

A segmentação linguística na legendagem para surdos e ensurdecidos (LSE) de 'Amor Eterno Amor': uma análise baseada em *corpus*¹

Linguistic segmentation in the subtitling for the deaf and hard-of-hearing (SDH) of the Brazilian TV soap opera *Amor Eterno Amor*: a *corpus*-based analysis

Vera Lúcia Santiago Araújo*
Ítalo Alves Pinto de Assis**

RESUMO: A segmentação na legendagem relaciona-se à divisão dos diálogos de uma produção audiovisual em legendas, a qual pode ser realizada a partir de três critérios: a) visual – pelo corte; b) retórico – pelo fluxo da fala; c) linguístico – pela sintaxe. Os resultados de uma pesquisa realizada pelo grupo LEAD (Legendagem e Audiodescrição) da UECE (ARAÚJO; NASCIMENTO, 2011) sugeriram que uma segmentação linguística adequada pode garantir uma leitura confortável à comunidade surda mesmo quando a legenda possui velocidades de 160 e 180 palavras por minuto. O presente trabalho teve por objetivo analisar e descrever como acontecem os problemas de segmentação linguística na LSE de um capítulo da telenovela 'Amor Eterno Amor'. A metodologia envolveu uma análise descritiva baseada em *corpus*. Os problemas de segmentação foram identificados a partir de etiquetas criadas com base no trabalho de Chaves (CHAVES, 2012; CHAVES; ARAÚJO, 2014). Após o processo de anotação, o *corpus* foi analisado com o auxílio do programa *Wordsmith Tools* 5.0. Os resultados apontaram para uma grande quantidade de problemas de segmentação linguística, em cerca de 25,5% do total de legendas. Os problemas apareceram com mais frequência na ordem dos sintagmas verbal e nominal em legendas de 3 linhas e com velocidade alta.

ABSTRACT: Segmentation in subtitling is related to the division of the dialogues on films and TV programs into subtitles, which can be carried out according to three criteria: a) visual – based on the cuts; b) rhetorical – based on the flow of speech; c) linguistic – on syntax. The results of a piece of research (ARAÚJO; NASCIMENTO, 2011) conducted by the research group called LEAD ('Subtitling and Audiodescription' in Portuguese) at State University of Ceará (UECE) suggested that an appropriate linguistic segmentation can ensure comfortable reading to the deaf community even when the subtitle rate is 160 and 180 words per minute. This study aimed to analyze and describe how linguistic-related segmentation problems occur in the SDH of a Brazilian TV soap opera called *Amor Eterno Amor*. The methodology involved a *corpus*-based descriptive analysis. The segmentation problems were identified through a tag setting created by Chaves (CHAVES, 2012; CHAVES; ARAÚJO, 2014). After the annotation process, the corpus was analyzed by *Wordsmith Tools* 5.0. The results showed a large amount of linguistic-related segmentation problems, in about 25,5% of the subtitles. The problems were more frequent in noun and verb phrases in three-line and fast subtitles.

¹ Este trabalho apresenta resultados parciais do Projeto 'Segmentação na Legendagem para Surdos e Ensurdecidos (LSE): Uma análise baseada em corpus' (Projeto CORSEL), financiado pelo CNPq. Ele também traz os resultados parciais da dissertação em andamento do coautor, oriundos de uma pesquisa de TCC (trabalho de Conclusão de Curso) desenvolvida no ano de 2013 como requisito parcial para obtenção do Grau de Bacharel em Letras/Inglês da Universidade Estadual do Ceará (ASSIS, 2013).

* Universidade Estadual do Ceará (UECE)

** Mestrando Programa de Pós-Graduação em Linguística Aplicada da Universidade Estadual do Ceará (UECE).

PALAVRAS-CHAVE: Tradução Audiovisual. Legendagem para Surdos e Ensurdidos (LSE). Linguística de <i>Corpus</i> . Segmentação Linguística.	KEYWORDS: Audiovisual Translation. Subtitling for the Deaf and Heard-of-Hearing (SDH). Corpus Linguistics. Linguistic Segmentation.
---	--

1. Introdução

A Tradução Audiovisual (doravante também TAV) tem se tornado um campo fecundo quanto à elaboração de recursos para o acesso a produtos audiovisuais por parte de surdos – ensurdidos e pessoas com deficiência visual-PcDV. No caso da comunidade surda, a Legendagem para Surdos e Ensurdidos (LSE) é um dos meios pelos quais seus membros podem desfrutar de novelas, peças de teatro, telejornais, programas de TV, filmes etc. Contudo, pesquisas de recepção realizadas pelo grupo de pesquisa LEAD (Legendagem e Audiodescrição) da Universidade Estadual do Ceará (UECE) (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004, 2007, 2008) com surdos² de todo o Brasil têm demonstrado que as legendas do tipo *closed caption* oferecidas pelas emissoras de TV brasileiras precisam de ajustes, sugerindo que os parâmetros utilizados na confecção dessas legendas não atendem às necessidades da comunidade surda brasileira.

Os resultados obtidos nessas pesquisas (ARAÚJO; NASCIMENTO, 2011; ARAÚJO, 2012), conduzidas com 34 surdos de 4 regiões brasileiras, sugerem que legendas bem segmentadas promovem uma recepção eficaz do programa televisivo legendado mesmo que sejam rápidas (entre 160 e 180 palavras por minuto), o que nos surpreendeu, pois acreditávamos, com base em pesquisas anteriores (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004, 2007, 2008), que o principal entrave para uma recepção eficiente de filmes e programas legendados por parte de surdos estaria na velocidade da legenda. Uma boa segmentação acontece, segundo a literatura da área (DIAZ-CINTAS; REMAEL, 2007; REID, 1990; GOTTLIEB, 1994; KARAMITROGLOU, 1998; IVARSSON; CARROL, 1998), quando as falas de uma produção audiovisual são transformadas em legendas a partir de critérios que levam em conta o corte ou mudança de cena³ (segmentação visual), o fluxo da fala (segmentação retórica) e a sintaxe, a semântica, a lexicogramática, a fonologia-fonética (segmentação linguística).

² Os ensurdidos não foram contemplados na pesquisa, porque acreditamos que uma legenda adequada para surdos contemplaria bastante suas necessidades, já que os surdos teriam menos acesso à língua portuguesa do que os ensurdidos, os quais, diferentemente dos surdos, tem o português como primeira língua.

³ Para fins desse trabalho, os dois termos estão sendo usados como intercambiáveis.

Tendo em vista essas questões, temos como objetivo do trabalho aqui relatado identificar os problemas de segmentação linguística da LSE de um capítulo da telenovela 'Amor Eterno Amor' e analisá-los para descrevê-los, classificando-os. Os outros tipos de segmentação não foram analisados, dado que não se apresentaram como problemáticos em pesquisas anteriores (CHAVES, 2012; CHAVES; ARAÚJO, 2012) envolvendo filmes e telenovelas. Além da identificação e descrição-classificação, pretendemos sugerir possíveis 'ressegmentações' dos problemas encontrados no *corpus* a partir de parâmetros preconizados por pesquisadores da área. A metodologia utilizada foi, em grande parte, uma replicação daquela adotada na pesquisa de Chaves (CHAVES, 2012; CHAVES; ARAÚJO, 2014) sobre os problemas de segmentação na LSE do filme brasileiro 'Nosso Lar'. Assim como na referida pesquisa, a identificação, análise e descrição-classificação dos problemas na segmentação linguística de 'Amor Eterno Amor' foi realizada com o auxílio de anotação por etiquetas criadas pela autora. Após a etiquetagem, o *corpus* anotado foi submetido ao *Wordsmith Tools* 5.0.

Este artigo está dividido nas seguintes seções além desta introdução: na Seção 2, apresentamos uma breve revisão de trabalhos relacionados à legendagem e sua segmentação e à legendagem em interface com a Linguística de Corpus; na Seção 3, dedicada à metodologia, discorremos sobre os processos de extração do *corpus* e sua anotação, dentre outras etapas; na Seção 4, apresentamos e discutimos os resultados da análise do *corpus* e, por fim, na Seção 5, traçamos algumas considerações finais acerca deles.

2. Legendagem

2.1 A Legendagem *Closed Caption*

O sistema norte-americano de legendagem, mais conhecido como *closed caption*, é um tipo de LSE, no qual a linha de legenda, para ser visualizada pelo espectador, necessita ser transformada em códigos eletrônicos e inserida na linha 21 do intervalo vertical em branco da TV, que é uma barra horizontal posta entre as imagens (ARAÚJO, 2008, p. 62). O sistema brasileiro, na maioria das vezes, não edita a fala, constituindo-se quase numa transcrição e, portanto, não atendendo à maioria das especificações técnicas esperadas numa legendagem voltada tanto para surdos e ensurdecidos como para ouvintes (ARAÚJO, 2008; ARAÚJO; NASCIMENTO, 2011). Esta abordagem difere-se da europeia no que diz respeito à confecção das legendas para surdos e ensurdecidos. Os europeus adotam os mesmos parâmetros para

ambos os tipos de legendagem, com exceção da inclusão de informações adicionais que não podem ser recuperadas pela audição como a identificação de falantes e de efeitos sonoros.

Aqui no Brasil, os procedimentos são diferentes dependendo do tipo de legendagem. Para os ouvintes, nos quais são legendadas produções audiovisuais estrangeiras, a legenda segue os padrões europeus, ou seja, tem no máximo duas linhas e velocidades de 145, 160 e 180 palavras por minuto (DIAZ CINTAS; REMAEL, 2007:96). Quando a velocidade da fala é maior do que 180 palavras por minuto, é necessário uma edição dessa fala por meio da condensação, redução (uso de palavras mais curtas) ou omissão de palavras quando seu sentido pode ser apreendido pelas imagens. Para os surdos e ensurdecidos, produz-se uma tipo de legendagem semelhante ao norte-americano utilizado somente para traduções intralinguísticas, já que o sistema foi criado exclusivamente para atender à comunidade surda dos Estados Unidos. A LSE produzida para os canais de TV brasileiros são também intralinguísticas, visto que são produzidas a partir das versões dubladas de filmes e programas de TV em língua estrangeira. A pesquisa realizada na UECE não concorda com essa diferença de procedimentos, aproximando-se mais da concepção europeia de legendagem.

A LSE do tipo *Closed Caption* pode aparecer em dois formatos: como *pop-on* ou como *roll-up*. A legendagem *roll-up* é destinada a programas ao vivo, podendo ser desenvolvida por estenotipia ou por refalamento (*respeaking*). No primeiro caso, por estenotipia, é produzida por meio de um teclado de digitação rápida chamado de estenótipo. No segundo caso, por refalamento, é produzida a partir de um software de reconhecimento de voz, no qual o legendista repete o que está sendo dito na tela para que as legendas possam ser visualizadas. As legendas do tipo *roll-up* aparecem na tela palavra por palavra da esquerda para a direita. Quando a linha de legenda se completa ela desloca-se para cima e permanece na tela até que a linha de baixo esteja também completa. A legendagem *pop-on*, por sua vez, também feita por meio do estenótipo, aparece na tela em bloco e sai em bloco. É semelhante às legendas exibidas tanto na TV quanto no cinema, as quais já estamos acostumados. É reservada para programas pré-gravados, como novelas e filmes. Este foi o formato analisado no presente trabalho, já que, supostamente, por ser editável, possibilitaria ao legendista a elaboração de LSE de acordo com os preceitos da literatura da área, especialmente no caso do parâmetro da segmentação, que, como já dito, foi o foco deste trabalho.

2.2 A Segmentação em Legendagem

A segmentação em legendagem diz respeito à divisão da fala traduzida em porções de texto escrito na parte inferior da tela. A segmentação, tal como indicamos anteriormente, pode ocorrer de três formas: pautada pela linguística (sintaxe, semântica, lexicogramática e fonologia-fonética), pela retórica (fluxo da fala) e pelo visual (corte ou mudança de cena) (DIAZ-CINTAS; REMAEL, 2007; REID, 1990; GOTTLIEB, 1994; KARAMITROGLOU, 1998; IVARSSON; CARROL, 1998). A segmentação linguística⁴ pode acontecer de duas formas: 1) segmentação entre linhas, a qual ocorre na divisão das falas de uma produção audiovisual em diferentes legendas; ou 2) segmentação quebra de linha (*line break*), a qual ocorre dentro da mesma legenda, envolvendo a divisão da informação entre as duas linhas dessa legenda (DIAZ CINTAS; REMAEL, 2007, p. 173). O foco aqui recairá sobre o segundo tipo de segmentação linguística.

Muitos teóricos da TAV enfatizam a importância de uma segmentação linguística dentro dos critérios supracitados para a recepção eficaz de um filme ou programa legendado. Diaz Cintas e Remael (íbidem, p. 172) alertam que “uma segmentação cuidadosa da informação pode ajudar a reforçar a coerência e a coesão na legendagem”⁵. Do ponto de vista da segmentação quebra de linha, o foco deste estudo, as legendas devem ser divididas de uma forma que sejam autossuficientes sintática e semanticamente, já que “quando segmentamos uma sentença, forçamos o cérebro a pausar o seu processamento linguístico por um momento até que os olhos captem a próxima informação”⁶ (KARAMITROGLOU, 1998, p. 10), sendo, assim, necessário que essa segmentação siga determinadas diretrizes linguísticas. Karamitroglou (1998) considera que o processo deve se dar no mais alto nível sintático possível. Para tal, o mesmo faz uso de nódulos sintáticos oriundos de uma descrição sintática em árvore de base gerativista, que apresentamos na Figura 1.

⁴ Reid (1990) utilizou o termo 'gramatical' ao invés de 'linguística'. Para além da questão teórica problematizada por Chaves (2012, p; 45), a qual sugeriu que a escolha por "linguística" deveu-se ao fato de que a palavra 'gramatical' está “arraigada de preconceito e por isso acaba conduzindo à discussão por um viés normativo da língua”, preferimos o termo 'linguística' por o mesmo ser mais abrangente, porque a segmentação não envolve somente a gramática, ou seja o léxico e a sintaxe, mas também a semântica e fonologia-fonética.

⁵ Tradução dos autores. Original: "A careful segmentation of the information can help reinforce coherence and cohesion in subtitling".

⁶ Tradução dos autores. Original: "When we segment a sentence, we force the brain to pause its linguistic processing for a while, until the eyes trace the next piece of linguistic information."

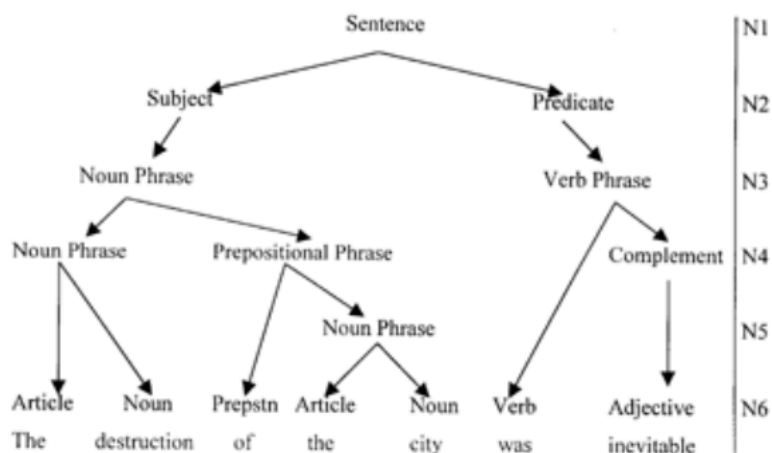


Figura 1 – Árvore sintática usada por Karamitroglou (1998) para explicar a segmentação linguística.
Fonte: Karamitroglou (1998, p. 9).

O autor divide a oração *the destruction of the city was inevitable* em seis nódulos: no primeiro, N1, está a oração em si, constituída de sujeito e predicado (N2), os quais são formados de sintagmas nominais e verbais respectivamente (N3). O sujeito *the destruction of the city* é constituído de dois sintagmas, um nominal (*the destruction*) e o outro preposicional (*of the city*), enquanto o predicado tem um complemento (N4). O sintagma preposicional também tem entre seus constituintes um sintagma nominal (*the city*, N5). Por fim, temos o léxico e a função dentro da estrutura gramatical (N6). Para o autor, se a oração em questão estivesse em uma legenda, ela deveria ser segmentada, idealmente, privilegiando as unidades *The destruction of the city* (sujeito) na linha de cima e *was inevitable* (predicado) na linha de baixo. Dessa maneira, a quebra aconteceria no mais alto nódulo sintático a partir do qual a segmentação de uma oração é possível, o N2, e, quando isso ocorre, “[...] maior é o agrupamento de carga semântica e mais completa é a informação a ser apresentada ao cérebro” (ibidem, p. 9)⁷, facilitando, assim, a compreensão da legenda como um todo.

Um pequeno adendo a fazer é a respeito da distribuição das legendas na segmentação quebra de linha. Ao invés de utilizar os critérios de segmentação pelos critérios discutidos anteriormente, alguns legendistas preferem segmentar pelo número de caracteres, produzindo três formatos de legendagem. No primeiro, aparece quase o mesmo número de caracteres para as duas linhas, tendo um formato semelhante ao de um retângulo. No segundo e no

⁷ Tradução dos autores. Original: "This occurs because the higher the node, the greater the grouping of the semantic load and the more complete the piece of information presented to the brain."

terceiro, aparecem mais caracteres na linha de cima ou na de baixo, respectivamente, lembrando um triângulo. O Quadro 1 mostra os três formatos:

Quadro 1: Formato das legendas na tela.

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

Fonte: Os autores.

Karamitroglou (ibidem, p. 10) acredita que o ideal seria que a legenda de duas linhas, a qual pode ter até 39 caracteres em cada linha, tivesse um formato retangular, ao invés do triangular, por ser aquele ao qual os telespectadores estão mais acostumados. Contudo, quando as questões linguísticas não permitirem esse formato, os triangulares podem ser acessados. No que diz respeito ao capítulo de 'Amor Eterno Amor' em foco, quando foram propostas novas legendas conforme as diretrizes dos pesquisadores da área para a segmentação quebra de linha (DIAZ-CINTAS; REMAEL, 2007; REID, 1990; GOTTLIEB, 1994; KARAMITROGLOU, 1998; IVARSSON; CARROL, 1998), o formato em forma de triângulo prevaleceu.

Por questões de espaço e, como foi dito anteriormente, dado que os outros tipos de segmentação não se apresentaram como problemáticos em pesquisas anteriores (CHAVES, 2012; CHAVES; ARAÚJO, 2014) envolvendo filmes e telenovelas, decidimos analisar neste trabalho apenas a segmentação quebra de linha, a qual será chamada aqui também de segmentação linguística.

2.3 Legendagem e Linguística de *Corpus*

O uso da Linguística de *Corpus* como metodologia nos Estudos da Tradução começou a ser preconizado por Baker (1993). Na Tradução Audiovisual, por sua vez, a análise baseada em *corpus* vem ganhando cada vez mais espaço entre os pesquisadores, sem ser ainda, contudo, grande o arcabouço teórico formado pela interface entre essas duas áreas disciplinares. A partir dos objetivos do presente trabalho, iremos detalhar, a seguir, as duas

pesquisas que embasaram nosso estudo. São, a saber: Perego (2008) e Chaves (2012; CHAVES; ARAÚJO, 2014).

Por meio da análise manual de um *corpus* de filmes, Perego (2008) examinou casos de má segmentação na legendagem interlinguística para ouvintes. Em seu trabalho, a autora investigou a segmentação linguística de um *corpus* heterogêneo de legendas de filmes para DVD e para cinema. A autora definiu categorias para a análise dessa má segmentação como quebra do sintagma nominal, quebra do sintagma preposicionado, quebra do sintagma verbal e quebra da oração complexa em seus constituintes, as orações coordenadas e subordinadas. Mesmo sem a utilização de uma análise baseada em *corpus*, como nos propusemos a fazer em nosso estudo, foi o trabalho de Perego que motivou e serviu como ponto de partida para a produção de etiquetas de Chaves (2012; CHAVES; ARAÚJO, 2014) para a análise de aspectos técnicos (número de linhas e caracteres, velocidade, tempo inicial e final da legenda) e, principalmente, de problemas de segmentação na LSE do filme brasileiro 'Nosso Lar'.

Na questão relativa à análise da segmentação linguística, Chaves (ibidem) utilizou os pressupostos de Karamitroglou (1998) sobre a importância de ela ocorrer no mais alto nível sintático possível. Para as definições de oração e dos vários tipos de sintagmas na sintaxe do português brasileiro, necessárias para a análise dos problemas de segmentação linguística, a autora fez uso da abordagem descritiva de orientação formalista de Perini (2010). Chaves (2012; CHAVES; ARAÚJO, 2014) formulou 3 etiquetas indicativas de problemas de segmentação (linguística, retórica e visual) e 8 para análise de parâmetros técnicos da LSE (número de linhas e caracteres, velocidade, tempo inicial e final da legenda). A partir dos conceitos, nomenclaturas e observações de Perini (2010) sobre a gramática do português brasileiro, a autora elencou um número de 19 subetiquetas referentes a categorias de problemas de segmentação linguística. Os resultados apontaram para uma maior quantidade de problemas de segmentação linguística no nível do sintagma verbal e nominal (41 e 26%, respectivamente), majoritariamente na quebra verbo+verbo, e em legendas com velocidade considerada alta (160 e 180 palavras por minuto). Abaixo, o Quadro 2 referente aos problemas de segmentação identificados por Chaves.

Quadro 2⁸ – Quadro de etiquetas de problemas de segmentação de Chaves.

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

Fonte: Chaves (2012, p. 58)

Usando como ponto de partida as etiquetas da autora e valendo-se do pressuposto teórico de que “[...] mesmo a ação mais simples imaginável, a de contar palavras ou identificar a pontuação pressupõe uma teoria linguística” (GREFENSTETTE; TAPANAINEN, 1994; NUNBERG, 1990, *apud* SANTOS, p. 58) o mais coerente possível com o objetivo e o desenho metodológico no âmbito do Projeto CORSEL – este será mais detalhado na seção 3.1 – como um todo, passamos a embasar a análise da segmentação linguística a partir de um viés funcionalista, já que essa vertente dos estudos linguísticos possui uma maior aproximação epistemológica com a metodologia baseada em *corpus*. Para tal, a ‘Nova Gramática do Português Brasileiro’ (CASTILHO, 2012) – pela grande riqueza de

⁸ Para mais detalhes sobre as diferenças entre as nossas etiquetas e as de Chaves, ver Assis (2013).

detalhes desta no que se refere à descrição dos sintagmas e das sentenças⁹ complexas¹⁰ da língua portuguesa brasileira, sendo menos sintética¹¹ que a descrição de Perini (2010) – foi adotada. Para Castilho (p. 249), “a estrutura sintática da sentença fundamenta-se nos arranjos lexicais de que ela é formada, os sintagmas, bem como nas funções que decorrem do relacionamento desses sintagmas”, sendo estes um somatório de constituintes que apresentam um lugar previsível dentro na estrutura da sentença. Ainda segundo o autor, o sintagma tem uma estrutura composta por um NÚCLEO, uma margem esquerda, preenchida ou não pelos chamados ESPECIFICADORES do sintagma, e por uma margem direita, preenchida ou não pelos ditos COMPLEMENTADORES. No Quadro 3, mostramos os vários tipos de sintagma reconhecidos por Castilho e a estrutura geral de cada um bem como a estrutura geral dos dois tipos de sentença.

Quadro 3: Estrutura geral dos sintagmas e sentenças complexas (CASTILHO, 2012).

Sintagma Verbal (SV)	O SV é a construção que tem como núcleo o verbo. Como a sentença é um verbo que articula seus argumentos, a única diferença entre ele e uma sentença é que no SV não figura o sujeito.
Sintagma Nominal (SN)	O SN é uma construção cuja estrutura tem por núcleo um substantivo ou um pronome, tendo por Especificador o artigo e os pronomes e por Complementadores os SAdjs e os SPs.
Sintagma Adjetivo (SAdj)	O SAdj tem por Núcleo o adjetivo, que é uma classe basicamente predicadora, funcionando como adjunto adnominal, enquanto constituinte do sintagma nominal, ou como predicativo, enquanto constituinte do sintagma verbal (CASTILHO, 2012, p. 516). O SAdj tem como Especificador advérbios predicativos qualificadores e como Complementador SPs e sentenças substantivas objetivas.

⁹ Segundo o autor, a “designação da sentença não é pacífica na literatura” (p. 58), ao passo que podemos encontrar “[...] termos tais como oração, frase, período (conjunto de orações) etc.” (ibidem). No Glossário de sua obra, Castilho, define o vocábulo ‘sentença’ da seguinte forma: “1. Sentença ou oração é a unidade da sintaxe estruturada por um verbo que seleciona o seu sujeito e seus complementos. Os adjuntos também integram uma sentença, mas não são selecionados pelo verbo [...]” (p. 691). Ao que nos parece, o autor considera os dois termos intercambiáveis, apesar de preferência pela utilização de ‘sentença’. Esta pressuposição confirma-se pelo fato de que no vocábulo ‘oração’, há uma remissão ao vocábulo ‘sentença’ expressa por ‘Veja **Sentença**’ (p. 686). Em nossa análise preferimos utilizar o termo ‘oração’, pois, se ‘sentença’ e ‘oração’ são intercambiáveis, a utilização deste último se torna menos obscura. Inclusive, o termo ‘sentença’ nem mesmo consta ‘Nomenclatura Gramatical Brasileira’, ao contrário de ‘oração’.

¹⁰ No capítulo intitulado ‘A Sentença Complexa e sua Tipologia’, Castilho (2012) afirma que o termo ‘sentença complexa’ é preferido no lugar de ‘período’, dado que este não apresenta uma unidade sintática diferente da sentença simples, “ou seja, tudo o que ocorre numa sentença simples ocorre numa sentença complexa”(p. 336). Dado que não há uma diferença além de terminológica em relação a esses dois termos, iremos utilizar em nossa análise os termos ‘oração coordenada’ e ‘oração subordinadas’ (períodos compostos) pelos mesmo motivo supracitado.

¹¹ Por ser menos sintética, ela nos permitiu compreender de forma mais substancial os problemas de segmentação linguística.

Sintagma Preposicional (SP)	O SP tem por núcleo a preposição, por Especificador o advérbio, enquanto que o Complementador em um SP também pode ser: (i) um verbo, como em ‘para comer’, (ii) um pronome, como em ‘para mim’, ‘para quem’; (iii) um quantificador definido, como em ‘para dois’; (iv) um quantificador indefinido, como em ‘para muitos’, dentre outras possibilidades de realização.
Sintagma Adverbial (SAdv)	O SAdv tem como núcleo o advérbio e como Especificador e Complementador outros advérbios. Sintaticamente, Castilho (2012) diz que os advérbios mantêm relação entre si, com e com os adjetivos verbos, adjetivos.
Sentenças Complexas	Enunciados que possuem mais de um verbo, ou seja, contém mais de uma sentença, podendo estabelecer relação de coordenação, subordinação e correlação entre si.

A partir dessa releitura, algumas etiquetas formuladas por Chaves (2012) permaneceram intactas, sendo outras mudadas em relação à nomenclatura, tendo acontecido, também, alguns casos de aglutinação de diferentes etiquetas em apenas uma. Outras surgiram ainda pelos casos inéditos encontrados no *corpus* do presente estudo. Nossa anotação fez uso, em relação aos tipos de problema de segmentação, de apenas u etiqueta, que foi para o problema do tipo linguístico – <PROSEGL>, pois não mais etiquetamos os problemas de segmentação dos tipos retórico e visual. Quanto aos problemas de segmentação linguística, fizemos uso de 12 subetiquetas, contra as 19 de Chaves. No que diz respeito aos parâmetros técnicos, optamos por replicar as etiquetas presentes no trabalho de Chaves, acrescentando a etiqueta <3L>, que não apareceu em seu *corpus*. Apresentaremos todas as etiquetas nas Subseções 3.2.1 e 3.2.2.

Na próxima seção, iremos tratar das etapas do percurso metodológico, embasado, principalmente, em ferramentas oriundas da Linguística de Corpus.

3. Metodologia

A pesquisa teve suporte teórico-metodológico, como já mencionado, nos Estudos da Tradução, nos estudos em TAV-LSE e na Linguística de Corpus. A metodologia envolveu uma dimensão descritiva pautada por análises quanti-qualitativas baseadas em *corpus*.

3.1 O *Corpus*

O *corpus* é do tipo especializado, composto da LSE do tipo *closed caption pop-on* de um episódio da telenovela ‘Amor Eterno Amor’, produzida e exibida pela Rede Globo entre 5 de março e 7 de setembro de 2012, em 161 capítulos, escrita por Elizabeth Jhin e com

direção geral de Pedro Vasconcelos. Essa LSE faz parte do Projeto CORSEL (Corpus, Segmentação e Legendagem), que tem como objetivo a identificação e descrição-classificação dos problemas de segmentação na LSE da TV brasileira e é vinculado ao grupo de pesquisa LEAD-UECE. Para o projeto CORSEL, foram gravadas duas semanas de programação dos canais de TV do Brasil que oferecessem LSE do tipo *closed caption pop-on*. A partir da nossa pesquisa sobre quais itens de programação da TV brasileira possuíam legenda com os requisitos acima descritos, chegamos a quatro principais tipos de programa: novelas; seriados e programas de humor. Optamos por utilizar como *corpus* para o nosso recorte a LSE de um episódio da novela ‘Amor Eterno Amor’. Escolhemos apenas um episódio da novela em questão, em um universo de 5 que serão analisados no âmbito do projeto, por questões de espaço e de não haver tempo hábil de análise de todos os episódios para essa publicação.

Consideramos a LSE de um episódio da novela em questão, com um total de 294 legendas e 5.181 palavras corridas¹² (*tokens*), como sendo representativa para a análise feita, mesmo constituindo-se em um *corpus* pequeno¹³. Esse nosso posicionamento se justifica pelo fato de que, como um *corpus* é uma amostra de um fenômeno linguístico cuja extensão não se conhece, não se pode estabelecer o tamanho ideal representativo do fenômeno a ser descrito, não existindo critérios objetivos para a determinação dessa representatividade, a não ser aqueles estipuladas pelo próprio analista. Ou seja, como pontua Leech (1991, p. 27 *apud* BERBER SARDINHA, 2004, p. 25), “os usuários de um *corpus* atribuem a ele a função de ser representativo de certa variedade”.

Para a extração das legendas do tipo *closed caption*, o software *CCExtractor 0.61*¹⁴ foi o escolhido. Apresentamos sua interface na Figura 2.

¹² Sendo computados os números presentes no corpus relativos à inserção da legenda e aos tempos iniciais e finais, que, se não contabilizados, somariam 2.534 palavras corridas.

¹³ Ao observar corpora realmente utilizados pela comunidade, Berber Sardinha (2002 *apud* BERBER SARDINHA, 2004, p. 26), descreve um corpus com menos de 80 mil palavras como pequeno, por exemplo. Mas é claro que, quão mais especializado o corpus, como é o nosso caso, menores são as proporções do que seria considerado como um corpus pequeno ou grande.

¹⁴ Programa *freeware* que pode ser baixado no domínio <http://ccextractor.sourceforge.net/>

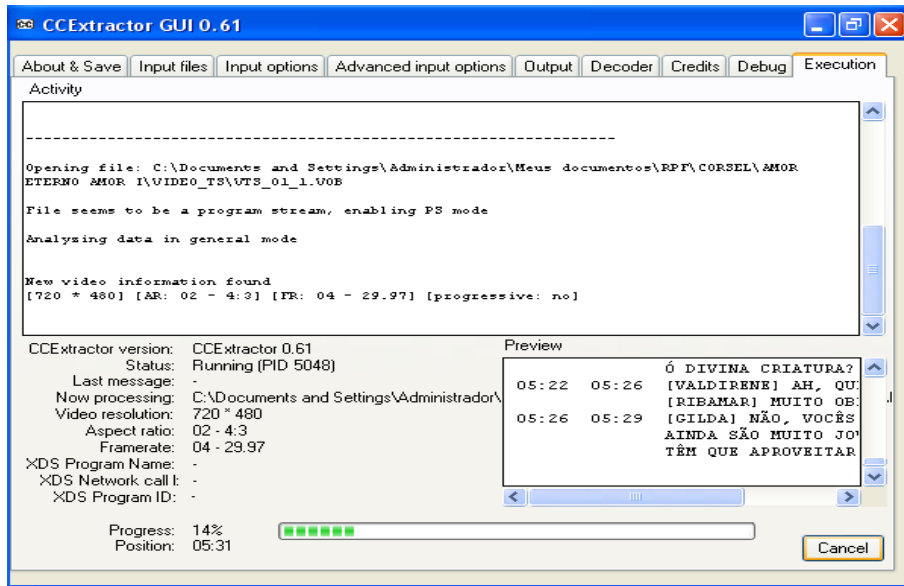


Figura 2: Interface *CCExtractor*.

Fonte: Os autores.

As vantagens do *CCExtractor* são inúmeras. Além de ter uma interface simples, sendo fácil de operar, o programa gera um arquivo de legendagem srt. com o número de inserção das legendas, assim como a marcação e sincronização das mesmas. Por vezes, contudo, o arquivo gerado apresenta pequenas falhas que precisam ser corrigidas manualmente, como no caso das legendas 1, 3, 6 e 7¹⁵ na Figura 3.

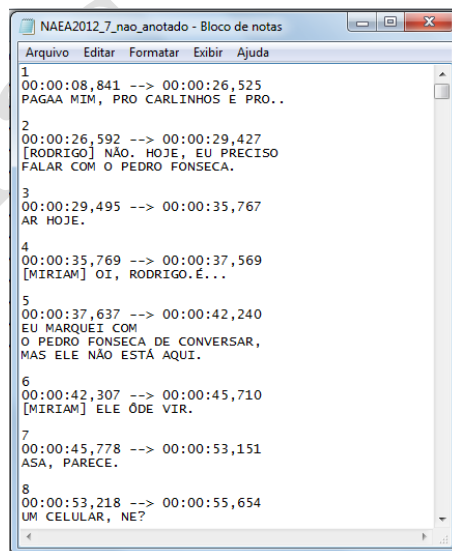


Figura 3: Arquivo de legenda em formato srt. aberto no Bloco de Notas.

Fonte: Os autores.

¹⁵ A nossa experiência nos sugere que as legendas 1, 3, 6 e 7 apresentaram problemas, provavelmente, devido a instabilidade do sinal em que é veiculado a legendagem *closed caption*. Contudo, há a possibilidade de o erro ter ocorrido na própria confecção das legendas por parte da emissora, sendo difícil detectar com absoluta certeza em que instância desse processo o erro aconteceu.

Utilizando como exemplo a primeira inserção de legenda, podemos observar a primeira linha composta do número '1', indicando a posição daquela legenda na sequência de inserções. Logo abaixo, vemos a indicação do tempo de entrada daquela inserção '00:00:08,841' e o tempo de saída '00:00:26,525', definindo assim a marcação e sincronização daquela legenda, ou seja, o tempo em que ela permanecerá na tela.

3.2 Etiquetagem

As etiquetas¹⁶ foram feitas a partir do padrão *SGML* (*Standard Generalized Markup Language*), que fornece códigos escritos no formato <etiquetas de abertura>informação</etiqueta de fechamento/>, caracterizando-as e delimitando o conteúdo a ser analisado. Ainda com o intuito de replicar a metodologia formulada por Chaves (2012; CHAVES; ARAÚJO, 2014), utilizamos, na confecção das etiquetas, o sinal de (+) de modo a indicar a quebra indevida da estrutura sintagmática, caso em que ocorrem os problemas de segmentação linguística na maioria das vezes.

Após o processo de anotação do *corpus*, feita manualmente no arquivo txt aberto no programa Bloco de Notas¹⁷, as etiquetas puderam ser tratadas de forma adequada pelo programa *Wordsmith Tools* 5.0. e analisadas com o auxílio ferramenta *Concord*¹⁸, que, de acordo com Berber Sardinha (2009, p. 9), pode ser utilizada para realizar concordâncias no texto, listando palavras específicas (nódulos) juntamente com parte do texto ou cotexto em que a mesma ocorreu.

Agora, passamos a apresentar as etiquetas.

¹⁶ Segundo a descrição de Tagnin (2010) sobre os tipos de etiquetagem, o tipo em que a nossa etiquetagem pode ser encaixado é no campo da "discursiva", processo pelo qual o conteúdo do corpus recebe etiquetas demarcando uma determinada parte do texto. Contudo, pela especificidade do projeto, podemos situar nossas etiquetas como 'etiquetas de segmentação'.

¹⁷ Editor de texto disponível no sistema operacional *Windows*, da *Microsoft*.

¹⁸ Em Assis (2013) foi utilizada também a *Wordlist*, que produz listas de palavras dos arquivos selecionados, contendo tanto suas frequências absolutas, quanto percentuais, de modo a fazer uma descrição geral das características quantitativas do corpus *per se*. Mas como esses resultados não apresentaram influência direta na análise da segmentação, optamos por deixá-los fora do estudo ora relatado.

3.2.1 Para os Parâmetros Técnicos

As etiquetas criadas para os parâmetros técnicos serviram, como já dito, para uma compreensão mais abrangente dos problemas de segmentação linguística. O Quadro 2 traz essas etiquetas.

Quadro 4: Etiquetas para parâmetros técnicos em legendagem.

ETIQUETAS DE ANÁLISE DE PARÂMETROS TÉCNICOS DA LEGENDAGEM	
Número da legenda	<sub1 ¹⁹ >legenda1</sub1>
Linhas por legenda	<1L>, <2L> e <3L>
Tempos inicial e final de cada legenda	<t>início --> final</t>
Número de caracteres por linha (aplicada em legendas de 2 e 3 linhas)	<cpl ²⁰ >
Velocidade da legenda baixa (145ppm ²¹)	<velocidade da legenda_baixa>
Velocidade de legenda média (160ppm)	<velocidade da legenda_média>
Velocidade de legenda alta (180ppm)	<velocidade da legenda_alta>

Fonte: Os autores.

As etiquetas relativas ao número de linhas de cada inserção de legenda foram as seguintes: <1L>, <2L> e <3L>. Vale lembrar que a terceira não foi utilizada por Chaves (2012; CHAVES; ARAÚJO, 2014) por ausência de legendas de 3 linhas em seu *corpus*²², característica esta comum na legendagem do tipo *closed caption pop-on* analisada aqui²³.

3.2.2 Para os Parâmetros Linguísticos

A base de sustentação para a formulação das etiquetas relacionadas à segmentação linguística, como já mencionado, são as considerações de Castilho (2012) acerca dos sintagmas e das orações coordenadas e subordinadas, as quais seguem a vertente linguística funcionalista-cognitivista do autor (ibidem, p. 32). As considerações do Castilho (ibidem), por sua vez, são respaldadas por exemplos a partir de excertos da norma culta da língua oral do

¹⁹ Abreviatura de *subtitle*, 'legenda' em inglês.

²⁰ Acrônimo de 'caracteres por linha'.

²¹ Acrônimo de 'palavras por minuto'.

²² A LSE do filme 'Nosso Lar' seguiu os padrões da legendagem comercial, que permite no máximo duas linhas de legenda.

²³ A quantidade de 3 linhas por legenda não é recomendada por estudiosos de TAV por ser, segundo pesquisas (D'YDEWALLE et al., 1987), um fator que dificulta uma boa recepção.

português brasileiro, sendo, assim, ideais para o nosso propósito de identificar e analisar problemas de segmentação linguística na legendagem *closed caption pop-on* de uma novela da TV brasileira: são ideais porque, no Brasil, via de regra, as telenovelas mostram a norma culta oral e a LSE se constitui de tradução, no meio escrito, de textos produzidos o meio oral. O Quadro 5 traz as etiquetas.

Quadro 5²⁴: Etiquetas para problemas de segmentação linguística em legendagem.

ETIQUETA INDICATIVA DE PROBLEMA DE SEGMENTAÇÃO LINGUÍSTICA
<PROSEGL>
SUBETIQUETAS PARA O SINTAGMA VERBAL (SV)
<SV_verbo+verbo> <SV_verbo+adv> → adv=advérbio <SV_(verbo)+oblíquo+verbo> <SV_negação+verbo>
SUBETIQUETA PARA O SINTAGMA NOMINAL (SN)
<SN_nominal_composto> <SN_especificador+subst> → subst=substantivo <SN_conj+subst> → conj=conjunção <SN_subst+adj> → adj=adjetivo
SUBETIQUETA PARA O SINTAGMA ADVERBIAL (SAdv)
<SAdv>
SUBETIQUETA PARA O SINTAGMA PREPOSICIONAL (SP)
<SP_prep+subst> → prep=preposição
SUBETIQUETA PARA A ORAÇÃO COORDENADA (COORD)
<COORD_conj+oração>
SUBETIQUETA PARA A ORAÇÃO SUBORDINADA (SUBORD)
<SUBORD_conj/pronome_rel+oração> → rel=pronome relativo

Fonte: Os autores.

Essas etiquetas foram elaboradas com o intuito de contemplarem aquela recomendação de Karamitroglou (1998) ao defender que, quando uma mesma legenda tiver que ser dividida em duas porções de texto em linhas distintas, essa quebra deve ocorrer no mais alto nível sintático. As etiquetas baseadas em Castilho (2012) viabilizam a operacionalização analítica da recomendação porque o que ela significa é que não devem ocorrer quebras entre os sintagmas constituintes de uma dada linha de legenda, como os verbais (SVs), nominais

²⁴ Não houve ocorrências de quebra do SAdj no corpus, sendo assim, a etiqueta correspondente a esse sintagma foi retirada do quadro de etiquetas.

(SNs), adjetivais (SAdj), adverbiais (SAdv) e preposicionais (SPs), além das orações coordenadas e subordinadas, e que não seja quebrada a estrutura interna destes.

4. Problemas de Segmentação Linguística

4.1 Resultados da Etiquetagem

Nas subseções a seguir, serão apresentados os problemas de segmentação linguística identificados a partir das ocorrências de quebra de linha nas legendas em que ou as estruturas completas de sintagmas e/ou de orações foram indevidamente separadas ou a estrutura interna de sintagmas e/ou de orações foi desconsiderada. As porções de texto que deveriam aparecer preferivelmente na mesma linha de legenda foram sublinhadas. Serão apresentados, ainda, os resultados da análise descritivo-classificatória dos problemas identificados, o que será feito pela alocação dos problemas por tipo de sintagma e de oração em subseções distintas. Pela impossibilidade de se colocarem todas as ocorrências encontradas no *corpus* dada a restrição de espaço, optamos por trazer exemplos que mostrem apenas uma só categoria de problema por tipo de sintagma e oração. Para cada exemplo, mostramos também nossa proposta de ‘ressegmentação’. Há ainda duas subseções dedicadas à análise da inter-relação dos parâmetros técnicos de velocidade e número de linhas da legenda com os problemas de segmentação.

4.1.1 Sintagma Verbal (SV)

Sobre as categorias de quebra problemática existentes no sintagma verbal, estas geralmente acontecem dentro de um sintagma verbal composto, normalmente composto de seu núcleo preenchido por verbo pleno numa forma nominal – infinitivo, particípio ou gerúndio – e especificado, ou seja, antecedido por um verbo auxiliar, como nos casos da quebra verbo+obliquo+verbo, verbo+verbo etc. Outro caso considerado é o da quebra verbo+advérbio porque, apesar de não levar em conta constituintes que estejam relacionados dentro de uma estrutura verbal complexa, eles são tão inter-relacionados semanticamente (o segundo modifica o primeiro) que, quando os mesmos não ocorrem vizinhos um do outro em uma linha de legenda, o sentido imediato da linha pode ser comprometido. Na Figura 4, mostramos um fragmento da tela da ferramenta Concord com as ocorrências de problemas de segmentação no SV.

31	E? SE VOCE DEIXAR<PROSEGL><SV_(verbo)+oblíquo+verbo> <cpl2/>EU	310	14	7%	0	6%
32	NÃO ESTOU PODENDO<PROSEGL><SV_(verbo)+oblíquo+verbo> <cpl24>ME	1,043	45	0%	0	1%
33	NÃO ESTOU PODENDO<PROSEGL><SV_verbo+verbo> <cpl28>CORRER, EU	1,038	45	2%	0	1%
34	FILHO, VOCÊ ESTÁ<PROSEGL><SV_verbo+verbo> <cpl26>PEGANDO	1,060	45	6%	0	2%
35	DEOLINDA VAI CUIDAR<PROSEGL><SV_verbo+adv> <cpl21>BEM DE MIM.	1,348	58	1%	0	7%
36	EU NÃO POSSO<PROSEGL><SV_verbo+verbo> <cpl28>ENGORDAR.	1,239	53	4%	0	5%
37	FICAR APARECENDO<PROSEGL><SV_verbo+adv> <cpl29>AQUI FORA DE	751	31	4%	0	5%
38	TÁ? PORQUE DEVE<PROSEGL><SV_verbo+verbo> <cpl31>SER SOBRA	661	27	1%	0	3%
39	ACORDO, VOCÊ SABE<PROSEGL><SV_verbo+adv> <cpl30>MUITO BEM O	796	32	1%	0	6%
40	EU NÃO ESTOU<PROSEGL><SV_verbo+verbo> <cpl>	839	35	4%	0	7%
41	QUE A SUA VÓ TÁ<PROSEGL><SV_(verbo)+oblíquo+verbo> <cpl29>ME	823	34	2%	0	7%

Figura 4: Fragmento de tela do *Concord* com ocorrências de problemas de segmentação linguística no SV.
Fonte: Os autores.

Ao todo, houve 41 problemas de segmentação na **ordem** do sintagma verbal, o que significa 50,6% dos problemas de segmentação na LSE de ‘Amor Estranho Amor’. O Quadro 6 traz um exemplo da ocorrência da categoria representada pela etiqueta²⁵ <SV_verbo+verbo>, a mais presente em todo o *corpus* com 16 ocorrências ou 19,7%:

Quadro 6: Problema de segmentação linguística - quebra no sintagma verbal.

Nº da Legenda	Número de Linhas	Velocidade da Legenda	Etiqueta	Legenda
Sub247	2	Baixa	<SV_verbo+verb o>	[GRACINHA] <u>JÁ VALE FICAR TRABALHANDO COM FEBRE, É?</u>

Fonte: Os autores.

Para exemplo dos casos de quebra da estrutura verbo+verbo, observemos a legenda 247, que possui um sintagma verbal complexo – ‘vale ficar trabalhando’ – e um advérbio anteposto – ‘já’, os quais deveriam permanecer juntos para que o legendista tivesse seguido a recomendação de Karamitoglou (1998). Se considerarmos o parâmetro da sintaxe na segmentação linguística, esta legenda deveria ficar assim:

[GRACINHA] JÁ VALE FICAR TRABALHANDO
COM FEBRE, É?

²⁵ Nos casos em que a legenda contém mais de um problema de segmentação linguística, apenas o problema do sintagma discutido na subseção correspondente é ressaltado com o sublinhado.

É interessante notar que, nessa ressegmentação, a segmentação linguística prevalece, como defende Karamitroglou (1998), em relação ao formato considerado ideal pelo teórico: linha de cima e de baixo quase do mesmo tamanho, conforme os tipos de formato descritos no Quadro 1. Como dissemos anteriormente, esse não foi o padrão observado na legendagem de 'Amor Eterno Amor' como um todo.

4.1.2 Sintagma Nominal (SN)

Quanto às categorias de quebra problemática no âmbito do sintagma nominal, identificamos quebras entre o i) Especificador e o núcleo do SN, ii) o núcleo do SN e o adjetivo que o modifica, iii) conjunções e os substantivos a eles relacionados, iv) além dos casos que denominamos 'nominais compostos'²⁶. As quebras problemáticas na ordem do sintagma nominal aconteceram 26 vezes, isto representando cerca de 32,1% do total de má segmentação linguística na LSE de 'Amor Estranho Amor'. O Quadro 7 apresenta um exemplo da categoria anotada pela etiqueta <SN_nominal_composto>, que, com 13 ocorrências ou 16%, ranqueou em segundo lugar:

Quadro 7: Problema de segmentação linguística - quebra no sintagma nominal.

Nº da Legenda	Número de Linhas	Velocidade	Etiqueta	Legenda
Sub121	2	Alta	<SN_nominal_comp osto>	NÃO VAI TER <u>EQUIPAMENTO DE SEGURANÇA</u> PRA TODO MUNDO.

Fonte: Os autores.

Nesse exemplo, relativo à legenda 121, a estrutura sublinhada que deveria estar unida em uma mesma linha corresponde ao sintagma nominal 'equipamento de segurança'. Defendemos a ideia de que o exemplo em questão se configura como uma quebra na ordem do SN se considerado o seguinte: no momento da leitura da legenda, o telespectador se depara com o substantivo 'equipamento' e se pergunta "que equipamento?". A resposta só surge na

²⁶ Essa etiqueta serviu de 'guarda-chuva', digamos, por ter aglutinado as seguintes etiquetas de Chaves: <SN_superlativo+adj> <SN_nominal+modif/modif+nominal>, <SN_nomepróprio>, <SN_título+nome próprio>, de modo que pudesse abarcar também a terminologia de Castilho (2012). Além disso, ela serviu para nomear os casos em que o SP, enquanto Complementador do sintagma nominal, que cumpre função adjetiva é separado do núcleo do SN. Contudo, em reuniões do nosso grupo de pesquisadores, essa etiqueta tem se demonstrado problemática, dado, que com um nome mais geral, tenta abarcar muitas situações específicas. É certo que em futuros trabalhos essa etiqueta será remodelada e diluída em outras.

linha seguinte, onde está o sintagma preposicional ‘de segurança’, que, na verdade, funciona como sintagma adjetival por conferir uma qualidade/característica ao substantivo. Uma possibilidade de quebra que conciliaria melhor a questão da sintaxe e a facilitação do processo de compreensão da porção de legenda em questão está logo abaixo:

NÃO VAI TER
EQUIPAMENTO DE SEGURANÇA PARA TODOS.

Neste caso, o formato não é o retangular preconizado por Karamitroglou como o que melhor representa uma boa segmentação. Porém, esse formato triangular permite uma boa visualização para uma legenda com a linha superior mais curta, conforme o Quadro 1.

4.1.3 Sintagma Preposicional (SP)

Utilizando a etiqueta <SP_prep+subst>, identificamos a categoria de quebra problemática do sintagma preposicional que diz respeito à separação entre o seu núcleo e o seu Complementador, que, no caso do *corpus* do estudo ora relatado, só se realizou por substantivo. Essa categoria ocorreu 6 vezes ou cerca de 7,4% do total de má segmentação linguística na LSE de ‘Amor Estranho Amor’. A categoria em apreço está exemplificada no Quadro 8.

Quadro 8: Problema de segmentação linguística - quebra no sintagma preposicional.

Nº da Legenda	Número de Linhas	Velocidade	Etiqueta	Legenda
Sub4	3	Alta	<SP_prep+subst>	O GUARDINHA ME PAROU POR CAUSA <u>DE</u> <u>UMA BOBAGEM</u> DA PLACA QUE CAIU!

Fonte: Os autores.

Podemos observar na legenda 4 que há a separação, indevida a nosso ver, da preposição ‘de’ e do sintagma nominal ‘uma bobagem’. A seguir, fazemos uma proposta de como a legenda poderia ficar se considerado o parâmetro linguístico:

O GUARDINHA ME PAROU POR CAUSA
DE UMA BOBAGEM DA PLACA QUE CAIU!

4.1.4 Sintagma Adverbial (SAdv)

Aqui, consideramos como quebra problemática relacionada ao sintagma adverbial a segmentação de estruturas adverbiais compostas, isto é, sempre que houve a separação entre dado advérbio e outro lhe acrescenta informação. Essa categoria foi anotada pela etiqueta <SAdv> e houve apenas 1 ocorrência, representando 1,2% dos problemas de segmentação linguística na LSE de ‘Amor Estranho Amor’. A única ocorrência encontra-se no Quadro 9.

Quadro 9: Problema de segmentação linguística - quebra no sintagma adverbial.

Nº da Legenda	Número de Linhas	Velocidade	Etiqueta	Legenda
Sub131	3	Alta	<SAdv>	ENTROU COM UM PROCESSO LÁ NA EMPRESA EM QUE ELE TRABALHAVA, GANHOU UMA GRANA FERRADA.

Fonte: Os autores.

A legenda 131 contém um advérbio de lugar – ‘lá’ – e um sintagma preposicional com função de SAdv igualmente de lugar – ‘na empresa’ – que, juntos, formam uma só estrutura adverbial complexa. Como os dois segmentos estão em linhas distintas, houve a separação entre um advérbio e aquele que o modifica. A nosso ver, portanto, o SAdv complexo ‘lá na empresa’ deveria permanecer inseparável em uma única linha. Por se tratar de uma legenda com densidade lexical alta (compõe-se de 83 caracteres quando o máximo deveria ter 78 caracteres ou 39 caracteres por linha como informamos na Subseção 2.2), propor uma ressegmentação sem considerar o tempo disponível para a marcação é uma tarefa complicada. A legenda 131, provavelmente, teria que ou ser bem condensada ou ser ressegmentada em duas porções de legenda. Como nossas ressegmentações são apenas a título de exemplificação, podemos propor uma que seja condensada e respeite a segmentação linguística sem dividir a mesma em mais legendas. Contudo, para isso, achamos necessário retirar o advérbio ‘lá’, redundante no contexto:

ENTROU COM UM PROCESSO NA EMPRESA
EM QUE TRABALHAVA E GANHOU BOA GRANA.

4.1.5 Orações Subordinadas (SUBORD) e Coordenadas (COORD)

A identificação de quebras problemáticas em orações coordenadas e subordinadas foi um dos aspectos analíticos mais complicados, pois há uma instrução clara nos manuais de legendagem de, sempre que possível, deixar cada uma das orações que formam um período composto em uma linha da legenda, de forma a colocar o sentido completo de uma oração em cada linha (IVARSSON; CAROLL, 1998). Conciliar essa instrução com uma boa segmentação linguística foi um desafio.

Interpretando o trabalho de Chaves (2012; CHAVES; ARAÚJO, 2014) e de Perego (2008) sobre a segmentação em orações coordenadas e subordinadas e utilizando as perspectivas gramático-funcionais de Castilho (2012) sobre as mesmas, definimos como problemas de segmentação apenas os casos em que as conjunções coordenativas e subordinativas – incluindo nestas os pronomes relativos – aparecem desconectadas da oração coordenada ou subordinada que iniciam. Quebras problemáticas na ordem da oração complexa relativamente à oração subordinada ocorreram 6 vezes, representando 7,4% das ocorrências totais na LSE de ‘Amor Estranho Amor’. O mesmo tipo de quebra, mas relacionada à oração coordenada, aconteceu apenas 1 vez ou cerca de 1,2% das ocorrências totais de quebras problemáticas no *corpus*. O Quadro 10 traz um exemplo da categoria representada pela etiqueta <SUBORD_conj/pronome_rel+oração>.

Quadro 10: Problema de segmentação linguística - quebra em oração subordinada.

Nº da Legenda	Número de Linhas	Velocidade	Etiqueta	Legenda
Sub21	3	Alta	<SUBORD_conj/pronome_rel+oração>	NÃO TEM NADA NESSE MUNDO <u>QUE</u> <u>SUBSTITUA</u> CARINHO DE MÃE, NÉ?

Fonte: Os autores.

Nesse exemplo, ocorre separação entre o pronome relativo ‘que’ e o restante da oração subordinada por ele introduzida: ‘que substitua carinho de mãe, né?’. Uma sugestão para a ressegmentação da legenda 21 seria:

NÃO TEM NADA NESSE MUNDO
QUE SUBSTITUA CARINHO DE MÃE, NÉ?

4.1.6²⁷ Quebras Problemáticas por Número de Linhas das Legendas

A análise mostrou que 72 inserções tiveram 1 linha de legenda ou 24,5% do total. A maior parte do *corpus* é composta de legendas de 2 linhas, num total de 137 ocorrências ou 46,6 % do total. As legendas de 3 linhas somaram 85 ocorrências ou 28,9% do total de inserções. A Tabela 1 mostra a relação entre os problemas de segmentação e o número de linhas das legendas.

Tabela 1: Ocorrências de problemas de segmentação por número de linhas das legendas.

Número de linhas das legendas	Percentual de ocorrência de quebras problemáticas
2L	33,3%
3L	66,7%

Fonte: Os autores.

Os problemas de segmentação linguística só apareceram em legendas de 2 e 3 linhas, já que a segmentação aqui analisada é a de quebra de linha, que só ocorre em legendas com mais de 1 linha. 54 ocorrências de problemas de segmentação de foram em legendas de 3 linhas, o que significa 66,7% do total. Os outros 27 problemas apareceram em legendas de 2 linhas, o que representa 33,3% das ocorrências.

4.1.7 Quebras Problemáticas por Nível de Velocidade das Legendas

Quanto à velocidade, em um universo de 294 legendas, a maior frequência de ocorrência no *corpus* foi de legendas com velocidade alta, em um total de 169 ou 57,5% do total. A segunda maior foi a de legendas com velocidade baixa, perfazendo 95 inserções ou 32,3% do total. Por último, as legendas com velocidade média foram em número de 30, ou 10,2% do total. Abaixo, apresentamos a Tabela 2 com os percentuais de problemas de segmentação por nível de velocidade.

²⁷ Para efeitos de análise, apenas alguns dos parâmetros técnicos foram levadas em conta, como caracteres por linha, número de linhas e velocidade da legenda, por apresentarem relações mais intrínsecas com o parâmetro da segmentação linguística. Os outros parâmetros etiquetados foram importantes para a localização da legenda no corpus, como a etiqueta que indica o número da legenda e a que delimita os tempos inicial e final da legenda. Apesar de ter sido importante para o cálculo da velocidade da legenda, o parâmetro técnico de 'caracteres por linha' (<cpl>), por não ter apresentado uma relação tão forte com a segmentação como os parâmetros de velocidade e número de linhas, decidimos, pelo espaço limitado não incluí-lo nas subseções de inter-relação entre parâmetros técnicos e problemas de segmentação.

Tabela 2: Ocorrências de problemas de segmentação por nível de velocidade.

Nível de velocidade das legendas	Percentual de ocorrência de quebras problemáticas
Alto	76,5%
Médio	12,4%
Baixo	11,1%

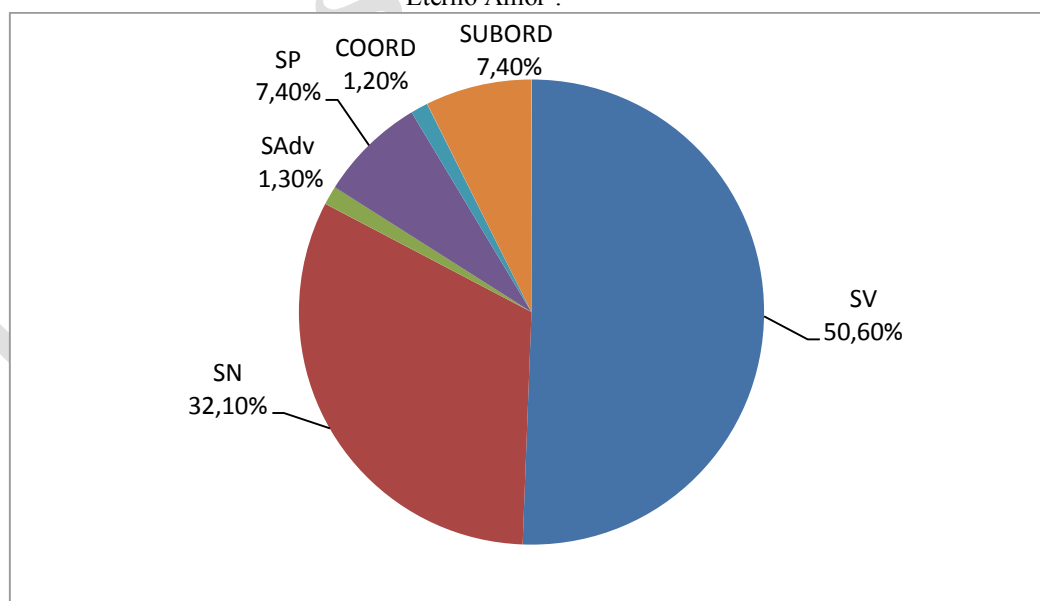
Fonte: Os autores.

A tabela mostra que, no que concerne à relação entre as velocidades das legendas e os problemas de segmentação encontrados, foi nas legendas com velocidade alta que os problemas predominaram: 62 ou 76,5% das ocorrências. O segundo lugar foi ocupado pelas legendas com velocidade média, com 10 problemas ou 12,4% das ocorrências. Em último lugar, ficaram as legendas com velocidade baixa, com 9 problemas ou 11,1% das ocorrências.

4.2 Sistematização e Discussão Geral dos Resultados

Sistematizando os resultados quantitativos decorrentes dos dados provenientes da análise baseada em *corpus* que viabilizou a identificação e a descrição classificatória dos problemas de segmentação linguística na LSE da novela ‘Amor Eterno Amor’, temos o Gráfico 1:

Gráfico 1: Sistematização dos resultados dos problemas de segmentação linguística na LSE da novela ‘Amor Eterno Amor’.



Fonte: Os autores.

Como é possível observar no Gráfico 1, o maior número de problemas ocorre nas ordens do SV, com contribuição de 50,6%, e do SN, com participação de 32,1%, o que corrobora os resultados de Chaves (CHAVES, 2012; CHAVES; ARAÚJO, 2014). Além disso, assim como no trabalho de Chaves, a estrutura em que houve mais ocorrências de quebras problemáticas foi em verbo+verbo, assim como a maior parte dos problemas de segmentação linguística encontrados aconteceram em legendas com velocidade considerada alta (cf. Subseção 2.3).

Os resultados semelhantes de que obtivemos em relação à Chaves sugerem, mesmo em gêneros ficcionais audiovisuais diferentes (telenovela e filme em DVD), o que pode se apresentar como uma tendência para a segmentação linguística em gêneros ficcionais. Porém, ainda é cedo para apontar de forma substancial padrões característicos da segmentação linguística no Brasil, o que será possível com o aumento de *corpus* analisado em tamanho e variedade de gênero.

Contudo, apesar das similaridades entre os resultados de Chaves e o presente trabalho, é inegável, se examinados os números provenientes da análise da LSE no capítulo da novela ‘Amor Eterno Amor’, a grande relevância que os problemas de segmentação linguística têm no *corpus*. Ao todo, identificamos 81 problemas, todos anotados pela etiqueta <PROSEGL>. Estes, de uma forma geral, aparecem em 75 legendas em um universo de 294, representando 25,5% do *corpus*. Ou seja, a cada quatro legendas, um pouco mais de uma possui problema de segmentação, teoricamente dificultando bastante a compreensão do texto traduzido. Nas legendas de 2 linhas, cerca de 19,7% delas (27 casos em 27 legendas diferentes) possuem algum tipo de problema de segmentação, seja na ordem do sintagma ou da oração complexa. Nas legendas de 3 linhas, a porcentagem foi ainda maior, com 56,4% destas (54 casos em 48 legendas diferentes) apresentando uma ou mais ocorrências de algum tipo de problema de segmentação linguística, já que algumas legendas de 3 linhas apresentaram mais de um caso.

O grande número de legendas com uma velocidade considerada alta e de 3 linhas não foi algo de inteiro surpreendente, já que isso só reforça a noção equivocada da legendagem para surdos da TV brasileira, em desacordo com os parâmetros estabelecidos pelos pesquisadores da área. A alta velocidade das legendas é resultado previsível, se considerado que a legenda *closed caption* no Brasil é muito mais uma transcrição da fala do

que um processo tradutório que siga os preceitos dos estudiosos da Tradução Audiovisual, especificamente a LSE

Com todos esses resultados, pudemos identificar, de uma forma mais geral, o padrão de má segmentação linguística no *corpus*: na maioria das vezes, nos sintagmas verbal e nominal em legendas de 3 linhas e com velocidade alta. O fato de a segmentação ter sido problemática nas legendas mais longas – as de 3 linhas – e de velocidade alta se impõe como fator complicador para a recepção de produtos legendados, de acordo com o que estipula a literatura da área em relação à LSE. A partir dos resultados obtidos, podemos caracterizar a LSE no capítulo de ‘Amor Eterno Amor’ em questão como fora dos padrões de uma legendagem confortável de ser consumida pelos telespectadores surdos e ensurdecidos.

5. Considerações Finais

Nossas conclusões sobre a segmentação linguística estudada na telenovela ‘Amor Eterno Amor’ só foram possíveis graças à metodologia baseada em *corpus* proposta por Chaves (2012; CHAVES; ARAÚJO, 2014) para a análise do parâmetro linguístico em questão, mesmo com os poucos dados analisados aqui. Isso foi motivado, contudo, pelo tamanho e tempo restritos da própria pesquisa. A partir disso, nos foi mais interessante estudar uma pequena quantidade de legendas, que nos possibilitou empreender uma análise que desse mais vazão à interpretação qualitativa de nossos dados. Isso seria difícil de fazer com um grande *corpus* em tão pouco tempo, sendo, portanto, preferível à fazer um simples empilhamento de dados quantitativos.

A metodologia baseada em *corpus* também nos permitiu observar, de forma concomitante à análise da segmentação linguística em ‘Amor Eterno Amor’, uma amostra das características técnicas da LSE da TV Brasileira, contribuindo, assim, com nossos resultados, para a pesquisa em LSE realizada na UECE desde 2002, a qual visa estudar como esse tipo de Legendagem caracteriza-se e como ela é recebida pelo seu público alvo, os surdos e ensurdecidos. O Projeto CORSEL, ao qual os resultados desta pesquisa estão atrelados, está filiado nessa perspectiva. Este, ao seu fim, permitirá uma caracterização mais conclusiva a respeito das características técnicas e de como acontece a segmentação linguística na LSE do Brasil a partir da abrangência em quantidade e diversidade genérica de narrativas audiovisuais que o projeto pretende abarcar.

A análise dos parâmetros técnicos da legendagem *closed caption* no *corpus* foi bastante importante para a compreensão dos problemas de segmentação linguística encontrados, tendo embasado nossas considerações acerca destes e possibilitado uma correlação entre ambos. Para além disso, essa análise propiciou o que parece ser as características técnico-linguísticas da LSE na TV brasileira: legendas em sua maioria de velocidade alta, de 3 linhas e com uma grande quantidade de problemas de segmentação de ordem linguística. Do ponto de vista dos parâmetros linguísticos, os resultados apontaram o sintagma verbal e o nominal como os mais problemáticos em relação aos casos de má segmentação no *corpus*, tal como aconteceu em Chaves (2012; CHAVES; ARAÚJO, 2014).

Os dados resultantes do estudo que acabamos de relatar e dos demais no âmbito do CORSEL poderão, no futuro, possibilitar novas pesquisas dentro da interface Linguística de Corpus/LSE. Entre estas, por exemplo, está a elaboração e compilação de um *corpus* eletrônico de LSE acessível por computador, visando a consulta por legendistas e/ou seu treinamento através da plataforma do Projeto CoMET – Corpus Multilíngue para o Ensino e Tradução²⁸, um *corpus* monolíngue e paralelo voltado para pesquisadores e profissionais de legendagem, com o qual já mantemos *link* acadêmico. Outra possibilidade de utilização dos dados do Projeto CORSEL, assim como da presente pesquisa, é na investigação experimental com o auxílio de rastreamento ocular para o estudo da real influência do parâmetro da segmentação na recepção de legendas por pessoas surdas-ensurdecidas e ouvintes. Essa pesquisa se faz necessária, pois, apesar de ser um parâmetro considerado como essencial para a confecção de uma boa legendagem, o mesmo ainda carece de pesquisas experimentais que possam mensurar, de fato, a sua contribuição no processo de compreensão do produto legendado, sendo este um dos próximos passos da pesquisa em LSE pelo grupo LEAD na UECE, com auxílio da metodologia fornecida pela Linguística de Corpus.

Referências Bibliográficas

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

ARAÚJO, V. L. S. Closed subtitling in Brazil In: ORERO, P. (org.). **Topics in Audiovisual translation**. Amsterdã: John Benjamins Publishing Company, v. 1, p. 199 - 212, 2004. **crossref** <http://dx.doi.org/10.1075/btl.56.20san>

²⁸ <http://www.fflch.usp.br/dlm/comet/>

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

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.

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

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

BAKER, M. Corpus Linguistics and Translation Studies: Implications and Applications. In: Baker, M.; Francis, G.; Tognini - Bonelli, E. (orgs.). **Text and technology**: In honour of John Sinclair. Philadelphia, Amsterdam: John Benjamins, p. 233- 250, 1993. **crossref** <http://dx.doi.org/10.1075/z.64.15bak>

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

BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009, 299 p.

CASTILHO, A. de. **Nova Gramática do Português Brasileiro**. São Paulo: Contexto, 2012, 768p.

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

CHAVES, E. G.; ARAÚJO, V. L. S. Segmentation tags: a proposal for the analysis of subtitles. In: ALUÍSIO, S. M.; TAGNIN, S E. O. (orgs.) **New language, technologies and linguistic research**: a two way road. Newcastle upon Thyne: Cambridge Scholar's Publishing, 2014, p. 62-75.

DIAZ- CINTAS, J.; REMAEL, A. **Audiovisual Translation**: Subtitling. Manchester, UK, Kinderhook, N Y, UK : St. Jerome Publishing, 2007, 272 p.

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

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.

GOTTLIEB, H. Subtling: Diagonal Translation. In: **Perspective in Translatology**, v.2, n.1, p.101-121, 1994.

IVARSSON, J.; CARROLL, M.; **Subtitling**. Simrishamm, Suécia: TransEditHB, 1998,184 p.

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

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.

PERINI, M. A. **Gramática do português brasileiro**. São Paulo: Parábola Editorial, 2010, 366 p.

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

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

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

WORDSMITH Tools 5.0: sítio do programa disponível em: <<http://www.lexically.net/wordsmith/index.html>>. Acesso em: 17 de março de 2013.

Artigo recebido em: 13.10.2014

Artigo aprovado em: 14.12.2014