

# A SEGMENTAÇÃO LINGUÍSTICA DAS LEGENDAS PARA SURDOS E ENSURDECIDOS (LSE) DE TELENOVELAS: UMA PESQUISA BASEADA EM CORPUS<sup>1</sup>

## THE LINGUISTIC SEGMENTATION OF SUBTITLES FOR THE DEATH AND THE HARD-OF-HEARING (SDH) OF SOAP OPERAS: A CORPUS-BASED RESEARCH

**Vera Lúcia Santiago Araújo\***  
**Ítalo Alves Pinto de Assis\*\***  
**Daniel de Albuquerque e Arraes\*\*\***

---

### RESUMO

Este trabalho tem o objetivo de descrever a segmentação na legendagem, ou seja, a distribuição das falas em duas ou mais legendas. Os dados colhidos em projetos anteriores realizados pelo grupo Legendagem e Audiodescrição (LEAD) da UECE sugerem que uma segmentação adequada pode garantir uma boa recepção por parte de surdos. Isso está acontecendo até mesmo na recepção a legendas muito rápidas. Com o suporte teórico-metodológico dos Estudos da Tradução, mais especificamente da Tradução Audiovisual (TAV) e da Linguística de Corpus, a metodologia envolveu uma dimensão descritiva pautada por análises quanti-qualitativas. Foram analisadas quatro novelas exibidas por emissoras brasileiras que disponibilizam legendas do tipo *pop-on*, usadas para transmitir programas pré-gravados. A legenda *pop-on* é aquela cujas frases ou sentenças surgem como um todo e não palavra por palavra, como acontece com a legenda rotativa. Os resultados mostraram que, para as novelas, os maiores problemas de segmentação encontrados estão na quebra entre os constituintes do sintagma verbal.

**Palavras-chave:** legendagem para Surdos e Ensurdidos (LSE); Segmentação Linguística; Linguística de Corpus.

### ABSTRACT

This paper aims at describing the parameter of segmentation in subtitling, that is, the distribution of the lines in two or more subtitles. The data collected in previous projects carried out by the UECE's Subtitling and Audiodescription (LEAD) group suggests that

---

<sup>1</sup> Este projeto teve financiamento do CNPq em forma de bolsa de produtividade de Vera Lúcia Santiago Araújo.

\* Universidade Estadual do Ceará, Fortaleza CE. Brasil. verainnerlight@gmail.com (UECE).

\*\* Universidade Estadual do Ceará, Fortaleza CE. Brasil. italoalves1991@gmail.com (UECE).

\*\*\* Universidade Estadual do Ceará, Fortaleza CE. Brasil. arraes.d@gmail.com (UECE).

adequate segmentation could guarantee good reception by the deaf. This is happening even in reception to very fast subtitles. With the theoretical-methodological support of Translation Studies, more specifically of Audiovisual Translation (AVT) and Corpus Linguistics, the methodology involved a descriptive dimension based on quantitative-qualitative analysis. Four soap operas exhibited by Brazilian TV Channels that provide pop-on subtitles, the type used to broadcast prerecorded programs were analyzed. The pop-on subtitle is the one whose phrases or sentences appear as a whole and not word for word as with the roll-up subtitle. The results showed that, for soap operas, the main problems of segmentation found are in the break between the constituents of the verbal phrase.

**Keywords:** subtitling for the deaf and hard-of-hearing (SDH); linguistic segmentation; Corpus Linguistics.

## INTRODUÇÃO

Desde meados do século XX, os meios midiáticos de imagem em movimento, como cinema e televisão, têm se desenvolvido constantemente e mantido um lugar preponderante nas produções culturais humanas, tanto globais, quanto locais. Especificamente sobre o cinema, a transição dos filmes mudos para sua contraparte falada interpôs um obstáculo à distribuição internacional das obras cinematográficas: tornou-se necessário o desenvolvimento de técnicas de tradução dos diálogos dos filmes na língua de partida para a língua de chegada do público consumidor.

Dentre as técnicas desenvolvidas está a legendagem, que

consiste em representar pela escrita a tradução, para uma língua alvo, das trocas de diálogos originais enunciados pelos diferentes falantes, bem como de todas as outras informações verbais transmitidas visualmente (cartas, banners, encartes) ou pela audição (letras de música, voz *off*). (DIAZ-CINTAS, 2010, p. 344)

A legendagem é uma modalidade de tradução estudada dentro dos Estudos da Tradução, mais especificamente na subárea Tradução Audiovisual (TAV), juntamente com outras três modalidades: a dublagem, o *voice-over* e a audiodescrição. Nas últimas décadas, além de seu uso já corriqueiro na disponibilização de produtos audiovisuais em diferentes idiomas, a legendagem tem sido empregada também como meio de acessibilidade, permitindo que pessoas com diferentes níveis de surdez tenham acesso às miríades de produções culturais televisivas, cinematográficas, digitais etc. Essa modalidade de legendagem tem sido chamada pelos pesquisadores da área de *legendagem para surdos e ensurdecidos* (LSE).

No Brasil, ainda estão sendo aperfeiçoados os critérios de qualidade necessários para que seu público alvo usufrua a contento das diferentes produções audiovisuais, tornando-as efetivamente acessíveis. Nesse sentido, os pesquisadores do grupo Legendagem e Audiodescrição<sup>2</sup> (LEAD) da Universidade Estadual do

Ceará (UECE) têm desenvolvido pesquisas no sentido de esboçar tais critérios de qualidade no contexto da comunidade surda brasileira. Sucintamente, as primeiras pesquisas (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004) consistiram em pesquisas exploratórias que testaram legendas do tipo *closed caption* produzidas pelos canais de televisão nacionais. Os depoimentos dos participantes apontavam que era a *velocidade* o fator responsável por problemas de recepção dessas legendas por parte dos surdos brasileiros.

Posteriormente, novas pesquisas exploratórias (ARAÚJO, 2008) foram realizadas no sentido de testar diferentes modelos de legendagem com surdos cearenses para chegar a uma proposta que atendesse ao público brasileiro. O parâmetro apontado como responsável pela má recepção dos surdos foi o da velocidade. Esse modelo foi testado novamente (ARAÚJO; NASCIMENTO, 2011; ARAÚJO; VIEIRA; MONTEIRO, 2013, ARAÚJO, 2012) em uma pesquisa de âmbito nacional, na qual foram feitos testes de recepção com 34 surdos distribuídos em sete capitais brasileiras – Porto Alegre e Santa Catarina, na região Sul; Rio de Janeiro e Belo Horizonte, no Sudeste; Salvador e Fortaleza, no Nordeste; e Rio Branco, no Norte.

As coletas consistiam em apresentar aos participantes trechos de produtos audiovisuais, com o objetivo de confirmar uma hipótese: legendas rápidas (180 palavras por minuto - ppm) causariam problemas na compreensão das informações ali contidas, prejudicando a recepção, ao passo que legendas lentas (145ppm) seriam facilmente lidas pelos participantes. Curiosamente, os resultados dessa nova pesquisa foram de encontro à hipótese levantada: nas diferentes capitais, os participantes dos testes de recepção acompanharam eficientemente os filmes legendados com legendas rápidas desde que bem segmentadas. Uma reflexão posterior levou à postulação de uma nova hipótese, qual seja: a *segmentação* tem influência na recepção das legendas pelos surdos brasileiros.

Segmentação, de acordo com Diaz-Cintas e Remael (2007), significa dividir os diálogos traduzidos em segmentos – *inserções de legenda* – que possam ser compreendidos pelos espectadores. A segmentação pode seguir três critérios: os cortes de cena (segmentação visual); as pausas respiratórias dos falantes (segmentação retórica); e as características sintáticas e semânticas da mensagem (segmentação linguística). Sobre esses três critérios, a pesquisa de Chaves (2012; CHAVES, ARAÚJO, 2014) sugeriu que a segmentação linguística seria aquela que traria mais dificuldades na recepção. Por essa razão, no triênio 2012-2014 foi levado a

---

2 <https://grupoleaduece.blogspot.com.br>

cabo o projeto de pesquisa denominado “A segmentação na legendagem para surdos e ensurdecidos (LSE): um estudo baseado em corpus” (ARAÚJO, 2015), abreviado como projeto CORSEL (Corpus, Segmentação e Legendagem). O projeto coletou legendas de diferentes programas televisivos – novelas, documentários e comédias de situação – visando a identificação de legendas com segmentação linguística problemática, sua classificação em categorias, bem como a elaboração de soluções para os problemas (*ressegmentação*).

O presente artigo apresenta os resultados obtidos para o gênero telenovela. Foram analisadas as legendas de quatro novelas. O corpus foi constituído pela extração das legendas de cinco episódios de cada uma dessas novelas. Além desta introdução, o artigo possui uma seção sobre as bases teóricas, a qual aborda, em maiores detalhes, a segmentação linguística e sua relação com outros parâmetros de legendagem; uma seção sobre a metodologia, a qual descreve os procedimentos realizados para coleta dos dados e posterior análise; uma seção de discussão dos resultados; e uma seção de considerações finais.

## 1. LEGENDAGEM PARA SURDOS E ENSURDECIDOS (LSE)

A LSE é um recurso de acessibilidade para espectadores surdos que consiste na inserção de legendas em produções audiovisuais. Essas legendas diferem daquelas para ouvintes por alguns aspectos: 1) introdução de informações adicionais dependentes do canal auditivo para que aqueles com deficiência possam acompanhar filmes e programas de televisão; 2) questões técnicas; e 3) concepção de tradução. (FRANCO; ARAÚJO, 2003; ARAÚJO, 2004; ARAÚJO, 2005; ARAÚJO, 2007). As informações adicionais são de dois tipos: efeito sonoro e identificação do falante (ARAÚJO, 2008). Sobre as questões técnicas, podemos expor que as legendas para ouvintes, como as conhecemos hoje, têm no máximo duas linhas; são normalmente de cor branca ou amarela; duram no mínimo 1 e no máximo 6 segundos (D'YDEWALLE ET ALLI, 1987); e possuem três velocidades que permitem que o espectador harmonize a leitura das legendas com o áudio e as imagens do filme: 145 palavras por minuto (ppm), 160 ppm e 180ppm (D'YDEWALLE ET ALLI, 1987; IVARSSON; CAROL, 1998); DIAZ CINTAS; REMAEL, 2007).

Essa harmonização entre imagem, áudio e legendas é conquistada com a redução do texto para que o espectador tenha tempo de ler as legendas, desfrutar das imagens e ainda ouvir o áudio original, no caso dos ouvintes. A redução é necessária porque os espectadores, surdos ou ouvintes, absorvem a fala muito mais rapidamente do que a escrita. Para que o telespectador assista à produção legendada

com conforto, é preciso que a velocidade da fala seja de 145, 160 ou 180ppm. No caso de a fala ser mais rápida, as legendas precisarão ser editadas para que a recepção seja eficiente.

Para que a edição aconteça, algumas vezes pode ser necessária a segmentação de uma fala em duas ou mais legendas. Segundo Diaz Cintas e Remael, “quando segmentamos uma sentença, forçamos o cérebro a pausar seu processo linguístico por um instante, até os olhos localizarem a próxima informação linguística” (2007, p. 173). Isso significa que se a segmentação não for cuidadosa, os leitores farão esforço para decodificar o texto e, conseqüentemente, poderão se cansar mais rapidamente e perder o prazer proporcionado pelo filme ou programa. As legendas precisam ter um pensamento completo para serem entendidas no curto espaço de tempo em que são exibidas. Diaz Cintas e Remael (2007, p.172) propõem que..... a segmentação acontece dentro da mesma legenda ou entre legendas diferentes. O primeiro caso, a QUEBRA DE LINHA, diz respeito à divisão interna de uma legenda quando esta precisa ser dividida em duas linhas. O segundo caso, chamado também de SEGMENTAÇÃO, relaciona-se ao fato de o fenômeno acontecer entre duas ou mais legendas.

No que concerne à segmentação linguística (REID, 1990, p. 100), Karamitroglou (1998) afirma que o texto legendado deve ser segmentado no mais alto nível sintático possível. Isso significa que, idealmente, cada legenda deve conter um pensamento completo composto por unidades semânticas nos níveis do período, da oração e dos sintagmas (nominais, verbais, preposicionais, adjetivais e adverbiais). Isso quer dizer que, quando uma oração ou um sintagma precisam ser enunciados em mais de uma linha, a segmentação deve acontecer de acordo com as regras da sintaxe (PEREGO, 2008).

Vejamos então o exemplo de uma segmentação problemática presente nas as legendas oferecidas por um partido político na campanha eleitoral de 2006 (ARAÚJO, 2009:165).

Quadro 1. Segmentação de um programa do horário eleitoral

	Legendas	Duração segundos: quadros	Caracteres
L1	Só neste ano, oitenta e cinco mulheres foram assassinadas	2:12	58
L2	no estado do Ceará. Enquanto isso, apenas três assassinos	2:20	56
L3	foram condenados. Essa impunidade aumenta os índices	2:16	52
L4	de violência doméstica e sexual contra as mulheres.	2:18	49

Fonte: ARAÚJO, 2009, p. 165

Como não houve cortes, a segmentação deveria ter sido orientada pela retórica e pela sintaxe. Então, levando em conta esses dois critérios, as legendas do Quadro 1 seriam assim ressegmentadas:

Quadro 2. Proposta de ressegmentação das legendas do Quadro 1

	Legendas	Duração	Caracteres
L1	Só neste ano, 85 mulheres foram assassinadas no estado do Ceará.	3:10	77
L2	Enquanto isso, apenas três assassinos foram condenados.	2:07	55
L3	Essa impunidade aumenta os índices de violência doméstica e sexual contra as mulheres.	3:20	83

Fonte: ARAÚJO, 2009, p. 166

Atendidos os requisitos da segmentação, vemos que mais dois parâmetros não foram observados: o da velocidade da legenda e o da condensação. De acordo com uma tabela proposta por Diaz Cintas e Remael (2007) para a obtenção da velocidade de 145 ppm, que foi a velocidade apontada nos estudos sobre LSE realizados até agora na UECE, o número de caracteres relacionados ao tempo de exibição da legenda seriam:

Tabela 1: Máximo de caracteres para a velocidade de 145ppm

Segundos: Quadros	Caracteres
01:00	16
01:04	17
01:08	18
01:12	20
01:16	23
01:20	25
02:00	29
02:04	32
02:08	34
02:12	36
02:16	38
02:20	40
03:00	44
03:04	46
03:08	48
03:12	50
03:16	52
03:20	54
04:00	58
04:04	60
04:08	62
04:12	64

Fonte: Dias Cintas e Remael, 2007, p.97

Nenhuma das três legendas atende aos critérios preconizados na Tabela 3. Elas estão muito densas para o seu tempo de exibição. Para que essas legendas pudessem ser lidas confortavelmente, elas precisariam ser reduzidas. Uma possível condensação seria:

Quadro 3. Uma proposta de edição

	Legendas	Duração	Caracteres
L1	Em 2006, 85 mulheres foram mortas no Ceará	3:10	43
L2	mas apenas 3 assassinos foram condenados.	2:07	41
L3	A impunidade piora a violência doméstica e sexual contra mulheres.	3:20	66

Fonte: ARAÚJO, 2009, p. 166

A redução de 34 caracteres na legenda 1 foi obtida substituindo “Somente este ano” e “no estado do Ceará” por expressões mais curtas “Em 2006” e “Ceará”. Em L2, também houve a substituição de “no entanto” por “” por “mas”. Nessas duas legendas, a fim de ganhar mais espaço, os números por extenso foram escritos em algarismos arábicos. Finalmente, em L3 foi excluída a palavra “índices”, por ter sido considerada redundante e por não afetar a compreensão do conteúdo.

## 2. METODOLOGIA

O projeto CORSEL como um todo, incluindo o corpus das novelas aqui analisadas, consistiu numa pesquisa descritiva de cunho quanti-qualitativo, pautada tanto por análises baseadas em Linguística de *Corpus*, quanto pelos parâmetros vigentes da pesquisa em legendagem. O projeto foi conduzido pelo grupo LEAD, uma das linhas de pesquisa do grupo de pesquisa ‘Tradução e Semiótica’<sup>3</sup>. As coletas e análises se deram principalmente nas dependências do Laboratório de Tradução Audiovisual da Universidade Estadual do Ceará (LATAV-UECE).

Para a composição dos diferentes *subcorpora* do projeto CORSEL, foram selecionados três gêneros diferentes – novelas, documentários e seriados tipo comédias de situação –, haja vista serem os itens televisivos nos quais comumente se utilizam as legendas *off line* ou *pop-on* (que entram e saem em bloco da tela), cujo uso principal é na legendagem de programas gravados.

<sup>3</sup> <http://dgp.cnpq.br/dgp/espelhogrupo/3293257816353168>

A intenção original era obter programas de diferentes emissoras de TV, mas, devido à falta de regularidade dos canais em exibir programação legendada, optou-se novamente, como nas pesquisas de recepção anteriores (ARAÚJO, 2004, 2008), pela mesma emissora de televisão. O *corpus* para o gênero novela, portanto, consiste de quatro itens: AEO; CC; CP; e AB. O Quadro 4 mostra as datas em que os capítulos foram televisionados e gravados.

Quadro 4. Períodos de coleta das legendas do corpus

Novela	Período da gravação
AEO	18 a 23 de junho e 25 a 30 junho de 2012
CC	16 a 27 de abril de 2012
CP	18 a 21 de junho e 26 a 28 de junho de 2012
AB	27 a 31 de junho de 2012

Fonte: arquivos do projeto CORSEL

Os capítulos foram obtidos pelo registro do sinal televisivo numa mídia digital através de um gravador de DVD. Depois de coletados vários capítulos de cada novela, em alguns desses verificou-se a impossibilidade de visualizar as legendas, acarretando no descarte dos capítulos defeituosos. A experiência adquirida em pesquisas anteriores sugere que pode ter havido instabilidade do sinal veiculador das legendas. Contudo, também é possível ter havido erro na própria confecção das legendas pela emissora, sendo difícil detectar onde o erro aconteceu. Ao final do descarte, o corpus consistiu de cinco capítulos para cada novela.

As legendas foram extraídas automaticamente por meio do software *CCExtractor 0.61<sup>4</sup>*, o qual gera um arquivo de legendagem na extensão *srt*, capaz de registrar o número de inserção das legendas, bem como a marcação dos tempos inicial e final de cada uma delas. Os arquivos de legenda na extensão *srt* podem ser facilmente convertidos para um formato de texto puro, como a extensão *txt*, que é mais adequada para o *Wordsmith Tools 5.0*, software utilizado para a análise das etiquetas do CORSEL. A conversão pode ser feita por qualquer aplicativo básico de edição de texto, como o *Bloco de Notas* do sistema operacional Windows, por exemplo. A Figura 1 traz uma amostra de arquivo *.txt* com as legendas iniciais extraídas do capítulo 2 de AB.

4 <http://ccextractor.sourceforge.net/>

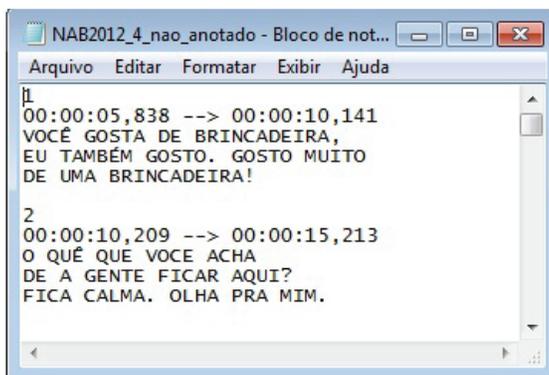


Figura 1. exemplo de arquivo txt aberto no Bloco de Notas com legendas extraídas pelo CCExtractor 0.61

Fonte: arquivos do CORSEL

Depois de convertidos para o formato txt, os arquivos sofreram um processo de etiquetagem, ou seja, o acréscimo de marcações (etiquetas) no corpo do texto para posterior análise. Essas etiquetas foram criadas com base no padrão *SGML* (*Standard Generalized Markup Language*), que fornece códigos escritos no formato <etiquetas de abertura>informação</etiqueta de fechamento/>. A Figura 2 retoma as legendas iniciais do capítulo 2 de AB, porém agora já etiquetadas.

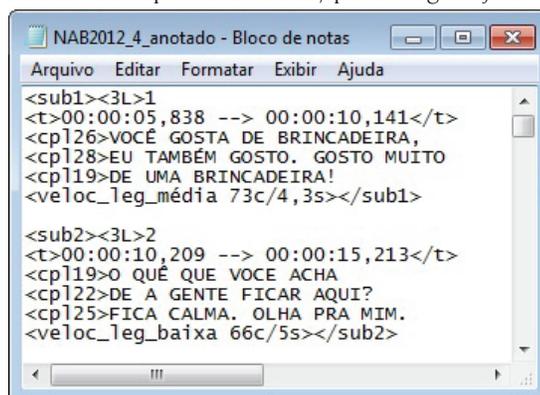


Figura 2. exemplo de arquivo txt após a etiquetagem das legendas

Fonte: arquivos do CORSEL

As etiquetas utilizadas foram de dois tipos: as etiquetas de parâmetros técnicos e as etiquetas de problemas de segmentação. O Quadro 5 resume as etiquetas de parâmetros técnicos.

Quadro 5. etiquetas para parâmetros técnicos da legendagem

Número da legenda	<sub1 <sup>5</sup> >legenda1</sub1 >
Linhas por legenda	<1L>, <2L> e <3L <sup>6</sup> >
Tempo inicial e final de cada legenda	<t>início --> final</t>
Número de caracteres por linha	<cpl>
Velocidade de legenda baixa (145ppm)	<veloc_leg_baixa>
Velocidade de legenda média (160ppm)	<veloc_leg_média>
Velocidade de legenda alta (180ppm)	<veloc_leg_alta>

Fonte: ARAÚJO, 2015

A etiqueta *número da legenda* delimita o início e o fim de uma legenda e permite encontrá-la mais facilmente por meio do mecanismo de busca do Bloco de Notas. As etiquetas da categoria *linhas por legenda* registram quantas linhas estão contidas na inserção de legenda. As etiquetas <t> e </t> delimitam o espaço indicador dos tempos inicial da legenda, enquanto <cpl> indica quantos caracteres existem em cada linha da legenda. Essa informação é então usada para calcular a velocidade da legenda, dividindo-se o total de caracteres da legenda pela sua duração.

Além das etiquetas de parâmetros técnicos, temos também as etiquetas para os problemas de segmentação linguística. A base de sustentação para a formulação desta segunda categoria foi a descrição gramatical do português oral culto do Brasil organizada por Castilho (2012), linguista teórico seguidor de uma vertente funcionalista-cognitivista. A escolha dessa perspectiva fundamenta-se na constatação de que essa vertente dos estudos linguísticos possui uma maior aproximação epistemológica com a metodologia baseada em corpus, bem como contempla melhor a natureza das legendas *closed caption*, na medida em que transcrevem quase literalmente o discurso oral dos programas de televisão.

5 Abreviatura de subtitle, 'legenda' em inglês.

6 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.

Em sua ‘Nova Gramática do Português Brasileiro’, Castilho (2012) detalha com grande riqueza os sintagmas e as orações<sup>7</sup> complexas<sup>8</sup> do português brasileiro. Para o autor (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 da 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 6, 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 oração.

Quadro 6. estrutura geral dos sintagmas e orações

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 Verbal (SV)	O SV é a construção que tem como núcleo o verbo. Como a oração é um verbo que articula seus argumentos, a única diferença entre o SV e uma oração é que no SV não figura o sujeito.
Sintagma Adjetival (SAdj)	O SAdj tem por Núcleo o adjetivo, que é uma classe basicamente predicadora, funcionando como adjunto adnominal, enquanto constituinte do sintagma nominal, ou como predicativo, enquanto constituinte do sintagma verbal (CASTILHO, 2012, p. 516). O SAdj tem como Especificador advérbios predicativos qualificadores e como Complementador SPs e orações substantivas objetivas.

7. 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.” (CASTILHO, 2012). 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 da preferência pela utilização de ‘sentença’. Essa 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 termo se torna menos obscura. Note-se que o termo ‘sentença’ nem mesmo consta na ‘Nomenclatura Gramatical Brasileira’, ao contrário de ‘oração’.

8. 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 subordinada’ (períodos compostos) pelos mesmos motivos supracitados.

Sintagma Preposicional (SP)	O SP tem por núcleo a preposição, por Especificador o advérbio, enquanto 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 e com os adjetivos, verbos e orações inteiras.
Orações complexas	Enunciados que possuem mais de um verbo, ou seja, contêm mais de uma oração, podendo estabelecer relação de coordenação, subordinação e correlação entre si.

Fonte: ARAÚJO e ASSIS (2014)

A descrição elaborada por Castilho (2012) permitiu que as etiquetas de problemas de segmentação linguística fossem elaboradas contemplando as recomendações de Karamitroglou (1998), segundo o qual, "quando segmentamos uma oração, forçamos o cérebro a pausar seu processo linguístico por um instante, até os olhos localizarem a próxima informação linguística" (p. 9-10). Logo, quando uma mesma legenda tiver que ser dividida em duas porções de texto em linhas distintas, "devemos tentar levar o cérebro a fazer essa pausa num ponto onde a carga semântica se configure numa informação completa" (p 9-10).

O autor organiza essas ideias utilizando a descrição gramatical de base gerativista, que representa as relações entre constituintes da oração através de nódulos sintáticos dispostos visualmente em esquemas arbóreos. Nesse sentido, diz-se que a quebra de linha numa legenda deve ocorrer nos mais altos níveis sintáticos. Porém, a descrição gerativista parece não lidar muito bem com a língua oral em uso, haja vista suas inúmeras pausas, hesitações, reformulações etc. Dessa sorte, a operacionalização analítica das recomendações de Karamitroglou (1998) só foi possível pela elaboração de etiquetas baseadas em Castilho (2012) e sua descrição do português oral culto do Brasil. O Quadro 7 lista as etiquetas de problemas de segmentação agrupadas por tipo de sintagma e de oração.

Quadro 7. Etiquetas para problemas de segmentação linguística em legendagem

ETIQUETAS DE ANÁLISE DO SINTAGMA NOMINAL (SN)
<SN_especif+SN>
<SN_subs+SAdj>
<SN_subs+SP>
<SN núcleo>

ETIQUETAS DE ANÁLISE DO SINTAGMA VERBAL (SV)
<SV_composto>
<SV_verbo+SAdv>
<SV_(verbo)+oblíquo+verbo>
<SV_verbo+SP>
ETIQUETAS DE ANÁLISE DO SINTAGMA ADJETIVAL (SAdj)
<SAdj_especif+SAdj>
ETIQUETAS DE ANÁLISE DO SINTAGMA PREPOSICIONAL (SP)
<SP_prep+SN>
<SP_prep+SV>
<SP_prep+SAdj>
ETIQUETAS DE ANÁLISE DO SINTAGMA ADVERBIAL (SAdv)
<SAdv_adv+adv>
ETIQUETAS DE ANÁLISE DE ORAÇÃO SUBORDINADA (SUBORD)
<SUBORD_conj/pron_rel+oração>
ETIQUETAS DE ANÁLISE DE ORAÇÃO COORDENADA (COORD)
<COORD_conj+oração>

Fonte: Arquivos do CORSEL

Terminado o processo de etiquetagem, as etiquetas foram analisadas pelo *Wordsmith Tools 5.0.*, o qual possui três ferramentas principais: *Wordlist*, *Concord* e *Keywords*. No projeto CORSEL, no entanto, apenas duas ferramentas foram necessárias: o *Wordlist*, para a obtenção do número de *tokens*<sup>9</sup> no subcorpus do gênero novela; e o *Concord*, para buscar o número de ocorrências de cada etiqueta e, no caso dos problemas de segmentação, mostrar o contexto das ocorrências encontradas para cada categoria.

### 3. RESULTADOS E DISCUSSÕES

Primeiramente, apresentamos o resultado das análises dos problemas de segmentação linguística na LSE do *subcorpus* das telenovelas, aliando os números gerados pelas ferramentas do *Wordsmith Tools* à interpretação qualitativa dos cotextos em que ocorrem os problemas de segmentação linguística. Em seguida, apresentamos possíveis ressegmentações dos segmentos problemáticos usados como exemplos.

9. "Palavras, ocorrências ou itens (*tokens*): contagem de palavras corridas (*running words*); cada palavra conta como uma ocorrência, mesmo que seja repetida". (BERBER SARDINHA, 2004, p. 165)

Uma vez analisado pelo *Wordsmith Tools*, o *corpus* apresentou um total de 7.703 inserções de legendas englobando 60.710 *tokens*, podendo ser considerado um *corpus* pequeno<sup>10</sup>. Poder-se-ia levantar uma questão sobre a representatividade do *corpus*; contudo, a representatividade de uma amostra só pode ser aferida estatisticamente caso se conheça o tamanho da população total de onde a amostra foi extraída. Como no caso das línguas naturais é impossível determinar o total de textos escritos e/ou falados numa dada cultura, não há um critério objetivo para determinar o tamanho ideal de um *corpus* representativo. A Tabela 2 apresenta em maiores detalhes, para cada novela: (i) o total de legendas, (ii) o número de *tokens* encontrados, (iii) o valor absoluto (VA) de problemas de segmentação (PROSEGL), (iv) o percentual de legendas problemáticas em relação a VA, e (v) o Índice de Frequência Simples<sup>11</sup> (IFS) de problemas de segmentação em cada novela.

Tabela 2. Dados quantitativos sobre número de legendas e de *tokens* em cada novela

Novela	Nº de legendas	Tokens	PROSEGL		
			VA	%	IFS (problemas de segmentação a cada mil palavras)
AEO	1849	14.856	333	17,9	22,4
CC	1725	13.112	243	14,0	18,5
CP	1282	9.744	324	25,2	33,2
AB	2847	22.998	310	10,8	13,5
TOTAL	7703	60.710	1210	15,7	19,9

Fonte: arquivos do CORSEL

Para as novelas, 15,7% das legendas apresentaram problemas de segmentação linguística e, se levarmos em conta o número de *tokens*, veremos que, de acordo com o IFS, a cada 1000 palavras ocorrem 19,9 problemas de segmentação. Se fôssemos considerar apenas esses parâmetros, parece que esses problemas não poderiam comprometer tanto a recepção; porém, conforme apontam as pesquisas

10. Para Berber Sardinha (BERBER SARDINHA, 2004, p. 26), um *corpus* é considerado pequeno quando o total de palavras for menor ou igual a 80 mil palavras.

11. A divisão do número de ocorrências de uma dada categoria pelo número total de palavras do *corpus* usado como referência, com o resultado multiplicado por 1.000 (PRAXEDES FILHO; MAGALHÃES, 2013a, p. 80). No contexto deste trabalho, o IFS serve para compensar os números de palavras diferentes nas LSEs das quatro novelas analisadas. Esse recurso estatístico foi incorporado ao Projeto CORSEL em 2014, com o projeto em andamento, de modo que fosse possível comparar os *subcorpora* localizados no seio do projeto, os quais tinham tamanhos distintos, sendo assim difícil, às vezes, comparar os resultados provenientes da análise dos *subcorpora* entre si por meio de porcentagens (ARAÚJO, 2015; ARRAES, 2015).

de recepção realizadas no âmbito do LEAD (ARAÚJO; NASCIMENTO, 2011; ARAÚJO; VIEIRA; MONTEIRO, 2013), um cuidado com a segmentação pode ter sido a razão pela qual os surdos tiveram boa recepção em legendas de velocidade alta (180ppm). Com relação à distribuição das 1210 legendas com problemas de segmentação pelas categorias de sintagmas e orações, houve maior incidência de problemas nos sintagmas verbal (SV) e nominal (SN), como mostra a Tabela 3.

Tabela 3. Valores absolutos e percentuais de PROSEGL do gênero novela

Novela	SV	SN	SAdj	SAdv	SP	COORD	SUBORD
AEO	146	109	6	4	32	6	30
CC	85	80	2	6	31	12	27
CP	150	61	8	1	27	22	55
AB	149	71	10	3	36	5	36
TOTAL 1 (VA)	530	321	26	14	126	45	148
TOTAL 2 (%)	44,35	26,86	2,17	1,17	10,54	3,76	12,38

Fonte: Arquivos do CORSEL

Os resultados corroboram aqueles de todas as pesquisas brasileiras que se debruçaram sobre a segmentação na legendagem de filmes e novelas (CHAVES, 2012; DINIZ, 2012; ASSIS, 2013, 2016; GABRIEL, 2013; 2015; CHAVES; ARAÚJO, 2014; ARAÚJO; ASSIS, 2014; ARRAES, 2015), ou seja, que os problemas de segmentação mais comuns na legendagem desse gênero estão na quebra do sintagma verbal. Houve também problemas de segmentação no nível dos sintagmas nominal e preposicional, assim como na quebra das orações subordinadas. Os resultados apontam para uma maior preocupação com a segmentação linguística desses tipos de estrutura, com maior ênfase no SV.

A seguir, discriminamos os resultados da análise linguística para cada tipo de problema de segmentação, enfatizando as categorias de SV e SN, dado que estes apresentaram maior número de problemas. Dessa forma, exemplificamos os nossos dados com dois exemplos desses sintagmas, e utilizamos apenas um exemplo na discussão dos outros sintagmas e das orações.

### 3.1. SINTAGMA VERBAL (SV)

Dos 530 problemas ocorridos no sintagma verbal, 285 (53,8%) foram do tipo <SV\_composto>, 175 (33%) foram do tipo <SV\_verbo+SAdv>, 62 (11,7%) foram do tipo <SV\_(verbo)+oblíquo+verbo> e 8 (1,5%) problemas do tipo <SV\_verbo+SP>. Para exemplificar os casos de problemas de segmentação linguística no Sintagma Verbal, tomemos uma ocorrência de <SV\_composto> no *corpus*, apresentado no Quadro 8. Esse problema ocorreu na legenda 89 do Capítulo 5 de AEO, que possui um sintagma verbal composto – ‘queria trazer’ – mal segmentado entre as duas linhas da legenda em questão.

Quadro 8. Problema de segmentação no sintagma verbal (SV\_composto)

TCR <sup>15</sup>	No de linhas	Duração	Vel.	Legenda
00:06:28,754 --> 00:06:31,555	2	2,8s	180ppm (alta)	EU NÃO QUERIA <u>TRAZER</u> PROBLEMA PRA VOCÊS

Fonte: Arquivos do CORSEL

Considerando o parâmetro da segmentação linguística, a legenda apresentada no Quadro 8, ressegmentada, ficaria da seguinte forma:

Eu não queria trazer  
problema para vocês.

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

Em relação aos problemas de segmentação abarcados pela subetiqueta <SV\_verbo+SAdv>, a segunda categoria com mais ocorrência no nível do sintagma verbal, o Quadro 9 traz como exemplo a legenda 164 do capítulo 2 de AEO.

<sup>15</sup> Sigla para Time Code Reader, que indica o tempo de entrada e saída da legenda.

Quadro 9. Problema de segmentação no sintagma verbal (SV\_verbo+SAdv)

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

Fonte: Arquivos do CORSEL

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

[Rodrigo]  
O Mauro se doou a isso.

### 3.2. SINTAGMA NOMINAL (SN)

Dos 321 problemas ocorridos no sintagma nominal, 176 (54,8%) foram do tipo <SN\_especif+SN>, 72 (22,4%) foram do tipo <SN\_subs+SP>, 56 (17,5%) foram do tipo <SN\_subs+SAdj> e 17 (5,3%) problemas do tipo <SN\_núcleo>. Vejamos o exemplo de estrutura <SN\_especif+SN> (legenda 36 do Capítulo 4 de AB), categoria com problemas no nível do sintagma nominal, no Quadro 10, em que há quebra entre o núcleo nominal e seu especificador. Neste caso, podemos ver a quebra entre o pronome “minha” e o substantivo “mulher”.

Quadro 10. Problema de segmentação no sintagma nominal (SN\_especif+SN)

TCR	No. de linhas	Duração	Vel.	Legenda
00:03:21,834 → 00:03:24,000	2	3,1s	180ppm  (Alta)	VOCÊ É MUITO MAIS QUE <u>MINHA</u>  <u>MULHER</u> , VOCÊ É MINHA <u>PARCEIRA</u> .

Fonte: Arquivos do CORSEL

Quando o personagem que toma a palavra se dirige ao seu interlocutor, o pronome “você” pode ser depreendido sem maiores problemas, não havendo necessidade do segundo pronome. Ademais, a prática comum de transcrição do diálogo inteiro acaba saturando o número de caracteres por linha. Para sanar este complicador e diminuir as chances de um segundo problema de segmentação, optou-se por condensar por omissão a legenda, retirando o advérbio ‘muito’. Para este caso, propõe-se a seguinte ressegmentação, na qual os 58 caracteres anteriores se transformaram em 46.

Você é mais que minha mulher,  
é minha parceira.

Passando para um exemplo de quebra da estrutura <SN\_subst+SAdj>, tome-se como exemplo a legenda 13 do Capítulo 2 da mesma novela, na qual o substantivo “Mauricinho” está separado do seu complementador “problemático”. Este exemplo está no Quadro 11.

Quadro 11. Problema de segmentação no sintagma nominal (SN\_subst+SAdj)

TCR	No. de linhas	Duração	Vel.	Legenda
00:00:49,800 --> 00:00:55,000	3	5,2s	145ppm (baixa)	MUITO MAIS FELIZ DO QUE ESSE MAURICINHO <u>PROBLEMÁTICO</u> DO JORGINHO

Fonte: Arquivos do CORSEL

Apesar da velocidade baixa devido aos generosos cinco segundos de duração, a legenda apresenta gordura textual característica da oralidade, a qual dificulta sanar o problema de segmentação sem ultrapassar o número máximo de caracteres por linha. Uma solução seria a seguinte:

Muito mais feliz  
que o mauricinho-problema do Jorginho.

A presença do adjetivo “feliz” precedido de advérbio de intensidade “mais”, bem como seguido da conjunção “que”, configura uma clássica estrutura de comparação, sendo desnecessária a presença da preposição “do”. Ademais, tendo em vista as restrições sobre o número de caracteres por linha, é interessante achar soluções que ocupem menos caracteres. Nesse sentido, o pronome demonstrativo “esse” pode ser substituído pelo artigo definido “o”, bem como “problemático” pode ser substituído por “-problema”.

### 3.3. SINTAGMA PREPOSICIONAL (SP)

Em relação aos problemas no nível do sintagma preposicional, a distribuição dos problemas entre categorias ficou da seguinte maneira: em um total de 126 problemas, 103 (81,7%) foram casos de <SP\_prep+SN> e 23 (18,3 %) casos de <SP\_prep+SV>. Um exemplo de quebra na estrutura <SP\_prep+SN>, ocorrência mais frequente no nível do sintagma preposicional, é a legenda 178 do capítulo 3 de AB presente no Quadro 12.

Quadro 12. Problema de segmentação no sintagma preposicional (SP\_prep+SN)

TCR	No. de linhas	Duração	Vel.	Legenda
00:32:07,825 --> 00:32:07,825	3	2,1s	180ppm (alta)	[JORGINHO] EU ESTAVA COM UM MEGA PROBLEMA NA MINHACASA. O QUÊ QUE VOCÊ QUERIA?

Fonte: Arquivos do CORSEL

Nesta legenda, há uma quantidade grande de caracteres por linha (78 caracteres, quando deveria ter no máximo 43), impedindo a mera realocação de uma palavra, além de sua duração ser bastante curta. Logo, há que ser condensada por omissão, como mostra um exemplo de possível ressegmentação abaixo.

[Jorginho] Tava com problema.  
Que você quer?

Haja vista a busca dos roteiristas da novela por emularem a oralidade do português brasileiro, não causaria prejuízo à legenda utilizar uma versão coloquial do verbo “estar”, a qual permitiria reduzir o número de caracteres por linha. Pode-se diminuir o número de caracteres também pela eliminação do adjetivo “mega” e do SP “minha casa”, além da condensação por meio de paráfrase da oração “O que você queria”.

### 3.4. SINTAGMA ADJETIVAL (SAdj)

Em relação à quebra no nível do sintagma adjetival, houve 26 ocorrências no corpus. Um exemplo de quebra relativa à categoria <SAdj\_especif+SAdj> presente na legenda 157 do Capítulo 5 de Amor Eterno Amor é apresentada no Quadro 13.

Quadro 13. Problema de segmentação no sintagma adjetival (SAdj\_especif+SAdj)

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

Fonte: Arquivos do CORSEL

Na ocorrência em questão, vemos a separação entre o Especificador “bem mais” e seu núcleo, o SAdj “tranquilos”. Para a ressegmentação, privilegiando uma segmentação linguística adequada, teríamos que fazer poucos ajustes de redução, eliminando repetições e redundâncias para que obtivéssemos o exemplo abaixo, de 54 caracteres.

Estamos mais tranquilos vendo o senhor,  
assim, forte, contente.

### 3.5. SINTAGMA ADVERBIAL (SAdv)

Houve 14 ocorrências de quebra no nível do sintagma adverbial no corpus. A legenda 126 do Capítulo 5 de CC (Quadro 13) ilustra esse tipo de quebra.

Quadro 14. Problema de segmentação no sintagma adverbial (SAdv\_adv+adv)

TCR	No. de linhas	Duração	Vel.	Legenda
00:27:15,032 --> 00:27:17,300	2s	2,2s	145ppm  (baixa)	[SAMUEL] <u>AINDA</u> <u>POR CIMA É DEDO DURO</u>

Fonte: Arquivos do CORSEL

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

[Samuel]

Ainda por cima é dedo duro.

### 3.6. ORAÇÕES COORDENADAS E SUBORDINADAS

Quanto às quebras no nível das orações, houve 148 casos em orações subordinadas e 45 ocorrências em coordenadas. Exemplificando o problema de segmentação em orações subordinadas, temos o exemplo da legenda 21 do capítulo 1 de CC, na qual ocorre separação da oração subordinada entre a segunda linha (que eu) e a terceira linha (tinha que pagar nessa casa). Além do problema de segmentação linguística, essa legenda apresenta inadequações em número de linhas (3 linhas) e caracteres (74) para o tempo de duração da legenda (3,4s).

Quadro 15. Problema de segmentação nas orações (SUBORD\_conj/pron\_rel+oração)

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

Fonte: Arquivos do CORSEL

Para essa legenda, os 3,4 segundos de duração da legenda só comportam 61 caracteres, havendo a necessidade de redução. O português brasileiro permite a construção de orações com sujeito oculto, ao que elidimos os dois pronomes ‘eu’ da legenda. Ademais, o constituinte ‘mesmo’ é apenas enfático neste trecho e sua supressão não prejudica a compreensão da mensagem. Depois dessas modificações, a oração subordinada se mantém completa na segunda linha, solucionando a quebra. Ainda nessa nova legenda, mantemos um formato apropriado, além de permanecer dentro dos parâmetros já preconizados e já discutidos em nossas análises anteriores.

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

### 3.7. RELAÇÃO DAS CARACTERÍSTICAS TÉCNICAS COM OS PROBLEMAS DE SEGMENTAÇÃO LINGUÍSTICA

No que diz respeito à relação dos problemas de segmentação linguística com os parâmetros técnicos de número de linhas e de velocidade de legenda, a Tabela 4.

Tabela 4. Relação entre PROSEGL e parâmetros técnicos do Gênero Novela

Novela	Legenda de 2L	Legenda de 3L	145ppm Vel. Baixa	160ppm Vel. Média	180ppm Vel. Alta
AEO	150	182	143	14	175
CC	115	127	75	11	156
CP	161	163	170	16	138
AB	131	178	77	23	209
TOTAL 1 (VA)	557	650	465	64	678
TOTAL 2 (%)	46,1	53,9	38,5	5,3	56,2

Fonte: Arquivos do CORSEL

No que diz respeito à relação dos problemas de segmentação linguística encontrados no corpus e o parâmetro de número de linhas, 557 (46,1%) dos problemas foram encontrados em legendas de 2 linhas e 650 (53,9%) em legendas de 3 linhas. Em relação aos problemas de segmentação linguística e à velocidade da legenda, 678 (56,2%) foram em legendas com velocidade alta, 465 (38,5%) em legendas com velocidade baixa e 64 (5,3%) em legendas de velocidade média.

Os resultados corroboram os de outras pesquisas sobre problemas de segmentação linguística, com exceção de Chaves (CHAVES, 2012; CHAVES; ARAÚJO, 2014), o que pode ser explicado devido ao fato de a pesquisadora ter

investigado a LSE em filmes em DVD, cuja legendagem recebe um controle de qualidade mais rigoroso. Em sua análise do filme ‘Nosso Lar’, Chaves (2012) encontrou poucos problemas de segmentação, além de não ter se defrontado com legendas rápidas. Esse não foi o caso dos pesquisadores que trabalharam com o gênero novela (ASSIS, 2016; GABRIEL, 2013), os quais se defrontaram com resultados próximos aos aqui apresentados.

## CONSIDERAÇÕES FINAIS

Ao definir parâmetros específicos para descrever a segmentação linguística na legendagem, o projeto CORSEL desenvolveu uma ferramenta que tanto mostra a legendistas profissionais como ocorrem os problemas de segmentação como instrumentaliza o legendista aprendiz com diretrizes para distribuição das falas nas legendas. A proposta de seguir os mesmos parâmetros da legenda para ouvintes, levando em conta as especificidades da LSE, pode contribuir para a melhoria da oferta de produtos audiovisuais na TV para a comunidade surda brasileira.

A interface entre a Linguística de Corpus e a TAV permitiu a criação de uma metodologia para a análise da LSE em língua portuguesa por meio da etiquetagem com parâmetros técnicos de problemas de segmentação. Essa metodologia permite analisar, além da segmentação linguística, as estratégias para a edição de uma legenda quando isso se fizer necessário. Por possibilitar a análise de uma grande quantidade de dados provenientes de três gêneros televisivos diferentes (novelas, seriados e documentários) e apresentar resultados conclusivos sobre cada um deles, fazemos uma proposta viável para a segmentação não só da LSE, mas de qualquer legendagem em língua portuguesa.

Quanto ao *corpus* analisado, tanto a grande quantidade de problemas de segmentação linguística, quanto inadequações em termos de concepção de tradução e de parâmetros técnicos concernentes à LSE indicam a necessidade de melhorias na confecção dessas legendas por parte dos veículos que oferecem esse produto audiovisual. Aliada à janela de LIBRAS, a LSE do tipo *closed caption* oferecida pela televisão brasileira é uma das maneiras pelas quais os surdos podem ter acesso a diversos itens da programação televisiva. Dessa forma, as pesquisas que caracterizam essa modalidade de TAV e que propõem critérios de aperfeiçoamento são importantes para que ocorram transformações na prática levada a cabo pelos canais de TV. Mais pesquisas são necessárias, para que sejam objetivamente mensurados os efeitos concretos dos problemas de segmentação linguística na compreensão da LSE por surdos, como fizeram Monteiro (2016) e Vieira (2016),

cujas pesquisas apontam para o efeito negativo da má segmentação na recepção de legendas por surdos e ouvintes brasileiros.

---

## REFERÊNCIAS BIBLIOGRÁFICAS

- ARAÚJO. (2004). Closed subtitling in Brazil. In: ORERO, P. (Org.). *Topics in Audiovisual translation*. Amsterdã: John Benjamins. pp. 199-212.
- ARAÚJO. (2007). Subtitling for the deaf and hard-of-hearing in Brazil In: *Media for All: Subtitling for the Deaf, Audio Description and Sign Language*. Kenilworth., Nova Jersey, EUA: Rodopi, 2007, v.30, 99-107.
- ARAÚJO. (2008). Por um modelo de legendagem para Surdos no Brasil. *Tradução e comunicação: Revista Brasileira de Tradutores*, n. 17, pp. 59–76. Disponível em: <http://sare.anhanguera.com/index.php/rtcom/article/view/148/147>. (24/03/2016).
- ARAÚJO. (2009). In Search of SDH Parameters for Brazilian Party Political Broadcasts. In: *The Sign Language Translator and Interpreter*, Manchester: St. Jerome Publishing Company, v. 3, n. 2, p. 157-167.
- ARAÚJO. (2012). *Legendagem para surdos: em busca de um modelo para o Brasil*. Relatório Técnico n. 306948/2008-7. Fortaleza: CNPq. Fev/2012. Não publicado.
- ARAÚJO. (2015). *A segmentação na legendagem para surdos e ensurdecidos (LSE): um estudo baseado em corpus*. Relatório Técnico n. 306441/2011-0. Fortaleza: CNPq. Fev/2015. Não publicado.
- ARAÚJO., NASCIMENTO, A. K. P. (2011). Investigando parâmetros de legendas para Surdos e Ensurdecidos no Brasil. *Tradução em Revista*, n. 11, pp. 1-18. Disponível em: <http://www.maxwell.lambda.ele.puc-rio.br/18862/18862.PDF>. (24/03/2016).
- ARAÚJO., VIEIRA, P. A.; MONTEIRO, S. M. M. (2013). Legendagem para surdos e ensurdecidos (LSE): um estudo de recepção com surdos da região sudeste. *TradTerm*, v. 22, pp. 283-302. Disponível em: <http://www.revistas.usp.br/tradterm/article/view/69132/71589>. (24/03/2016).
- ARAÚJO; ASSIS. (2014). A segmentação linguística na legendagem para surdos e ensurdecidos (LSE) de ‘Amor Eterno Amor’: uma análise baseada em corpus. *Letras & Letras*, v. 30, n. 2, jul/dez, pp. 156-184. Disponível em: <http://www.seer.ufu.br/index.php/letraseletras/article/download/28562/15798>. (19/01/2015).

- ARRAES. (2015). *Legendagem para surdos e ensurdecidos: análise baseada em corpus da segmentação linguística do filme "Virada Radical"*. 71f. Trabalho de conclusão de curso (Bacharelado em Letras). Curso de Letras, Universidade Estadual do Ceará, Fortaleza.
- ASSIS. (2013). *A segmentação na LSE de amor eterno amor: uma análise baseada em corpus*. 2013. 72f. Trabalho de conclusão de curso (Bacharelado em Letras). Curso de Letras, Universidade Estadual do Ceará, Fortaleza.
- ASSIS. (2016). *Legendagem para Surdos e Ensurdecidos (LSE): análise baseada em corpus da segmentação linguística em Amor Eterno Amor*. 2015. 107f. Dissertação (Mestrado em Linguística Aplicada). Universidade Estadual do Ceará, Fortaleza, 2016. Disponível em: [http://www.uece.br/posla/dmdocuments/Disserta%C3%A7%C3%A3o\\_%C3%8Dtal0%20Alves.pdf.pdf](http://www.uece.br/posla/dmdocuments/Disserta%C3%A7%C3%A3o_%C3%8Dtal0%20Alves.pdf.pdf). (20/10/2016).
- BERBER SARDINHA, T. (2004). *Linguística de Corpus*. Barueri, São Paulo: Manole.
- CASTILHO, A. T. (2012). *Nova gramática do português brasileiro*. São Paulo: Contexto.
- CHAVES, E. G. (2012). *Legendagem para Surdos e Ensurdecidos: 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.
- CHAVES, E. G.; ARAÚJO. (2014). 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, pp. 62-75.
- DIAZ-CINTAS, J. (2010). Subtitling. In: GAMBIER, Y.; DOORSLAER, D. (Eds.). *Handbook of Translation Studies*. Amsterdam/Philadelphia: John Benjamins, pp. 344-349.
- DIAZ-CINTAS, J.; REMAEL, A. (2007). *Audiovisual Translation: Subtitling*. Manchester, UK, Kinderhook, NY, UK: St. Jerome Publishing.
- DINIZ, N. S. L. (2012). *A Segmentação em Legendagem para Surdos e Ensurdecidos: um Estudo Baseado em Corpus*. 149f. Dissertação (Mestrado em Linguística). Universidade Federal de Minas Gerais, Belo Horizonte – MG.
- D'YDEWALLE, G.; PRAET, C.; VERFAILLIE, K.; VAN RENSBERGEN, J. (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In: O'REGAN, J. K.; LÉVY-SCHOEN, A. (Eds.).

- Eye Movements: From Physiology to Cognition*. Amsterdam, New York: Elsevier Science Publishers, pp. 313-321.
- FRANCO, E.; ARAÚJO. (2003). Reading Television: Checking deaf people's Reactions to Closed Subtitling in Fortaleza, Brazil. *The Translator: studies in intercultural communication*, v. 9, n. 2, pp. 249-267. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/13556509.2003.10799156?src>. (15/03/2015).
- FRANCO, R. P.; CHAVES, E. G. (2015). Legendagem para Surdos e Ensurdecidos (LSE): Um Estudo da Segmentação Linguística nas Séries da TV Brasileira. *Mosaico*, São José do Rio Preto, v. 14, pp. 247-265.
- GABRIEL, M. H. C. (2013). *A segmentação linguística na legendagem para surdos e ensurdecidos (LSE): uma análise baseada em corpus*. 2013. 59f. Trabalho de conclusão de curso (Especialização em Estudos da Tradução). Universidade Federal do Ceará, Fortaleza.
- GABRIEL, M. H. C. (2015). *Problemas de segmentação linguística na legendagem para surdos e ensurdecidos (LSE) de "Cheias de Charme": uma análise baseada em corpus*. 109f. Dissertação (Mestrado em Linguística Aplicada). Universidade Estadual do Ceará, Fortaleza, 2015. Disponível em: <http://www.uece.br/posla/dmdocuments/DISSERTA%C3%87%C3%83O%20-%20MARIA%20HELENA%20CLARINDO%20GABRIEL.pdf>. (10/01/2016).
- IVARSSON, J.; CARROLL, M. (1998). *Subtitling*. Simrishamm, Suécia: TransEditHB.
- KARAMITROGLOU, F. (1998). A Proposed Set of Subtitling Standards in Europe. *Translation Journal*, v. 2, n. 2, pp. 1-15. Disponível em: <http://translationjournal.net/journal/04stndrd.htm>. (10/03/2014).
- MONTEIRO, S.M.M. (2016). *Legendagem para Surdos e Ensurdecidos (LSE): um estudo sobre a segmentação na LSE da campanha política de 2010*. 230f. Tese (Doutorado em Linguística Aplicada). Universidade Estadual do Ceará, Fortaleza-CE, 2016
- PRAXEDES FILHO, P. H. L.; MAGALHÃES, C. M. (2013) A neutralidade em audiodescrições de pinturas. In: Org. ARAUJO, V. L. S; ADERALDO, M. F. (Orgs). *Os novos rumos da pesquisa em audiodescrição no Brasil*. 1 ed. Editora CRV. Curitiba, PR. 2013a, p. 73-87, ISBN: 978-85-8042-592-5.
- REID, H. (1990). 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.

VIEIRA, P. (2016). A influência da segmentação e da velocidade na recepção de legendas para surdos e ensurdecidos. 248f. Tese (Doutorado em Linguística Aplicada). Universidade Estadual do Ceará, Fortaleza-CE, 2016.

VIEIRA, P.; ARAÚJO, CHAVES, E. G.; MONTEIRO, S. M. M. (2015). A segmentação na legendagem do documentário Globo Repórter: uma análise baseada em corpus. In: PONTES, V. O.; CUNHA, R. B.; CARVALHO, E. P.; TAVARES, M. G. G. (Orgs.). *A tradução e suas interfaces: múltiplas perspectivas*. Curitiba: CRV, 2015: 65-86.

Recebido: 14/05/2017

Aceito: 01/08/2017