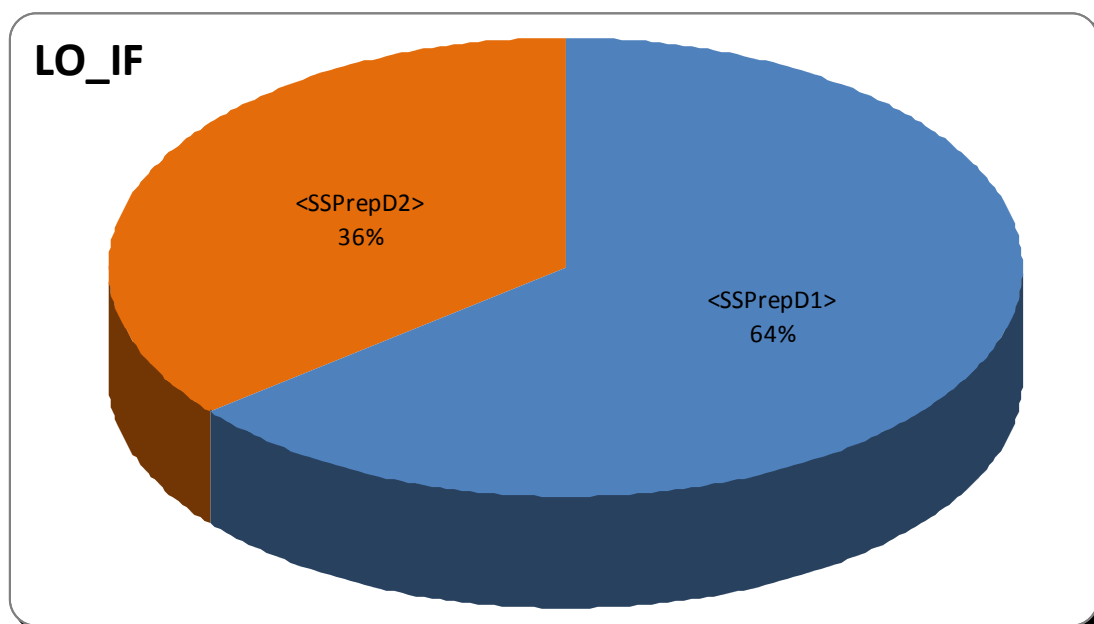


GRÁFICO 24 – Distribuição de <SSPrepD*> em LO_IF em valores percentuais

O gráfico acima mostra que mais da metade – 64% – dos desvios de segmentação de sintagmas preposicionais ocorre na separação de uma locução prepositiva (<SSPrepD1>). A separação de um sintagma preposicional de um substantivo que o segue aparece como segunda causa de desvios de segmentação, com um percentual de 36% (<SSPrepD2>).

O Gráfico 25 a seguir reapresenta a distribuição das subcategorias de desvios de segmentações do sintagma preposicional em LSE_IF, em valores percentuais.

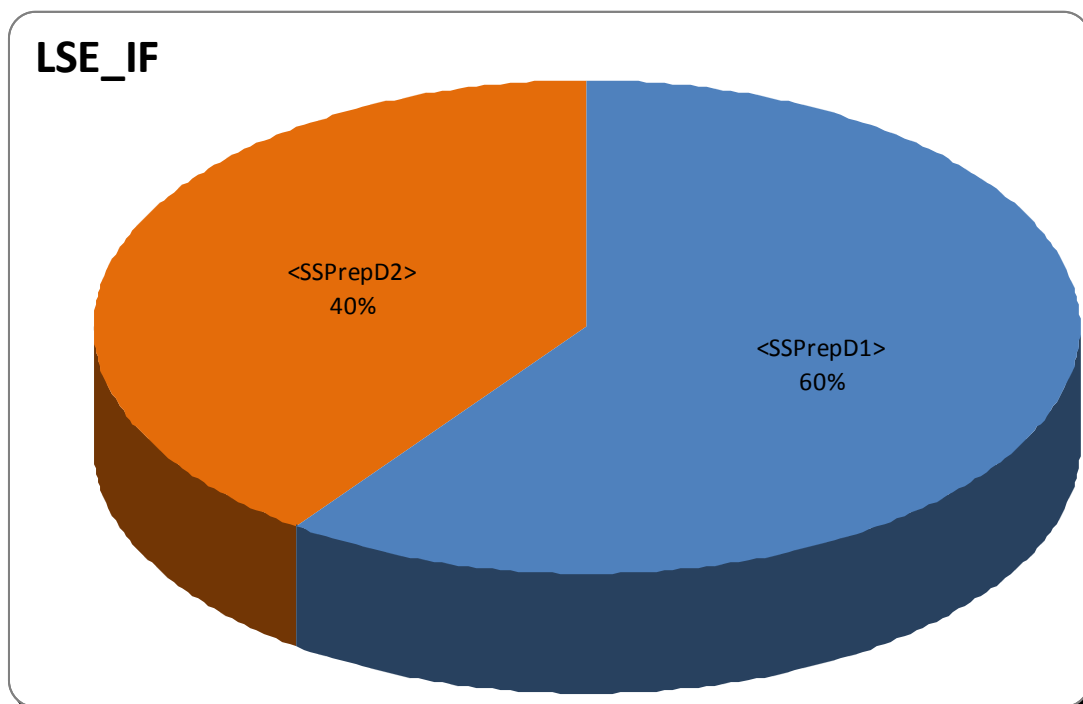


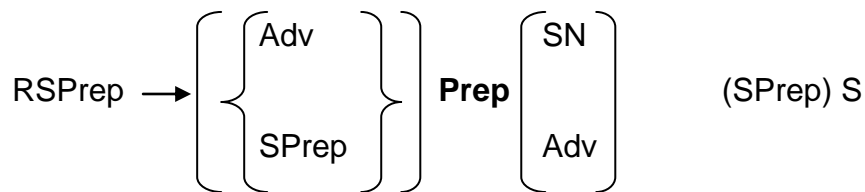
GRÁFICO 25 – Distribuição de <SSPrepD*> em LSE_IF em valores percentuais

O gráfico anterior mostra que mais da metade – 60% – dos desvios de segmentação de sintagmas preposicionais ocorre na separação de uma locução prepositiva (<SSPrepD1>). A separação entre sintagma preposicional e substantivo aparece como segunda causa de desvios de segmentação, com um percentual de 40% (<SSPrepD2>).

A seguir as duas categorias de segmentação <SSPrepD> estão descritas e estudadas detalhadamente, a fim de esclarecer onde ocorrem e investigar possíveis razões para as rupturas na estrutura do sintagma adjetivo ao longo dos textos do corpus.

3.2.6.1) SUBCATEGORIA <SSPREPD1>

Em cada um dos subcorpora há 08 casos dessa etiquetagem, sendo que as ocorrências incidem nas mesmas sentenças nas duas legendagens do corpus. Essa etiqueta marca os casos em que a preposição núcleo do sintagma é composta por mais de uma preposição e essas foram separadas pela quebra de linha. A regra geral segundo Lemle (1989) para o sintagma preposicional é:



Não está representado na regra, mas Lemle (1989) prevê que a preposição obrigatória do SPrep, representada por **Prep**, pode ser composta por uma locução prepositiva. A etiqueta <SSPrepD1> portanto incide dentro da estrutura do núcleo em **Prep**, quando esse é formado por mais de uma preposição.

Como são apenas oito os casos distintos ocorridos no corpus, todos estão apresentados no Quadro 36 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 36
Ocorrências de <SSPrepD1> em IF

C40	Na verdade, a morada Dele é dentro de nós, não nas pedras ou nas lonas.
C41	Meu caminho é para além das fronteiras de Jerusalém.
C42	Minha missão é correr atrás de outras ovelhas.
C43	A gente podia se sentir livre em cima de um jumentinho, não é, Barnabé?
C44	corri perigo por parte de meus irmãos de sangue,
C45	em nome de Nosso Senhor Jesus Cristo.
C46	Mas me converti depois de meu encontro com o Senhor.
C47	Este homem que junto a nós levou a Palavra do Cristo,

Em todas as ocorrências apresentadas acima, a locução prepositiva do SPrep começa na primeira linha de legenda e continua na segunda linha de legenda. O diagrama em árvore da Figura 24 abaixo exemplifica o tipo de ruptura marcada por essa etiquetagem.

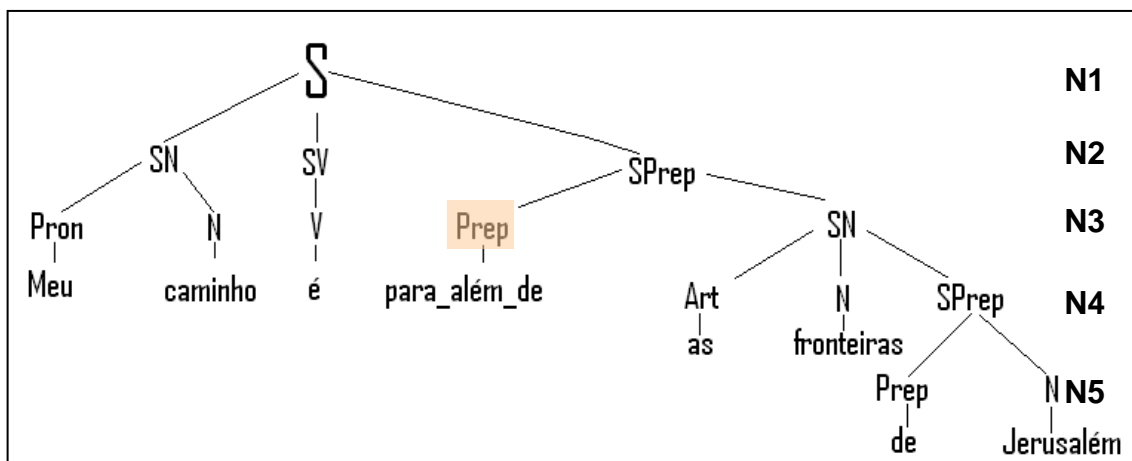


FIGURA 24 – Diagrama em árvore da sentença em C41

O diagrama anterior mostra que a segmentação da sentença em C41 deu-se no nó N3 dentro do nó terminal Prep destacado. A mesma ruptura acontece em outros casos: os elementos constituintes do núcleo do SPrep ficam divididos entre as linhas de legenda. Em todos os casos, um exame do vídeo permitiu que as legendas fossem ressegmentadas (dentro e entre legendas) e editadas para que a estrutura linguística das sentenças fosse restaurada. As reformulações sugeridas estão apresentadas no Quadro 37 a seguir.

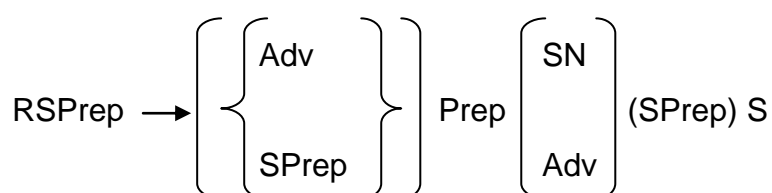
QUADRO 37
Reformulações das ocorrências de <SSPrepD1> em IF

C40R	<L463>	Na verdade, nós somos a casa Dele,	</L463>
	<L464>	não as pedras ou as lonas.	</L464>
C41R	Meu caminho é para além daqui.		
C42R	Minha missão é correr atrás de outras ovelhas.		
C43R	<L773>	A gente podia se sentir livre	</L773>
	<L774>	em cima de um jumentinho, não é, Barnabé?	</L774>
C44R	corri perigo por parte de irmãos de sangue,		
C45R	em nome de Jesus Cristo.		
C46R	Mas me converti depois de encontrar o Senhor.		
C47R	Este homem que junto a nós levou a Palavra de Cristo,		

Das 16 ocorrências desse desvio em IF, foi dada precedência sobre considerações linguísticas ao critério da geometria em 06 delas; em 08 delas ao critério da linha superior mais curta; e em 02 delas nenhum critério estético parece ter sido o motivador da segmentação.

3.2.6.2) SUBCATEGORIA <SSPREPD2>

Nos subcorpora há 05 casos em comum dessa etiquetagem, sendo que as ocorrências incidem nas mesmas sentenças nos dois textos do corpus, e há ainda mais um caso, porém exclusivo de LSE_IF. Essa etiqueta marca os casos em que a preposição aparece separada do SN subsequente pela quebra de linha. A regra geral segundo Lemle (1989) para o sintagma preposicional:



Foram marcadas as quebras de linha que incidiram assim:

Prep <SSPrepD2> SN

Como são apenas 06 os casos distintos ocorridos no corpus, todos estão apresentados no Quadro 38 abaixo. Para melhor visualização de como aparecem na tela, as ocorrências foram centralizadas e tiveram suas etiquetas retiradas.

QUADRO 38
Ocorrências de <SSPrepD2> em IF

C48	Mariana, cuidado com o que você fala. Não seja negativa.
C49	É difícil de acreditar, mas vim com a missão de levá-lo à casa de Pedro.
C50	E devo confessar que estar aqui com vocês é a segunda grande emoção.
C51	Há problema em levar a palavra de Cristo a algum homem na Terra?
C52	E como judeus, estamos aqui na sinagoga para repartir com vocês
C53	Somos testemunhas do poder de Cristo. O que mais podemos temer?

Em todas as ocorrências anteriores a primeira linha de legenda termina com uma preposição cujo SN associado está no início da segunda linha de legenda. Como pode ser visto no quadro acima, 04 ocorrências foram segmentadas por geometria e 02 delas foi segmentada pelo critério da linha superior mais curta. A ocorrência em C51 é exclusiva da LSE_IF, visto que em LO_IF essa aparece com um desvio na segmentação do sintagma nominal (ocorrência C18), conforme apresentado anteriormente na subseção Subcategoria <SSND3>. Na Figura 25 a seguir se apresenta a árvore sintática da sentença em C52, que ilustra a ruptura entre preposição e SN.

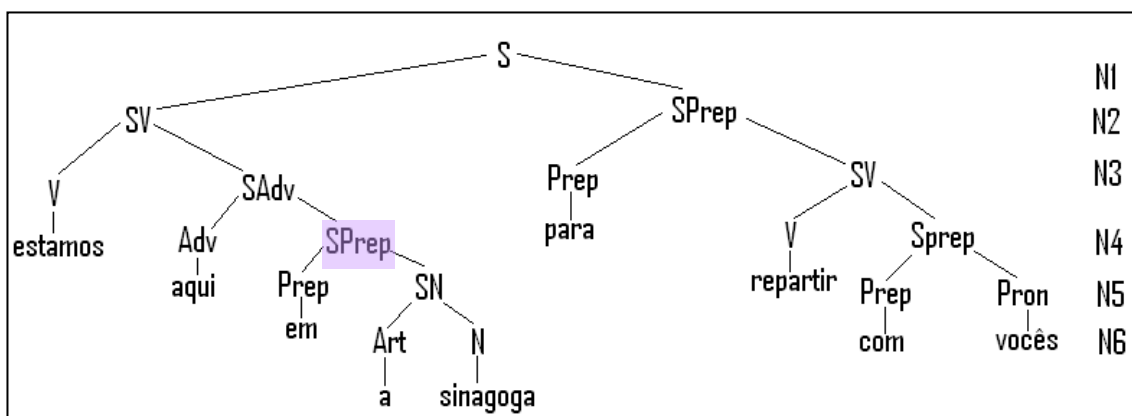


FIGURA 25 – Diagrama em árvore da sentença em C52

A segmentação da sentença em C52 incide, conforme a árvore acima, no nó sintático N5. O diagrama explicita a estrutura do SPrep destacado e expõe como preposição e sintagma nominal, nesse caso, estão fortemente associados, visto que parte do SN (o artigo) de fato se funde à preposição precedente. Nos outros casos de <SSPrepD2> não ocorre essa fusão, mas a ruptura aparece no mesmo ponto da estrutura. Em todos os casos, um exame do vídeo permitiu que as legendas fossem ressegmentadas (dentro e entre legendas) e editadas para que a estrutura linguística das sentenças fosse restaurada. As reformulações sugeridas estão apresentadas no Quadro 39 abaixo.

QUADRO 39
Reformulações das ocorrências de <SSPrepD3> em IF

C48	<L280>	Mariana, cuidado com o quê fala.	</L280>
	<L281>	Não seja negativa.	</L281>
C49	<L613>	É difícil de acreditar,	</L613>
	<L614>	mas vim com a missão de levá-lo à casa de Pedro.	</L614>
C50	<L634>	E devo confessar que estar aqui com vocês	</L634>
	<L635>	é a segunda grande emoção.	</L635>
C51 (cf.C18)	Há problema em levar a palavra de Cristo a algum homem na Terra?		
C52	E como judeus, viemos à sinagoga para repartir com vocês		
C53	<L1100>	Testemunhamos o poder de Cristo.	</L1100>
	<L1101>	O que mais podemos temer?	</L1101>

A próxima subseção apresenta os números gerais dos desvios de segmentação em IF.

3.3) DESVIOS DE SEGMENTAÇÃO EM IF

O número total de segmentações marcadas com desvios em IF é de 351, sendo 175 deles em LO_IF e 176 deles em LSE_IF. Do total de legendas de duas linhas em IF a quantidade de desvios representa aproximadamente 27%. A Tabela 9 abaixo apresenta em números absolutos a quantidade de segmentações etiquetadas com desvios por categoria.

TABELA 9
Desvios de segmentação por categoria

	LO_IF	LSE_IF	IF
<SOSD>	29	28	57
<SOCD>	3	3	06
<SSND*>	46	46	92
<SSVD*>	80	81	161
<SSAdjD*>	04	04	08
<SSPrepD*>	13	14	27

Para facilitar a visualização, o Gráfico 26 a seguir reapresenta os dados da tabela acima, em valores percentuais.

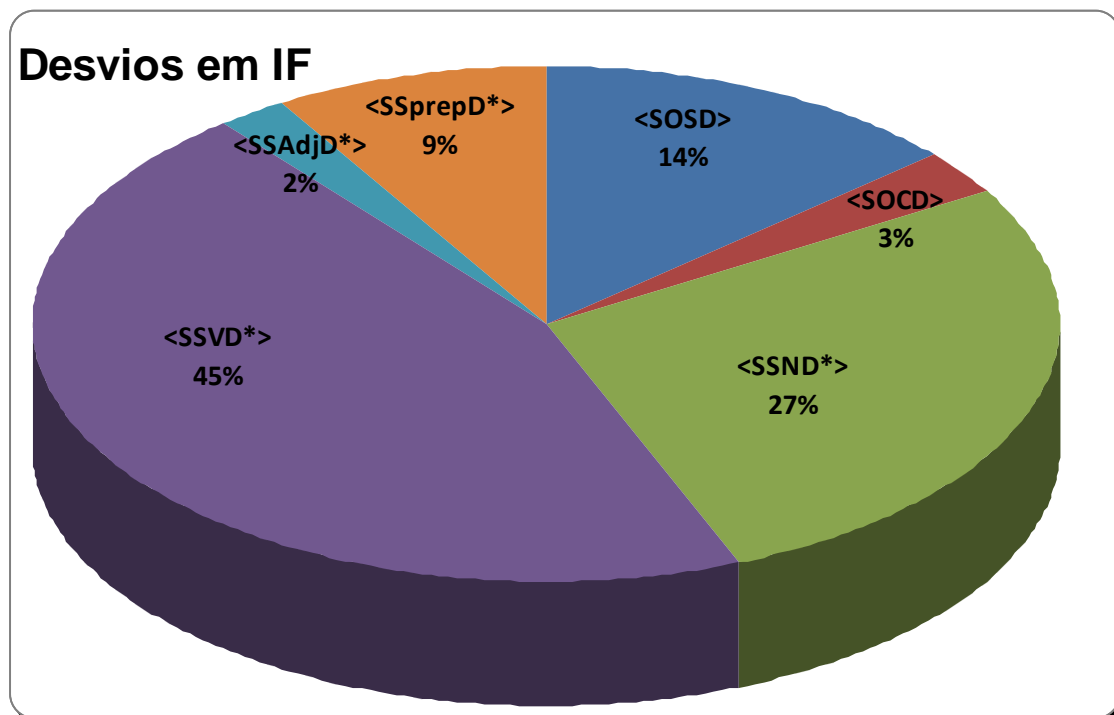


GRÁFICO 26 – Desvios de segmentação distribuídos em IF

Segundo a tabela e o gráfico acima, no sintagma verbal encontra-se o maior número de etiquetas de desvios, quase a metade (45%). A seguir, está a quantidade de desvios etiquetados no sintagma nominal, com 27%. Em terceiro, estão as estruturas subordinadas com 14% dos desvios marcados. Aqueles etiquetados nos sintagmas preposicionais aparecem em quarto representando 9% da quantidade de desvios. A seguir, estão as estruturas coordenadas, com 3%. Por fim, estão os sintagmas adjetivos, com 2% dos desvios encontrados no corpus. Vale destacar que não ocorreram casos de segmentação de sintagmas adverbiais em IF.

Visto que a distribuição da quantidade de desvios nos subcorpora de IF é muito semelhante, com exceção das diferenças de 01 caso de desvio na segmentação do sintagma verbal, e 01 caso de desvio na segmentação de oração subordinada, os valores percentuais da distribuição da quantidade de desvios por subcorpora são idênticos aos valores para o total do corpus, portanto, iguais aos mostrados no gráfico acima.

A Tabela 10 a seguir apresenta em números absolutos a quantidade de segmentações etiquetadas com desvios no nível sintagmático por subcategoria.

Tabela 10
 Desvios de segmentação sintagmática por subcategoria

	LO	LSE	IF
<SSVD1>	51	52	103
<SSVD2>	21	21	42
<SSVD3>	8	8	16
<SSND1>	04	04	08
<SSND2>	09	10	19
<SSND3>	33	32	65
<SSPrepD1>	08	08	16
<SSPrepD2>	05	06	11
<SSAdjD1>	03	03	06
<SSAdjD2>	01	01	02

Para facilitar a visualização, o Gráfico 27 a seguir reapresenta os dados da tabela acima, em valores percentuais.

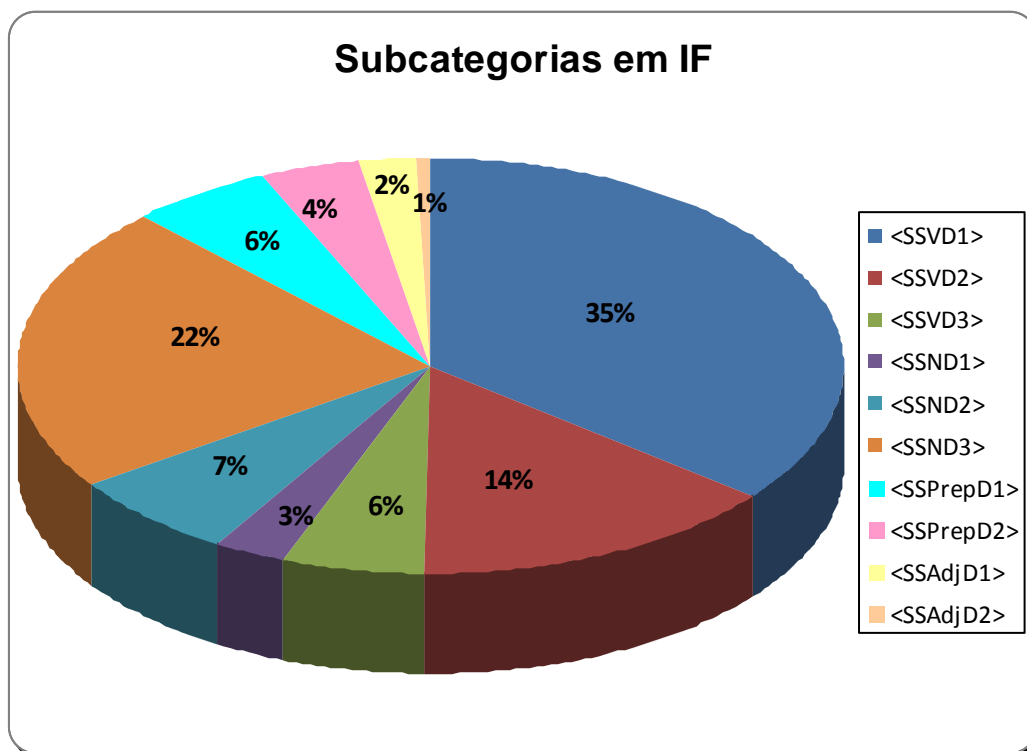


GRÁFICO 27 – Desvios distribuídos em IF por subcategoria

Como mostrado na tabela e no gráfico anteriores, a segmentação entre verbo e objeto ou complemento é o maior número de desvios encontrados, representando 35% em IF (<SSVD1>). Em segundo lugar, está a separação entre o SN e SPrep que o complementa, com 22% (<SSND3>). O verbo auxiliar separado do verbo principal na estrutura do SV está em terceiro lugar, representando 14% dos desvios de IF (<SSVD2>). Com 7%, aparecem em quarto lugar, as ocorrências de segmentação entre determinante/quantificador e SN (<SSND2>). Com 6% para cada etiquetagem, aparecem, em quinto lugar, os casos de segmentação entre verbo e advérbio (<SSVD3>); e os casos de segmentação dentro de locuções prepositivas (<SSPrepD1>). Em sexto lugar, estão as ocorrências de segmentação entre preposição e substantivo (<SSPrepD2>), representando 4% do total de desvios em IF. Com 3% do total dos desvios aparecem em sétimo lugar os casos de segmentação entre substantivo e adjetivo (<SSND1>). Em oitavo lugar com 2% aparece separação entre adjetivo e SPrep (<SSAdjD1>). Finalmente, com 1%, em nono lugar estão os casos de segmentação entre advérbio e adjetivo (<SSAdjD2>).

3.4) DISCUSSÃO DOS RESULTADOS

A primeira pergunta que norteou este estudo promoveu a investigação sobre o fator que motivou a segmentação na LO e na LSE das legendas de duas linhas que continham algum desvio gramatical no filme *Irmãos de Fé*. O exame da segmentação entre linhas adotada em IF buscou identificar os critérios que guiaram a quebra de linha e competiram com a integridade linguística das estruturas, principalmente com a dos sintagmas do PB nas legendas de IF. A partir dos resultados encontrados em Kalantzi (2008), este trabalho buscou identificar se o critério estético da geometria é também preferido na segmentação das legendas – tanto LO quanto LSE - dos filmes, e se esse critério também pode ser considerado o principal fator motivador de desvios linguísticos na quebra de linhas da legendagem estudada aqui.

Kalantzi (2008) identificou em seu corpus de pesquisa quatro fatores que influenciaram a segmentação de linhas de legenda: (i) geometria; (ii) preferência por linha superior mais curta; (iii) preferência por legenda verbatim; (iv) indeterminado. Os resultados revelaram que a geometria foi o principal motivador da maior parte dos desvios linguísticos na segmentação, seja em legendas editadas ou verbatim. Em seguida, a linha superior mais curta foi preferida, principalmente em casos quando uma segmentação linguisticamente apropriada resultaria em uma legenda com aspecto estético oposto. Contudo, nos casos em que esse aspecto colidia com a geometria, Kalantzi (2008) mostra que a geometria teve precedência mesmo quando a linha superior mais curta garantisse uma segmentação adequada. O estudo aponta a preferência por legenda verbatim como o terceiro fator que provocou desvios linguísticos na segmentação de legendas, principalmente em casos em que pequenas edições manteriam a integridade linguística da legenda. Por fim, há uns poucos casos em que nem critérios linguísticos, nem estéticos foram considerados, e, portanto, não se pôde determinar o porquê da segmentação adotada.

Todos os fatores identificados em Kalantzi (2008) foram encontrados nas legendas do corpus da presente pesquisa, porém influenciaram a segmentação em IF de maneira diferente. Uma característica da legendagem em IF – seja LO ou LSE – é a predominante preferência por legenda verbatim: a legenda apresenta na tela uma transcrição de tudo que é falado, com pouca ou

nenhuma edição. Todas as legendas etiquetadas com desvio linguístico de segmentação em IF são legendas verbatim. A distribuição dos outros três fatores recorrentes em IF estão apresentados no Gráfico 28 abaixo em valores percentuais.

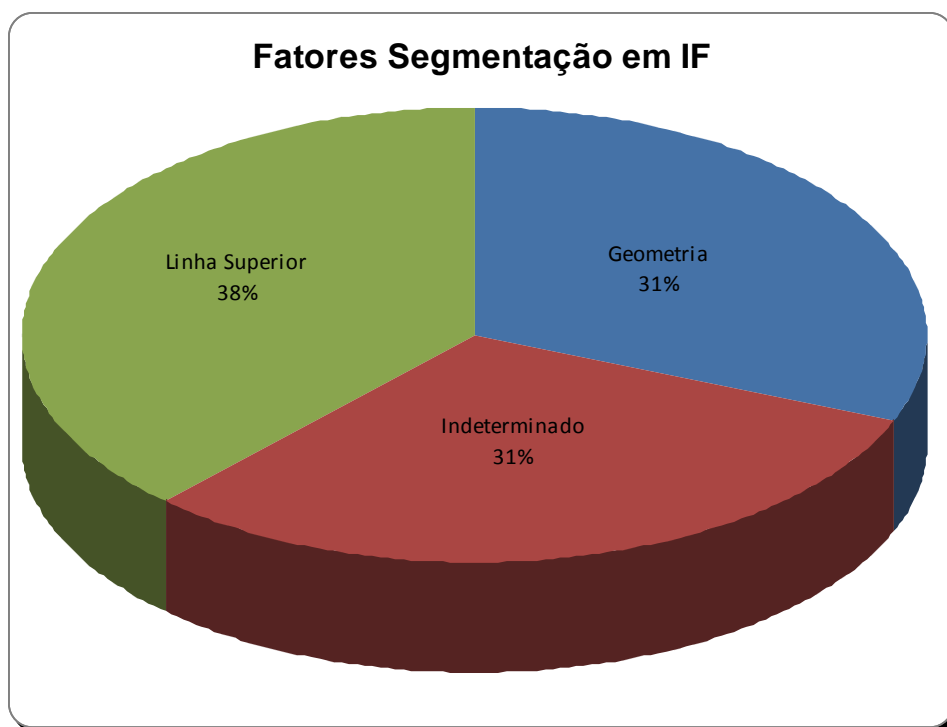


GRÁFICO 28 – Distribuição dos fatores de segmentação em IF

Como se pode observar no gráfico anterior as diferenças entre os critérios de segmentação são bastante pequenas. Diferentemente do que apontaram os resultados em Kalantzi (2008), o fator que mais competiu com considerações linguísticas (38%), influenciando as quebras de linha marcadas com desvio em IF foi a preferência por uma linha superior mais curta. Em Kalantzi (2008) a geometria predomina em mais da metade das legendas com desvio, enquanto em IF a preferência por linhas simétricas é responsável por 31% dos casos de desvios linguísticos na segmentação. Com esse mesmo percentual (31%), há um fator aleatório que não pôde ser determinado, mas influenciou a quebra de linha; isto é, não parece ter havido nenhum critério estético (nem visual, sintático ou retórico) que motivou a quebra de linha. Aparentemente, a preferência por legendagem verbatim acarreta lotação máxima no número de caracteres por linha, e assim o limite de caracteres foi o que ditou a segmentação nesses casos.

A partir da segunda pergunta de pesquisa, buscou-se identificar se a segmentação da LSE apresenta mais desvios que a da LO devido às informações adicionais que devem ser transmitidas na tela. Apesar de incipiente, a literatura da área de legendagem como AVT reconhece desde seus primórdios (De Linde e Kay, 1999) as legendagens para ouvintes e para SE como duas modalidades diferentes, que têm objetivos específicos distintos que as caracterizam e que a observação dos critérios técnicos e textuais peculiares a cada uma delas é determinante para o alcance desses objetivos. Os dados gerais da legendagem e posteriormente os dados específicos da segmentação em IF, no entanto, apontam que as duas modalidades de legendagem estudadas nesta pesquisa foram abordadas da mesma maneira pelo legendista. Isto é, quanto ao aspecto textual, nas duas legendagens houve grande preferência por legenda verbatim, e assim, também a quebra de linha quase não variou entre textos de legenda; quanto ao aspecto técnico, por exemplo, não houve diferença na prática de velocidades de leitura para cada modalidade. Sendo assim, conclui-se que as duas modalidades de legendagem em IF receberam o mesmo tratamento. É verdade que a prática da LSE em IF contou com a inclusão de elementos extralinguísticos fundamentais para o público alvo que a caracterizam como LSE. Para tal, o legendista usou notações próprias, aparentemente por desconhecer as convenções já testadas e aprovadas pelo público SE brasileiro. Os experimentos conduzidos no Brasil apontaram que a identificação de falantes na tela e a representação de sons da trilha sonora deve ser feita entre colchetes; o legendista de IF usou os colchetes para os sons, mas para os falantes, usou dois pontos:

Mariana: Padre, eu preciso
dizer uma coisa

Os resultados desses experimentos também indicaram a preferência do público pelo uso de itálico para marcar vozes de fora da tela; o legendista de IF aplicou o uso de itálico nessas ocasiões, porém o fez da seguinte forma:

[Paulo, em off]
Ainda que eu fale as línguas

Como o uso de itálico já indica voz fora da tela, a indicação “em off” não só é redundante, como também desfavorece a legibilidade. Como revelado em literatura da área, SE raramente aprendem uma terceira língua (FRANCO e

ARAÚJO, 2003), portanto, considerando que o PB já é uma segunda língua para SE, ter que lidar com uma terminologia cinematográfica em inglês complicar a leitura e piorar a recepção da legenda (ARAÚJO, 2004).

Apresentar informações adicionais na tela é uma particularidade que caracteriza a LSE e é, sim, essencial para que a audiência SE aprecie o filme, porém não é o suficiente para adequar o texto da legenda às necessidades do público alvo. No caso de IF, o uso de legendagem verbatim gerou legendas muito rápidas e longas que não endossam a leiturabilidade e enfraquecem o objetivo de promover a acessibilidade crucial à modalidade.

O estudo conduzido em Kalantzi (2008) fundamentou o uso da metodologia de análise de corpus adotado neste trabalho. Entretanto, o corpus investigado em Kalantzi (2008) constou de transcrições e legendas – e não o vídeo em si - dos programas estudados, uma limitação reconhecida na pesquisa que em muitas ocasiões dificultou a interpretação dos dados encontrados. Na presente pesquisa, por sua vez, em muitas ocorrências a etiquetagem, a ressegmentação e a sugestão de uma nova legenda foram baseadas no exame do vídeo e áudio originais, mostrando que a interpretação dos dados revelados pela metodologia de corpus de fato se beneficiou de uma parceria com a possibilidade de analisar o produto audiovisual como um todo. Uma grande contribuição dessa parceria foi apontar para um uso provavelmente desnecessário de legendas de duas linhas em várias ocasiões em IF. Muitas das ocorrências etiquetadas com desvio, além de terem sido causadas pelo uso da legendagem verbatim, parecem ter sido assim realizadas também porque o legendista possivelmente ou não observou a retórica do discurso na cena, que poderia definir a criação de legendas de uma linha (ao invés de legendas de duas linhas); ou então, não observou as informações contidas na imagem que poderiam fundamentar edições de texto para a elaboração de legendas de uma linha. A provável inobservância desses fatores parece explicar uma maior quantidade de desvios linguísticos, que podem desfavorecer a leiturabilidade, mas que podem ser facilmente ajustáveis por meio da ressegmentação ou edição baseadas no exame do vídeo, como mostrado em vários dos exemplos examinados na análise de dados. A identificação desse tipo de recorrência na segmentação da legendagem em IF

salienta a importância da associação da busca por padrões textuais à análise do texto fílmico na interface ETBC e TAV.

CONSIDERAÇÕES FINAIS

Este trabalho afilia-se aos estudos da tradução baseados em corpus em interface com estudos da tradução audiovisual, integrando as pesquisas em TAV desenvolvidas no LETRA-FALE/UFMG em parceria com o Grupo LEAD (Legendagem e Audiodescrição) da Universidade Estadual do Ceará (UECE). Neste contexto, esta pesquisa executou a análise das legendagens para ouvinte e para surdos e ensurdecidos do corpus comparável monolíngue IF, com foco principal nas quebras de linhas de legenda, cujo aspecto linguístico foi descrito e analisado a partir das regras gerais de descrição dos sintagmas do PB de Lemle (1989). Os objetivos gerais desta pesquisa foram contribuir com a consolidação das pesquisas em TAV desenvolvidas no LETRA, direcionando-a para a questão da acessibilidade através da LSE. Os objetivos específicos desta pesquisa foram descrever, analisar e comparar as duas modalidades de legendas que constam do corpus IF. A análise do corpus concentrou-se especificamente no aspecto da segmentação entre linhas de legenda, rastreando, contabilizando e descrevendo os desvios de segmentação localizados entre orações, mas com foco principal nos desvios localizados no nível sintagmático. Este trabalho pretendeu também, com o exame dessas ocorrências, identificar as possíveis causas dos desvios linguísticos ocorridos nos textos. Por fim, visou aferir a quantidade e a qualidade dos desvios encontrados na LSE partindo do pressuposto de que a tomada de decisão sobre segmentação na LSE seria mais complicada devido à necessidade de transmitir informações adicionais, e assim, conteria significativamente mais ocorrências de desvio que a LO.

Para atingir os objetivos propostos, esta pesquisa buscou aporte teórico-metodológico nos estudos descritivos da tradução, mais especificamente nos estudos baseados em corpus em interface com a tradução audiovisual, cujo objeto de pesquisa é a descrição dos vários aspectos textuais e técnicos envolvidos na legendagem com a preocupação de identificar os parâmetros e padrões que têm sido adotados na prática de legendagem, a fim de elaborar modelos de legendagem que realmente promovam acessibilidade e considerem as preferências do público alvo.

Os procedimentos metodológicos aplicados para a obtenção dos dados estudados dividiram-se em etapas. A primeira etapa foi a extração e correção dos arquivos de legenda para que estivessem em formato compatível com as ferramentas eletrônicas de análise do corpus. A etapa seguinte envolveu a anotação manual das ocorrências de quebras de linha no corpus de acordo com as regras descritivas dos sintagmas do PB de Lemle (1989). Com o auxílio das ferramentas eletrônicas, foram extraídos conjuntos de dados quantitativos cuja interpretação forneceu uma análise primária através de uma visão panorâmica das características das legendagens de IF. O rastreamento das etiquetas, por fim, forneceu dados quantitativos exclusivos da segmentação os quais foram investigados à luz de exemplos extraídos das legendas e colocados em diálogo com o referencial teórico adotado.

Os resultados obtidos revelaram que as duas legendagem em IF são praticamente iguais, isto é, apesar de cada modalidade de legendagem ter características distintivas e peculiares que as faz mais adequada à sua audiência alvo, o legendista, aparentemente, tratou-as de maneira igual tanto em termos textuais quanto técnicos. Em relação à segmentação, identificou-se que a estrutura que mais concentrou desvios marcados foi o sintagma verbal. Houve umas poucas sentenças cuja quebra de linha divergiu entre textos, porém, em nenhum dos casos uma particularidade da modalidade – LO ou LSE – parece ter motivado a diferença de decisão. Entre os fatores que influenciaram a segmentação, a preferência por linha superior mais curta é o que mais apareceu precedendo considerações linguísticas nos casos de segmentação com desvio. Por fim, a metodologia de corpus aliada à análise do filme mostrou-se produtiva para o estudo de características específicas do texto de legendagens.

Considerando as perguntas iniciais que motivaram a pesquisa, foi possível respondê-las em parte. No que diz respeito ao critério que predomina na segmentação de legendas de duas linhas marcadas com desvio linguístico em IF, pode-se dizer que, o aspecto que mais competiu com considerações linguísticas foi a preferência estética. Uma vez que desvios sintáticos são passíveis de ocorrer na quebra de linhas, é possível esperar que a maneira como o falante ritma seu discurso (segmentação retórica); ou a importância da informação contida na imagem que a legenda esconderia, ou ainda os cortes

de cena ou movimentos de câmera (segmentação visual) sejam motivos – plausíveis – para ocasional ruptura linguística na legenda. Entretanto, em IF identificou-se apenas uma ocorrência que parece ter sido segmentada por retórica; muitos casos de ruptura linguística só puderam ser justificados por razões estéticas; e ainda, em tantas outras ocorrências nem o fator estético explica a segmentação adotada. No que concerne à possibilidade de identificarem-se significativamente mais desvios linguísticos na LSE_IF devido à necessidade de transmitir informações adicionais, pode-se dizer que a igualdade de tratamento dada às duas modalidades de IF fez com que a investigação dessa pergunta não fosse muito produtiva neste corpus. A LSE de IF mostrou-se, na verdade, como uma LO com a adição dos elementos extralinguísticos. Não houve diferença quantitativa significativa ou qualitativa na quebra de linha em IF que se pode atribuir às peculiaridades das modalidades de legenda envolvidas nesta pesquisa. Essa constatação e a irregularidade na aplicação de critérios de segmentação realçam a necessidade de mais investigações com corpora maiores para que seja possível verificar se há comportamentos distintos em outras legendagens. O resultado de um trabalho investigativo de dimensão maior, por sua vez, poderia contribuir numa formação mais adequada de profissionais da área a partir de bases bem informadas.

Esta pesquisa contribuiu para os estudos da tradução ao utilizar as ferramentas dos ETBC em interface com os estudos de TAV com foco específico na LSE para analisar as legendas do corpus IF com relação ao aspecto linguístico da segmentação entre linhas. Contribuiu também para consolidar as pesquisas direcionadas para a questão da acessibilidade sensorial desenvolvidas no LETRA em cooperação com o LATAV. O estudo apontou que, com algumas ressalvas, a prática da LSE foi semelhante à da LO em IF e que, por isso, não foi possível identificar, e talvez até solucionar, questões concernentes à linguística da segmentação entre linhas na LSE que pudessem emergir devido a peculiaridades da modalidade. Essas questões são fundamentais na busca por estabelecer um modelo brasileiro de LSE por isso é importante que futuras pesquisas da área proponham-se a investigar as especificidades da quebra de linha na LSE, considerando especialmente as preferências e necessidades do público alvo. É essencial reiterar aqui que esta

pesquisa é apenas um estudo de caso, restrito pela falta do tipo de material necessário para uma pesquisa nos moldes propostos aqui e limitado pelo escopo de um estudo de mestrado e, portanto, os resultados alcançados não podem fundamentar generalizações e devem ser revisitados por futuras pesquisas da área.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, V. L. S.; NASCIMENTO, A. K. P. Investigando parâmetros de legendas para Surdos e Ensurdidos no Brasil. In: FROTA, M. P.; MARTINS, M. A. P. (orgs.). **Tradução em Revista**, v. 2, p. 1-18, 2011. Disponível em: <<http://www.maxwell.lambda.ele.puc-rio.br/18862/18862.PDFXXvmi=IHL9lwndfdjuJUEFIVRE5UIWICqnQePAWBji87volQT5MzL8uDNBzvzU90i0f8BcJlvSs3iOTdUHMGPhkETTasPjt8kJhVW2hsVLveBqrw8edZwzDqtNUj8oss8Lo5NHZjt5X5wqCFFQEdqhFU6AfeZJH7u7diPM3jS2ops269SKmj32faCupBe7JKfX36L7pF9P5DRcMLXa61cmcPRCr7KcR1aV7V4wBenHRfTF2RP9T3v7ToJlvtxovJlDGpbi>>. Acesso em: 20 de março de 2012.

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

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

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

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

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

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

BALDRY, Anthony; THIBAUT, Paul. **Multimodal Transcription and Text Analysis**, London e New York: Equinox 2006, 288 p.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri, São Paulo: Manole, 2004, 410 p. ISBN: 8520416764.

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

DE LINDE, Z.; KAY, N. *The Semiotics of Subtitling*. Manchester: St Jerome Publishing, 1999.

DÍAZ-CINTAS, J. Subtitling: the long journey to academic acknowledgment. **The Journal of Specialised Translation**, n. 1, 2004. Disponível em: <http://www.jostrans.org/issue01/art_diaz_cintas.php>. Acesso em: 15 de março de 2012

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

DINIZ, N.S.L.; MAGALHÃES, C.M. **Towards a model of subtitling for the deaf and hard-of-hearing (SDH): A study in multimodal transcription and audiovisual translation**. 2009. 42f. Trabalho de Conclusão de Curso. (Graduação em Letras Bacharelado) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.

DVD Decrypter. Software gratuito. Versão 3.5.4.0. Reino Unido: Lightning UK!, 2005. Disponível em: < <http://www.dvddecrypter.org.uk/>>. Acesso em: 12 mar 2011.

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

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

FORMAT FACTORY. Software gratuito. Versão 2.50. China: Free Time, 2008. Disponível em: < <http://www.pcfreetime.com/>>. Acesso em: 12 mar 2011.

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

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

KALANTZI, D. **Subtitling for the Deaf and Hard of Hearing: A corpus-based methodology for the analysis of subtitles with a focus on segmentation and deletion**. 366f. Tese (Doutorado): School of Languages, Linguistics and Cultures of the University of Manchester, UK, 2008.

KARAMITROGLOU, F. A Proposed Set of Subtitling Standards in Europe. In: **Translation Journal**, v. 2, n. 2, p. 1-15, 1998. Disponível em: <<http://translationjournal.net/journal//04stndrd.htm>> Acesso em: 07 de julho de 2011.

KENNY, D. Corpora in translation studies. In: BAKER, Mona (ed.) **Routledge Encyclopedia of Translation Studies**. London: Routledge, pp. 50-53, 1998.

LAVIOSA, S. **Corpus-based translation studies: theory, findings, applications**. Amsterdam, New York: Rodopi, 2002.

LEMLE, M. *Análise Sintática: (teoria geral e descrição do português)*. 2. Ed. São Paulo: Ática, 1989. 212p. ISBN 8508035535

NEVES, J. **Audiovisual Translation: Subtitling for the Deaf and Hard - of - Hearing**. Tese (Doutorado em Estudos da Tradução), Universidade de Surrey Roehampton, London, UK, 2005.

NEVES, M. H. M. *Gramática de usos do português*. São Paulo: UNESP, 2003, 1037p. ISBN 8571392889

OLOHAN, M. **Introducing corpora in Translation Studies**. London: Routledge, 2004.

PEREGO, E. Evidence of explicitation in subtitling: towards a characterization. In: **Across Languages and Cultures**, A Multidisciplinary Journal for Translation and Interpreting Studies. Budapest: Adadémiai Kiadó, v. 4, n. 1, p. 63-88, 2003. 116

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

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

PEREGO, E.; DEL MISSIER, F.; PORTA, M.; MOSCONI, M. The Cognitive Effectiveness of Subtitle Processing. In: **Media Psychology**. Philadelphia, PA: Routledge, p. 243-272, 2010. ISSN: 1521-3269 print/1532-785X online. Disponível em: <[http://www2 .units. it/delmisfa/ papers/ Subtitles Processing 2010.pdf](http://www2.units.it/delmisfa/papers/Subtitles%20Processing%202010.pdf)>. Acesso em: 23 de março de 2012.

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

SALDANHA, G. Principles of corpus linguistics and their application to translation studies research. In: **L'aplicació dels corpus lingüístics a la traducció**. Revista Tradumàtica: Traducció i Tecnologies de la Informació i la Comunicació, n. 7, dezembro de 2009.

SCOTT, M. *WordSmith Tools version 5*, Liverpool: Lexical Analysis Software, 2008.

SUBTITLE Workshop: sítio do programa disponível em: <www.urusoft.net>. Acesso em: 12 de novembro de 2011.

WURM, S. "Intralingual and interlingual subtitling: a discussion of the mode and edium in film translation". *The Sign Language Translator and Interpreter*, v. 1, n. 1, 2007, 115-141.