

ETIQUETAS DE SEGMENTAÇÃO: UMA PROPOSTA PARA ANÁLISE DA SEGMENTAÇÃO EM LEGENDAS INTRALINGUÍSTICAS DE FILMES BRASILEIROS

CONTEXTUALIZAÇÃO

Este trabalho se insere no campo dos estudos da tradução baseados em *corpora* (ETBC) inspirado pelos estudos seminais de Baker (1996). No âmbito da tradução audiovisual (TAV), principalmente na área de legendagem, os trabalhos baseados em *corpus* são recentes e pouco numerosos, no entanto a interação entre essas duas áreas já provou ser eficiente e promissora. O foco do presente trabalho consiste numa proposta de etiquetagem para a investigação da segmentação nas legendas intralinguísticas de filmes em língua portuguesa (variante brasileira). A motivação surgiu da inquietação provocada pela falta de parâmetros sistemáticos para análise e produção da segmentação na legendagem, uma vez que a segmentação é determinante na compreensão de legendas por parte dos espectadores.

A segmentação diz respeito ao parâmetro da legendagem relacionado à divisão e distribuição de legendas, que ocorre tanto entre duas legendas diferentes, quanto dentro da mesma legenda. Neste último, é chamada de quebra de linha. Para se segmentar uma fala em legendas, pode-se adotar três critérios: 1) linguístico - pautado pela sintaxe, ou seja, cada linha de legenda deve conter um pensamento o mais completo possível; 2) retórico - pelo fluxo da fala, ou seja, após uma pausa, uma nova legenda deve ser produzida; 3) visual - pautado pelo corte de cena, ou seja, sempre que a mudança de cena coincidir com o tempo de saída de uma legenda, esta não deve durar até a cena seguinte. Algumas pesquisas sugerem que uma segmentação fora desses padrões pode causar problemas de recepção, pois o espectador pode se cansar mais rapidamente e perder o interesse em continuar assistindo ao filme devido o esforço despendido (KARAMITROGLOU, 1998; DÍAZ-CINTAS; REMAEL, 2007).

Com o objetivo de testar a aplicabilidade dessas etiquetas, foi realizada uma análise eletrônica da segmentação na legendagem para surdos e ensurdecidos (LSE) do filme brasileiro em DVD, *Nosso Lar* (2010). A escolha de *Nosso Lar* se deve ao fato de que esse foi um dos últimos filmes em DVD lançado no Brasil que ofereceu tal recurso. A maioria dos filmes traz legendas em português, mas não se configuram como legenda

para surdos, pois para isso deve conter informações que só podem ser recuperadas pelo canal auditivo (tradução de efeitos sonoros e identificação do falante).

REVISÃO DE LITERATURA

Sobre a segmentação na legendagem, Perego (2008) investigou a segmentação linguística entre linhas de legendas de um *corpus* heterogêneo de legendas de filmes, cuja análise focou casos de má segmentação baseando-se nos mais altos níveis sintáticos propostos por Karamitroglou (1998), o qual sugere que a quebra de linha deve acontecer num ponto da legenda em que a carga semântica se configure num pensamento completo, ou seja, no mais alto nível sintático possível. A partir desse ponto, Perego (2008) começou a investigar a natureza dos problemas de segmentação na legendagem de filmes diversos e observou que os problemas aconteciam nos sintagmas nominal, preposicionado, verbal e nas orações coordenada e subordinada.

Sobre o processo de etiquetagem Berber Sardinha (2004, p. 145) afirma que este auxilia na desambiguação lexical e permite a descrição de padrões léxico-gramaticais. Em outras palavras a etiquetagem consiste numa codificação do *corpus* passível ou não de análise eletrônica através de concordanciadores. No glossário de Linguística de *Corpus* elaborado por Tagnin (2010) os níveis de etiquetagem linguística são definidos como morfossintático, sintático, semântico e discursivo, no entanto as etiquetas de segmentação propostas são específicas para a análise da segmentação em legendas em língua portuguesa e portanto não se enquadram inteiramente nos tipos encontrados na literatura.

METODOLOGIA

A etiquetagem do presente *corpus* foi feita manualmente, sem auxílio de programas etiquetadores, no editor de textos Microsoft Word 2007 e salva como texto sem formatação (formato .txt). São 4 as etiquetas indicativas de problema de segmentação, que informam se o problema foi de natureza linguística, retórica ou visual. São 19 as etiquetas de segmentação linguística que correspondem a problemas localizados nos níveis dos sintagmas nominal, verbal, preposicionado, adjetivo, adverbial e no nível das orações coordenada e subordinada.

ETIQUETAS DEFINIDAS

Etiquetas indicativas de problema de segmentação

- Indicativa de problema de segmentação linguística (gramatical) <PROSEGG>
- Indicativa de problema de segmentação retórica, porque antecipou informação <PROSEGR_antecipouinformação>; <PROSEGR_atrasouinformação>
- Indicativa de problema de segmentação visual, porque a legenda vazou para a cena subsequente <PROSEGV_vazou>.

Problemas de segmentação no Sintagma Nominal

- Quando há quebra entre pre-nucleares e substantivo <SN_pre-nucleares+subst>
ex: <sub142>Pelo menos tira aquela/sensação de fome.
- Quando há quebra entre nominal e modificador, ou na ordem inversa, modificador e nominal <SN_nominal+modif/modif+nominal>
ex: <sub227>Seu aparelho/gastrointestinal estava
- Quando há quebra entre superlativo e adjetivo <SN_superlativo+adj>
ex: <sub145>Muito mais/bem disposto, pelo visto.
- Quando há quebra entre relativo e oração incompleta <SN_relativo+oração incompleta>
ex: <sub288>A vida na Terra é que/é uma cópia daqui, André.
- Quando há quebra entre nome próprio <SN_nome próprio>
ex: Ela vai se chamar Ana/Rosa Belo Fernandes
- Quando há quebra entre título e nome próprio <SN_título+nome próprio>
ex: <sub389>Você pode procurar o irmão/Genésio no Ministério do Auxílio,
- Quando há quebra da estrutura interna de colocações, expressões idiomáticas e convencionais <SN_colocações/idiom/conv>
ex: <sub98>que vinha do fundo/da minha alma, mas fui ouvido.

Problemas de segmentação no Sintagma Preposicionado

- Quando há quebra entre preposição e substantivo <SP_prep+subst>
ex: <sub523>O que sabe sobre/a medicina espiritual?

Problemas de segmentação no Sintagma Verbal

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

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

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

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

- Quando há quebra entre colocações verbais <SV_colocações>

ex: Vamos tomar/providências sobre o assalto

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

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

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

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

Problemas de segmentação no sintagma Adverbial

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

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

Problemas de segmentação no Sintagma Adjetivo

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

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

Problemas de segmentação na Oração Coordenada

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

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

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

ex: vamos conversar, mas não/quero brigas.

Problemas de segmentação na Oração Subordinada

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

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

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

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

RESULTADOS

Após a etiquetagem das 1132 legendas do filme *Nosso Lar*, foram identificados 88 problemas de segmentação em 87 legendas. Os resultados sugeriram que os maiores problemas de segmentação em *Nosso Lar* estão concentrados, em sua maioria, nos níveis dos sintagmas, em especial no sintagma verbal, quando há quebra da estrutura verbo + verbo. O gráfico abaixo ilustra o resultado da análise

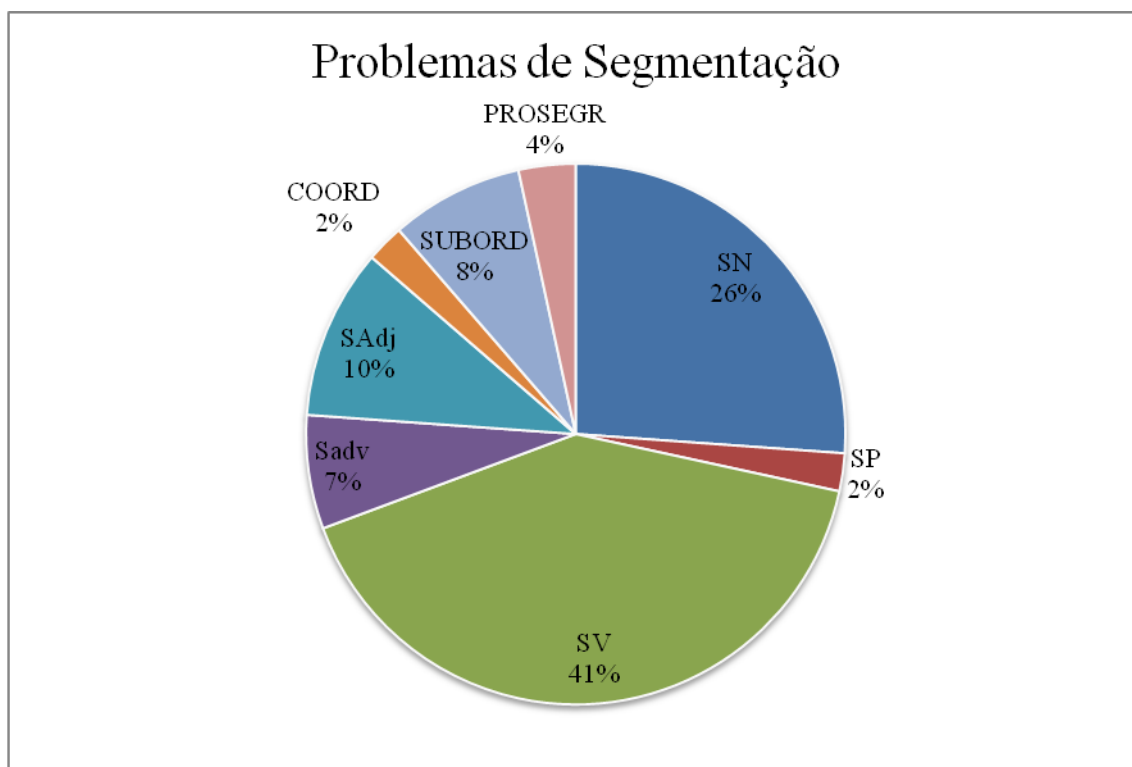


Gráfico com resultado geral dos problemas de segmentação na legendagem do filme *Nosso Lar*

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

CONSIDERAÇÕES FINAIS

É possível vislumbrar que os problemas de segmentação podem ser sanados a partir do desenvolvimento de estratégias de segmentação fundamentadas nas categorias linguísticas encontradas no *corpus*. Uma proposta de etiquetas de segmentação para análise de legendas é, além de inovador, de grande valia tanto para os estudos da tradução quanto para a linguística de *corpus*, uma vez que as etiquetas se tornam instrumentos de análise de segmentação em legendas.

Os resultados também sugerem que maior atenção deve ser atribuída à análise de constituintes no nível do sintagma, não excluindo os constituintes no nível da oração nem os problemas de segmentação retórica e visual. Desta forma, um melhor treinamento pode ser dado a tradutores em formação com ênfase na quebra dos sintagmas. Em conclusão, as etiquetas de segmentação definidas nesta pesquisa servem de recurso para profissionais legendistas, tradutores em formação e estudiosos da área que buscam desenvolver práticas de tradução mais conscientes e melhor fundamentadas.

REFERÊNCIAS

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

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

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

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

CHAVES, É. G. **Legendagem para Surdos e Ensurdecidos: Um estudo Baseado em Corpus da Segmentação nas legendas de filmes brasileiros em DVD**. 130f. Dissertação (Mestrado). Universidade Estadual do Ceará, Fortaleza-CE, 2012.

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

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

NOSSO LAR. Direção: Wagner de Assis. Brasil: Fox do Brasil, 2010. 1 DVD (102min), região 4, NTSC, color., legendas (para surdos em português) e audiodescrição.

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

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

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