# Análise Multimodal de Sentimentos para Estimação da Polaridade Tensiva de Notícias em Vídeos de Telejornais


## ABSTRACT

This paper presents a multimodal approach to perform content-based sentiment analysis in TV newscasts videos in order to assist in the automatic estimation of polarity tension of TV news. The proposed approach aims to contribute to the semiodiscoursive study relative to the construction of ethos of those TV shows. In order to achieve this goal, it is proposed the application of computational methods of state-of-the-art that, through the processing of newscasts' videos of interest, perform the automatic emotion recognition in facial expressions. Moreover, they extract modulations in the participants' speech (e.g., news anchors, commentators, reporters, among others) and apply sentiment analysis techniques in their text obtained from closed caption, therefore making possible to estimate the emotional tension level in the enunciation of the TV news. In order to evaluate the accuracy and the applicability of the system, we use an actual dataset composed by 358 videos from three Brazilian newscasts. The experimental results are promising, which indicate the potential of the approach to support the analysis of TV newscasts discourse.



## Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Audio input/output, Evaluation/methodology, Video.

## General Terms

Algorithms, Measurement, Design, Experimentation.

## Keywords

TV Newscasts; Tension Levels; Multimodal Sentiment Analysis; Prosody Features of Speech; Facial Expressions; Closed Caption.


## 1. INTRODUÇÃO

O telejornal é a espinha dorsal de todas as redes de televisão no mundo. Quando se assiste a um telejornal, presencia-se, com certa frequência, a espetacularização da informação por parte dos telejornalistas como uma manobra para preparar o modelo mental dos espectadores, mesmo que isso não seja percebido pelo público, a fim de transmitir credibilidade [1, 2].

O estilo dos telejornalistas é construído, principalmente, pela redundância, ou seja, pela repetição, ainda que existam alguns traços de imposição individual do sujeito. Dessa forma, como a fala é única, o sujeito de informação escolhe os recursos de estilo que irá utilizar em uma situação específica de comunicação, tais como modulações vocais e expressividade corporal, que permite ao programa tomar uma posição sobre aquilo que é noticiado e estabelecer uma identificação com o telespectador [3]. Agregando-se a essa dinâmica, o telejornal exibe a chamada das notícias, depois algumas reportagens com conteúdo emocional alto ou moderado, variando a tensão à medida que mais reportagens são apresentadas durante o programa. Geralmente, exceto em momentos de grande comoção social, as últimas notícias possuem conteúdo emocionalmente distenso [4].



*WebMedia'15*, October 27–230, 2015, Manaus, Brazil.



Nesse contexto, a subjetividade no sequenciamento das notícias em telejornais e na articulação verbal e corporal dos âncoras vem sendo objeto de estudo [5]. Existem diversos esforços em descobrir padrões que revelem a estratégia comunicativa pretendida por meio da organização das notícias, dos diferentes planos fílmicos empregados e da influência dos jornalistas na construção do *ethos* discursivo do telejornal [6, 7]. Ademais, os jornalistas devem causar impacto e despertar sentimentos diversos no telespectador por meio da expressividade corporal, de acordo com os objetivos pautados pelas emissoras de televisão [3].

### 1.1 Definição do Problema

No estudo sobre os telejornais, percebe-se que o problema da tensão está na percepção semântica do telespectador sobre o conteúdo da notícia, tratando-se de um problema característico de análise de subjetividade. Dessa forma, os esforços concentram-se sobre as instâncias de produção, visto que a Análise do Discurso procura encontrar, nos padrões do telejornal, as estratégias para convencer o telespectador.

Nesse cenário, a proposta deste artigo é identificar esses padrões do sequenciamento das notícias na composição de um telejornal qualquer a fim de cruzar as respectivas informações para os analistas do discurso. Acredita-se que esse sequenciamento é determinado pelo nível de tensão emocional na enunciação das notícias durante a exibição do programa [8, 21]. Por se tratar de uma percepção subjetiva, pode-se modelar diferentes níveis de tensão para agrupar as notícias. Fundamentando-se na proposta de [8] em categorizar as notícias em *Distensão*, *Tensão Moderada* e *Alta Tensão*, este trabalho classifica as notícias considerando as polaridades de *Baixa Tensão* e *Alta Tensão* emocional por meio da análise de conteúdo dos respectivos vídeos.

Nas reportagens sob a influência da *Alta Tensão* (AT), tem-se a percepção de que a condução do discurso em relação à temática nos remete à sensação de conflito, de violência, de tragédia e de morte (homicídios), revelando problemas do mundo que podem produzir uma patemização no telespectador. As reportagens sob *Tensão Moderada* (TM) também promovem uma implicação patêmica no telespectador, mas o local do fato possui certa distância do cotidiano do público ao qual se endereçam por conta de um nível de empatia médio em relação aos envolvidos e à temática abordada na notícia. Em contrapartida a essas duas categorias, tem-se as notícias imergidas ao nível da *Distensão* (DT) em que a percepção sobre a condução do discurso em relação à temática remete-nos à semântica de conteúdo leve ou meramente informativo, com baixa espetacularização, tais como eventos esportivos, comemorações, dicas de culinária, avanços tecnológicos, dentre outros [8].

A partir dos recursos audiovisuais e textuais disponíveis nos vídeos de telejornais, este artigo propõe automatizar e combinar técnicas computacionais para o reconhecimento de emoções em expressões faciais, determinar os planos fílmicos, extrair modulações sonoras nas falas dos indivíduos participantes nos

vídeos e identificar a polaridade do texto obtido do *closed caption* desses vídeos. Os modelos devem ser robustos tanto para os ambientes internos e controlados dos estúdios quanto para os ambientes externos, onde ocorrem as reportagens.

## 1.2 Motivação

A análise de telejornais é essencial para os analistas de mídia em diversos domínios, especialmente no campo do jornalismo [22, 23]. Como um telejornal constitui um tipo específico de discurso e de prática sociocultural, técnicas de análise do discurso [24] têm sido aplicadas para analisar a estrutura do telejornal em vários níveis de descrição, considerando algumas propriedades, tais como os temas gerais abordados, as formas esquemáticas usadas e suas dimensões estilístico-retóricas [26].

Normalmente, os discursos são analisados sem o apoio de ferramentas computacionais, tais como softwares de anotação automatizados e programas de análise de vídeo. No entanto, com o rápido desenvolvimento de áreas como a análise de sentimentos, a linguística computacional, os sistemas multimídia e a visão computacional, foram propostos novos métodos para apoiar a análise do discurso, especialmente de conteúdos multimídia como, por exemplo, telejornais [25].

Como um passo em direção a esse objetivo, apresentamos uma abordagem computacional inovadora para apoiar o estudo dos padrões de sequenciamento de notícias em telejornais, a partir da estimativa e avaliação das polaridades de tensão das notícias durante a narrativa dos acontecimentos. A abordagem proposta contribui para a identificação das estratégias comunicativas desses programas, permitindo aos pesquisadores realizar a análise semiodiscursiva sobre a linguagem verbal e a linguagem não-verbal utilizadas nas respectivas instâncias de produção.

A linguagem verbal coloca-se como a principal forma de comunicação, mas a linguagem não-verbal também é responsável por exercer um papel relevante no entendimento das mensagens e contribuir para o processo comunicacional. No telejornalismo, a comunicação não-verbal pode manifestar-se pela postura, pelas expressões faciais e pelos movimentos gestuais do jornalista diante da câmera, tendo-se uma significância maior quando esse profissional enuncia um fato no programa [27]. Assim, tem-se a premissa de que os gestos coverbais fornecem um canal para a expressão visual de ideias [8, 28, 29]. Além disso, diante das diversas abordagens apresentadas, mostram-se promissores os estudos do conteúdo emocional nos telejornais, muitas vezes submergidos na hiperemoção [27].

## 1.3 Contribuições

Do ponto de vista prático, este artigo apresenta importantes contribuições para a área de análise de conteúdo em vídeos, em especial vídeos de telejornais, e para a Análise do Discurso sobre o nível de tensão contido no sequenciamento de suas notícias:

- Novas técnicas de análise multimodal de conteúdo objetivo e subjetivo em vídeos de telejornais;
- Nova metodologia para complementar a análise do nível de tensão em notícias orientadas a componentes visuais e emocionais extraídos de forma automática por meio da combinação de técnicas de análise de sentimentos e de visão computacional.

Do ponto de vista teórico, o trabalho é importante, pois fornece evidências empíricas que, por meio da identificação de padrões no sequenciamento de notícias em telejornais, pode-se desenvolver estudos interdisciplinares inéditos, com grandes impactos no campo da Análise do Discurso midiático.

# 2. TRABALHOS RELACIONADOS

Nesta seção são apresentados alguns dos principais trabalhos relacionados, os quais contribuíram para alavancar estudos sobre: (i) níveis de tensão e padrões de sequenciamento de notícias nos telejornais, (ii) análise de sentimentos em notícias, (iii) modelos para análise multimodal de sentimentos em vídeos e (iv) técnicas para o reconhecimento audiovisual de emoções em expressões faciais e modulações sonoras em sinais de áudio.

## 2.1 Níveis de Tensão em Telejornais

A pesquisa [1] analisa a evolução da linguagem utilizada nos telejornais espanhóis desde o final das décadas de 80 e de 90 até as exibições dos últimos anos. Essas exibições foram analisadas exaustivamente por meio de questionários aplicados à população que contemplavam a análise de conteúdo, tanto na descrição dos programas, especificando-se a data de transmissão e o número total de notícias de cada telejornal, quanto na construção dos telejornais, analisando-se a serialização e composição de notícias, a personalização de informações e a importância dada ao impacto visual. Os resultados mostraram que os noticiários foram alterando o estilo de enunciação ao longo do tempo, passando de uma narrativa objetiva para uma narrativa dramatizada, que usa conteúdo visual com níveis de tensão altos a fim de alcançar sensibilização e identificação do público com os protagonistas das notícias. A espetacularização da informação tornou-se o recurso diário dos telejornais atuais como uma resposta à necessidade de alcançar maiores índices de audiência.

No estudo sobre o sequenciamento das notícias em telejornais quanto às temáticas, o trabalho [8] identifica certo padrão em relação ao nível de tensão do assunto tratado, independente da temática em que esse foi classificado. Esse trabalho analisou quatro telejornais, dois brasileiros e dois franceses, encontrando semelhanças entre eles quanto às temáticas abordadas e ao sequenciamento das notícias, com a tendência de ir de um ponto máximo de tensão, normalmente sobre notícias que demonstram a desordem do mundo, para certa sensação de leveza ao abordar notícias de conteúdo esportivo, de lazer, dentre outros assuntos. Com vasto levantamento das notícias, em várias datas e extensos tempos de exibição, foram modelados três níveis de tensão com base no conteúdo temático e na repercussão do sentimento envolvido: *Distensão*, *Tensão Moderada* e *Alta Tensão*.

## 2.2 Análise de Sentimentos em Notícias

Um grande número de pessoas acessa notícias virtuais online no site das grandes portais de comunicação. A interatividade e o imediatismo presente em notícias online estão mudando a forma como as notícias estão sendo produzidas e expostas pelas corporações de mídia. Sites de notícias têm que criar estratégias eficazes para chamar a atenção das pessoas para os respectivos conteúdos. Esforços recentes têm explorado técnicas de análise de sentimentos para examinar artigos de notícias ou para criar novas aplicações [9]. Sobre esse contexto, os autores de [10] investigam possíveis estratégias utilizadas por empresas de notícias online na concepção de suas manchetes. Foi analisado o conteúdo de 69.907 manchetes produzidas por quatro grandes empresas de mídia durante um mínimo de oito meses consecutivos em 2014. Foram extraídas características do texto das notícias relacionadas com a polaridade do sentimento da respectiva manchete, descobrindo-se que o sentimento da manchete está fortemente relacionado com a popularidade das notícias e com o sentimento nos comentários postados em cada notícia. Notícias com conteúdo de sentimento

negativo tendem a gerar muitos acessos e comentários também negativos. Em uma análise dos resultados, os autores apontaram que quanto maior a tensão negativa das notícias, maior a necessidade que o usuário sente em emitir sua opinião que, em assuntos polêmicos, possui grande possibilidade de ser um comentário que contradiz a opinião de outro usuário, promovendo discussões nas postagens.

No trabalho [11] foi apresentada uma abordagem na construção de um corpus de notícias em alemão sobre política para a mineração de opinião. O corpus foi treinado utilizando-se técnicas do estado-da-arte no aprendizado de regras para relacionar a extração dos títulos das notícias e do conteúdo da opinião, bem como a classificação de polaridades. O aprendizado de regras foi realizado com o suporte de um framework de aprendizagem de máquina minimamente supervisionado, obtendo-se sentimento negativo para 86,2% das notícias recuperadas numa estratégia de tensão para acesso ao respectivo conteúdo.

## 2.3 Análise Multimodal de Sentimentos

No processamento de arquivos multimídia, apenas se debruçar em uma modalidade de informação não é suficiente para realizar, de forma assertiva, a análise de sentimentos desse tipo de conteúdo. A multimodalidade implica no uso de várias mídias além de texto, tais como áudio e vídeo, para aumentar a precisão da classificação do sentimento pelos analisadores de conteúdo emocional. A integração desses recursos permite combinar os resultados obtidos da análise de sentimentos em metatados textuais, geralmente determinada pela polaridade e intensidade de dicionários léxicos, com a classificação emocional do sinal de áudio por meio de características prosódicas e a análise de conteúdo emotivo de vídeos com base nas posturas, gestos e expressões faciais. Nesse contexto, o trabalho [30] aborda o problema da análise multimodal de sentimentos em vídeos coletados na Web, extraindo as informações de cada modalidade de recurso para, em seguida, combinar essas saídas por meio de métodos de fusão de classificadores. Em experiências preliminares usando o conjunto de dados do YouTube, foi obtida uma precisão de 78,20% na extração de polaridade dos vídeos, superando muitos sistemas relatados no respectivo estado da arte.

Os autores de [31] descrevem uma abordagem para a análise baseada em conteúdo de mídias sociais, combinando mineração de opinião em texto e em recursos multimídia (imagens, vídeos, dentre outros), centrando-se no reconhecimento de entidade e evento para auxiliar arquivistas na seleção de material para inclusão em mídias sociais a fim de preservar a memórias de comunidades em categorias semânticas, resolver a ambiguidade e fornecer mais informações contextuais. A abordagem textual é baseada em regras, considerando questões inerentes às mídias sociais, tais como texto incorreto gramaticalmente, uso de palavrões e sarcasmo. Além da nova combinação de ferramentas para mineração de opinião em texto e em recursos multimídia, as ferramentas de Processamento de Linguagem Natural (PLN) foram adaptadas para a mineração de opinião no domínio específico de mídias sociais.

## 2.4 Reconhecimento de Emoções em Vídeos

Em [12], os autores demonstraram evidências de que as expressões faciais de emoções podem ser inferidas por sinais rápidos da face. Estes sinais são caracterizados por mudanças na aparência da face que duram segundos ou frações de segundo, uns mais visíveis que outros. Com isso, os autores formularam o modelo de emoções básicas, chamado Facial Action Coding System (FACS), fundamentado sobre seis expressões faciais (Alegria, Surpresa, Aversão, Raiva, Medo e Tristeza) que são encontradas em diversas culturas e são exibidas da mesma forma, desde crianças até idosos. O FACS é um método objetivo de quantificação da expressão facial por meio das ações faciais que a compõem. Foram mapeados os pontos de singularidade de cada tipo de expressão facial com base em testes realizados sobre um vasto banco de imagens.

No trabalho [13], o modelo facial Candide, usado para rastreamento e extração de características, foi adotado. Na etapa de classificação a primeira pose de cada face foi determinada e, em seguida, as expressões foram reconhecidas por meio de uma abordagem estocástica sobre um banco de dados próprio com vários vídeos. Os autores formularam experimentos com vídeos contendo 1.600 frames em que os voluntários foram autorizados, na gravação dos vídeos, a exibir qualquer expressão facial em ordem e duração aleatórias. Os resultados apresentaram melhor desempenho no seguimento da boca nas expressões faciais.

No âmbito do reconhecimento de características prosódicas nas modulações sonoras em sinais de áudio, os autores de [14] apresentaram o desenvolvimento do openSMILE, um arcabouço para extração de características emocionais em discurso, música e sons em geral, existentes em vídeos e em sinais de áudio. A detecção de atividade e acompanhamento de voz e a detecção de face também são recursos oferecidos pelo framework.

Observa-se, então, que existe uma demanda em analisar o conteúdo emocional de notícias telejornalísticas nos diversos tipos de mídias, cujos recursos informativos são obtidos, atualmente, de forma manual. Neste contexto, este artigo aplica técnicas robustas que permitem, de forma automática, extrair recursos multimodais para determinar a polaridade de tensão emocional do conteúdo telejornalístico e identificar o padrão no sequenciamento de notícias para a análise semiodiscursiva desses programas.

# 3. ABORDAGEM PROPOSTA

Esta seção apresenta a abordagem proposta neste trabalho para análise multimodal de sentimentos no padrão de sequenciamento de notícias em telejornais a partir da determinação automática e avaliação das polaridades tensivas no conteúdo da narrativa dos acontecimentos exibidos nos vídeos. Fundamentalmente, este artigo trabalha com os dados extraídos da modulação sonora do áudio, das emoções reconhecidas nas expressões faciais dos indivíduos nos vídeos e da análise de sentimentos sobre o *Closed Caption*, determinando-se a partir de tais dados os níveis de tensão correspondentes.

Para a etapa de reconhecimento de emoções, este artigo se baseia na análise de expressões faciais, conforme as abordagens propostas por [15] e [16], bem como na extração de características prosódicas referentes às modulações sonoras detectadas em sinais de áudio, conforme proposto por [14]. Para a etapa de análise de sentimentos sobre o texto obtido do *Closed Caption*, foi utilizado o software iFeel [33] que retorna a polaridade das sentenças com base em diversas técnicas do estado da arte. Adicionalmente, os parâmetros semiodiscursivos propostos no trabalho [8] permitiram modelar a polaridade tensiva das notícias em *Baixa Tensão* e *Alta Tensão*. Estas polaridades de tensão são ponderadas conforme a intensidade da emoção inferida por meio das expressões faciais dos indivíduos, seus respectivos planos fílmicos, os valores de modulações do áudio correspondentes às falas de tais indivíduos e, finalmente, do sentimento extraído de cada trecho das falas registradas no *Closed Caption* desses vídeos.

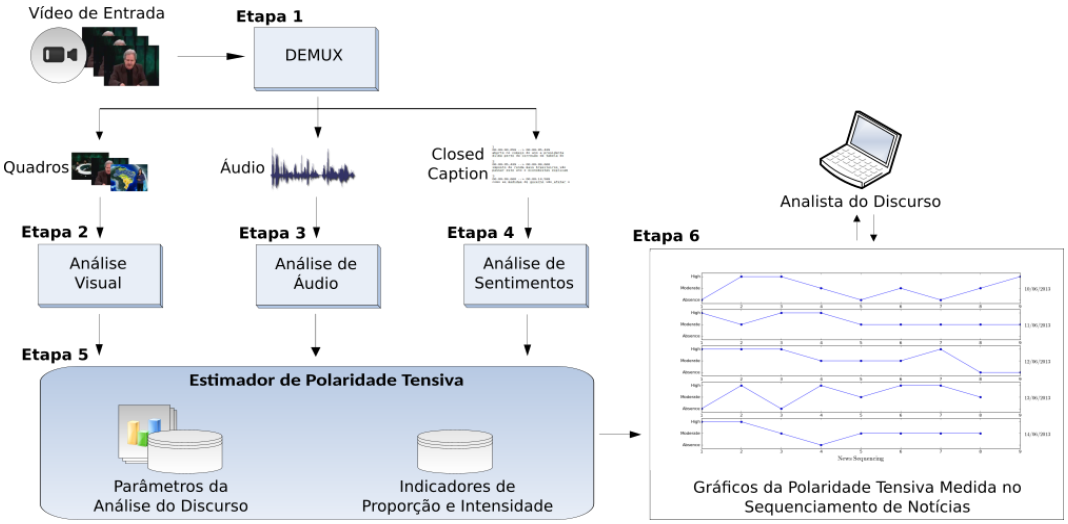


**Figura 1. Visão geral da abordagem proposta para a estimação automática da polaridade tensiva de notícias em vídeos de telejornais.**

A Figura 1 apresenta uma visão geral da abordagem proposta para a determinação automática dos níveis de tensão nas sequências de notícias exibidas em vídeos de telejornais por meio da análise multimodal de sentimentos. Na etapa 1, realiza-se a extração dos recursos multimodais do vídeo submetido ao sistema. Na etapa 2 de Análise Visual, aplicam-se os métodos para o reconhecimento de emoções em cada frame. Paralelamente, na Análise de Áudio, o sinal de áudio é processado, extraindo-se a modulação sonora nos instantes de fala dos indivíduos (etapa 3) e, na Análise de Sentimento, o texto obtido do *Closed Caption* é tratado e submetido ao iFeel (etapa4). Em seguida, a partir das ocorrências contabilizadas para cada emoção, realiza-se uma soma ponderada com os indicadores obtidos nas três etapas de análise multimodal (etapa 5) e agrupam-se as emoções reconhecidas nos níveis de tensão considerados. A etapa 6 consiste em apresentar essas informações em gráficos de acordo com a ordem em que as notícias foram exibidas no programa.

## 3.1 DEMUX

O sistema desenvolvido realiza a extração dos recursos audiovisuais e textuais dos vídeos de telejornais, em que cada vídeo é uma notícia, obtendo-se uma lista das imagens que compõem o vídeo (frames), o sinal de áudio em formato WAV e o texto obtido do *Closed Caption* utilizando-se das ferramentas OpenCV [17], FFmpeg [18] e Google2SRT [32].

Os frames extraídos são transformados do espaço de cores RGB para escala de cinza com o intuito de facilitar o processo de reconhecimento de faces na etapa de Análise Visual [15, 16]. Já o texto obtido do *Closed Caption* é tratado para gerar sentenças somente com o conteúdo falado, excluindo-se os metatados de marcação de tempo da aparição de cada trecho do texto no vídeo. Dessa forma, espera-se a ocorrência de sentenças neutras, considerando-se, em teoria, o formalismo do ambiente jornalístico e que alguns dos trechos que aparecem são muito curtos.

Esses recursos multimodais são organizados em três filas, uma para os frames de cada vídeo, outra para os sinais de áudio e a terceira para o conteúdo textual obtido do *Closed Caption*, que disparam os processos paralelos de Análise Visual, Análise de Áudio e Análise de Sentimentos, respectivamente.

## 3.2 Análise Visual

Sobre os frames extraídos e tratados na etapa anterior, aplicam-se os métodos para o reconhecimento de expressões faciais e o reconhecimento de emoções propostos em [15] e [16], obtendo-se os valores dos indicadores de intensidade sobre a expressividade emocional da face.

Como o modelo determina que as ações faciais de cada parte da face sejam detectadas a partir de uma face neutra, definiu-se neste trabalho marcar como neutra a primeira face detectada no vídeo, partindo-se da premissa de imparcialidade emotiva em que os apresentadores devem iniciar e se manter. Conforme a taxa de frames por segundo (fps) de cada vídeo, determina-se qual a emoção predominante naquele conjunto de frames para cada segundo de vídeo.

Além dos indicadores de intensidade sobre as expressões faciais, são extraídos os valores de proporção das faces no enquadramento da câmera (plano fílmico) em cada frame. Para isso, calcula-se a relação entre a área de reconhecimento da maior face detectada no frame e a resolução desse frame a qual ela está submetida. Acredita-se que no universo telejornalístico, em uma situação com vários indivíduos sob diferentes planos de enquadramento de câmera, estará em foco o indivíduo sob o plano fílmico mais fechado, ou seja, cuja face ocupa uma área maior do frame. Além disso, quanto maior a área que a face ocupa no frame, maior é a intensidade emocional que se pretende aplicar naquela instância de produção como estratégia comunicativa do programa no uso variado de planos fílmicos durante a exibição [7, 34].

A emoção reconhecida em cada frame, bem como os respectivos valores de intensidade e de proporção das faces detectadas, são gravados em um arquivo de extensão XML utilizando-se o padrão da linguagem de marcações sobre emoção EmotionML [19] a fim de permitir a interoperabilidade de futuras aplicações que tenham esse enfoque de pesquisa. Esses dados são submetidos para o Estimador de Polaridade Tensiva que, ao combinar os indicadores sobre a quantidade de emoções ocorridas e os indicadores sobre a polaridade do texto obtido do *Closed Caption*, computará a polaridade de tensão para aquele vídeo de notícia.

## 3.3 Análise de Áudio

Na etapa de Análise de Áudio, que ocorre em paralelo às etapas de Análise Visual e Análise de Sentimentos, o sinal de áudio é processado, extraindo-se as modulações sonoras nos instantes de fala dos indivíduos, ou seja, quando ocorre o discurso. Para isso, utiliza-se o framework openSMILE [14] para extrair o espectro do componente de áudio e obter as características prosódicas de intensidade sonora, probabilidade de vocalização e a frequência fundamental das modulações desses sinais.

A intensidade sonora reflete o percentual de percepção da amplitude da onda sonora pelo ouvido humano medida em decibéis (dB). A probabilidade de vocalização evidencia a chance de haver uma diferenciação na entonação durante o discurso para o próximo instante. Já a frequência fundamental corresponde ao primeiro harmónico de uma onda sonora, sendo a frequência mais influente na percepção de um som específico e um dos principais elementos que caracterizam a voz [14].

Essa etapa é muito importante para a análise dos dados referentes às modulações no discurso do telejornal que podem influenciar o ritmo na locução de notícias, incluindo os aspectos semânticos das estruturas emotivo-verbais que devem ser testadas quanto à eficácia da transmissão de informação [20].

## 3.4 Análise de Sentimentos do Closed Caption

Na etapa de Análise de Sentimentos, o arquivo de *Closed Caption* do vídeo é tratado a fim de se extrair apenas o conteúdo textual dependente do conteúdo referente aos trechos falados e transcritos durante a enunciação da notícia, eliminando os metadados independentes de conteúdo, tais como as enumerações de cenas e frames, bem como as marcações de tempo, em que os trechos são exibidos no vídeo. Em seguida, cada trecho, comumente exibido em duas linhas de texto ao longo de alguns segundos na tela, geralmente em sincronismo com o áudio [35], é transformado em uma única sentença. Cada sentença é gravada, na ordem em que é exibida no vídeo, em um arquivo de texto que será submetido ao processo de análise de sentimento. Para alcançar esse objetivo, foi utilizado o iFeel para extrair as polaridades de sentimento de cada sentença, dentre positivo, neutro e negativo, para 20 métodos do estado da arte e o método combinado implementado em [33].

Sob a abordagem de Votos por Maioria, em que a resposta final de polaridade é aquela que recebe o maior número de votos dentre todos os métodos executados, as ocorrências de sentenças classificadas como positivas, neutras e negativas são contabilizadas, nessa ordem, em um vetor tridimensional para representar cada dimensão de polaridade.

## 3.5 Estimador de Polaridade Tensiva

Depois que os indicadores de intensidade e de proporção são obtidos, a próxima etapa é a estimação dos níveis de tensão e a análise estatística dos dados obtidos. Inicialmente, calcula-se a quantidade de ocorrências de cada emoção no vídeo de cada notícia e as polaridades do texto do *Closed Caption*. Em seguida, a partir das ocorrências contabilizadas, realiza-se uma soma ponderada com os valores dos indicadores obtidos nas etapas anteriores, conforme apresentado na Equação 1, e, finalmente, agrupa-se as emoções reconhecidas nas polaridades de tensão consideradas, ou seja, para cada polaridade de tensão, soma-se a quantidade das ocorrências de cada emoção ponderadas pelos valores dos indicadores multimodais. As polaridades de tensão estimadas são apresentadas em gráficos de acordo com a ordem em que as notícias apareceram, contribuindo para a análise semiodiscursiva do padrão de sequenciamento de notícias em telejornais.

$$p = max\left(\sum_{j}\sum_{k} b_j * v_k \quad , \quad \sum_{m}\sum_{k} a_m * v_k\right) \quad (1)$$

em que:

- $p$ – representa a polaridade de tensão com a soma ponderada de maior valor no vídeo;
- $b_j$ – representa a frequência de cada uma das $j$ emoções (Alegria, Neutro, Surpresa e Aversão) que caracterizam a polaridade de *Baixa Tensão*;
- $a_m$ – representa a frequência de cada uma das $m$ emoções (Desprezo, Raiva, Medo e Tristeza) que caracterizam a polaridade de *Alta Tensão*;
- $v_k$ – corresponde a cada um dos $k$ valores dos indicadores obtidos nas etapas de análise multimodal.

A abordagem proposta realiza o reconhecimento das expressões faciais associadas às emoções de Alegria, Surpresa, Aversão, Raiva, Medo, Tristeza e Neutro. Conforme a Etapa 5, essas emoções são contabilizadas e agrupadas nas polaridades tensivas modeladas, em que as ocorrências de Alegria, Surpresa, Aversão e de Neutro referem-se às ocorrências de *Baixa Tensão* e as ocorrências de Raiva, Medo e Tristeza referem-se à *Alta Tensão*. A quantidade de faces neutras e de trechos de texto neutros extraídos do *Closed Caption* foram contabilizadas e agrupadas com a polaridade de *Baixa Tensão*, conforme experimentos estatísticos descritos na Seção 4.1. A própria neutralidade de carga emocional condiciona uma sensação de alívio por parte do telespectador, categorizando as notícias com baixa ou nenhuma tensão, mas vale ressaltar que o apelo pela espetacularização da informação pode levar a uma estratégia comunicativa em agregar certa patemização a uma notícia de temática leve e o modelo proposto pode recuperar reportagens que seriam distensas como conteúdo de tensão emocional moderada, justificando o agrupamento proposto, conforme ilustrado na Figura 2.

Os algoritmos para extração de características audiovisuais foram aplicados a fim de fornecer os recursos necessários para a detecção das faces dos indivíduos atuantes nos vídeos. Sobre as faces detectadas, atuam os módulos para o reconhecimento de expressões faciais e a identificação dos planos fílmicos de enquadramento da câmera correspondentes.

# 4. RESULTADOS EXPERIMENTAIS

Esta seção apresenta o corpus utilizado neste trabalho, descreve os experimentos realizados sobre o sistema implementado e discute os principais resultados obtidos, visando-se avaliar a aplicabilidade e a eficácia do modelo proposto.

## 4.1 Corpus

O corpus utilizado neste trabalho para avaliar a abordagem proposta é constituído por amostras de vídeos de quatro telejornais, sendo três telejornais brasileiros (Jornal da Record, Jornal da Band e Jornal Nacional) e um telejornal norte-americano (CNN). Mais especificamente, utilizou-se nos experimentos 94 vídeos de notícias do Jornal da Band (Band News) exibidas de 10 a 14 de junho de 2013 e de 23 a 28 de abril de 2015; 237 vídeos do Jornal da Record (JR) exibidas 24 de maio de 2013, 10 de janeiro de 2015, 05 de fevereiro de 2015 e 02, 10, 16 e 18 de

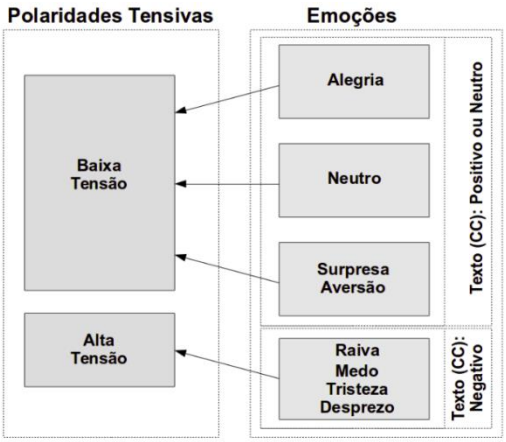


**Figura 2. Agrupamento das emoções nas polaridades tensivas modeladas.**

março de 2015; e 27 vídeos de notícias exibidas pelo Jornal Nacional (JN) em 20 de janeiro de 2015.

Cabe ressaltar que o corpus considerado é composto exclusivamente por vídeos de notícias (cada arquivo de vídeo corresponde a uma notícia), em especial Notas Peladas, comentários dos apresentadores em Notas Pé e reportagens com Notas Cobertas. Vídeos que continham apenas Notas Cobertas não foram utilizados, visto que a detecção de faces na notícia é um recurso crucial para a aplicação do modelo proposto.

## 4.2 Testes do Agrupamento de Neutralidade

No âmbito do telejornalismo, tem-se como recomendação teórica a postura centrada na imparcialidade e neutralidade emocional por parte dos profissionais atuantes nesses programas [3], porém essa tentativa se esvai ao longo da enunciação por conta do fator humano. Dessa forma, optou-se por analisar se os momentos de neutralidade dentro de uma notícia contribuem para evidenciar a ocorrência de alguma carga de tensão emocional que poderá caracterizar, com alguma predominância, aquele conteúdo na medida em que ele é enunciado.

A Figura 3 apresenta os valores de acurácia para 264 vídeos de notícias referentes à 8 exibições de telejornais submetidos a três métodos implementados no sistema: (i) sem neutralidade, ou seja, as ocorrências de faces neutras e de trechos de texto neutros do *Closed Caption* não foram contabilizadas; (ii) o processamento dos recursos neutros nas instâncias de vídeos com baixa tensão; e (iii) a contagem dos recursos de neutralidade nas instâncias ditas com alta tensão. Percebe-se, visualmente, que o método que agrupa os recursos neutros como *Baixa Tensão* possui os maiores valores de acurácia. A fim de comprovar esses resultados, os Testes de Hipótese dois a dois sobre os valores de acurácia medidos resultaram em hipótese alternativa para todos os casos, ou seja, existem evidências estatísticas de que a acurácia média de 0.81 computada para o método a medição de *Neutralidade na Baixa Tensão* é realmente maior que as acurácias médias de 0.32 e de 0.61 para as abordagens sem neutralidade e com neutralidade em instâncias de alta tensão, respectivamente.

## 4.3 Estimação das Polaridades de Tensão

No trabalho [8], os telejornais brasileiros Jornal da Band e Jornal Nacional, bem como os franceses TF1 e France 2, foram analisados sob o olhar da Análise do Discurso quanto à forma em que esses programas são construídos. Ao avançar nesses estudos, por meio da anotação manual dos níveis de tensão percebidos ao assistir cada vídeo, a autora percebeu certo padrão na sequência em que as notícias eram exibidas, independente das temáticas abordadas, ao categorizar os respectivos conteúdos com ausência de tensão, com nível de tensão moderado e com alta tensão.

Norteando-se nessa expectativa, realizou-se esse tipo de anotação manual, assistindo-se cada vídeo, com o intuito de comparar os resultados do processamento automático de níveis de tensão e computar o desempenho do modelo computacional proposto. A Figura 4 apresenta o resultado da estimação automática das polaridades de tensão em relação à anotação manual da polaridade de cada notícia durante o Jornal da Band nas exibições do dia 10 ao dia 14 de junho de 2013, à esquerda, e do dia 23 ao dia 28 de abril de 2015, à direita. O sistema executado sobre a abordagem proposta de análise multimodal de sentimentos obteve acurácia entre 0.62 e 0.89, conforme ilustrado. Pela anotação manual realizada, percebe-se forte padrão das polaridades de tensão na

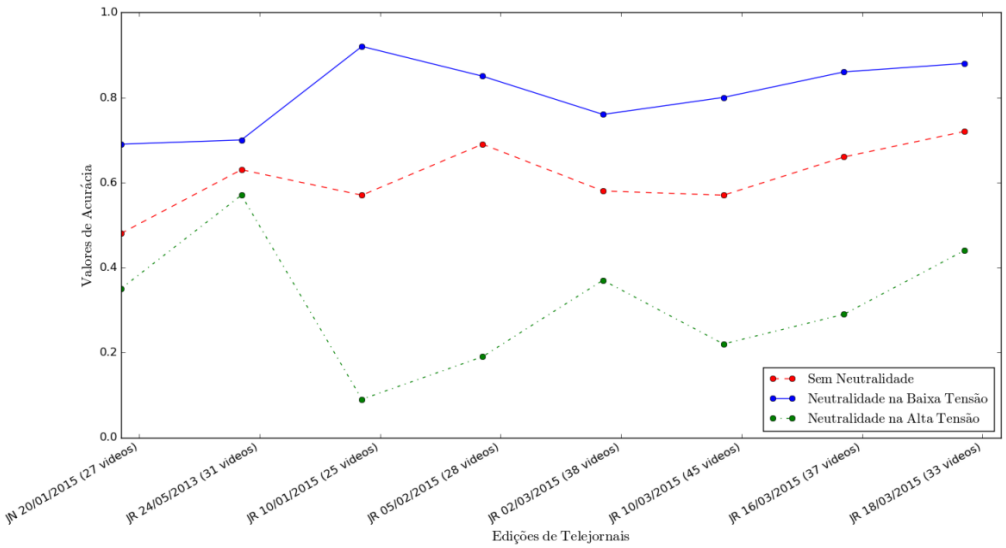


**Figura 3. Análise dos valores da acurácia média entre os métodos de neutralidade facial e de sentimento em texto.**

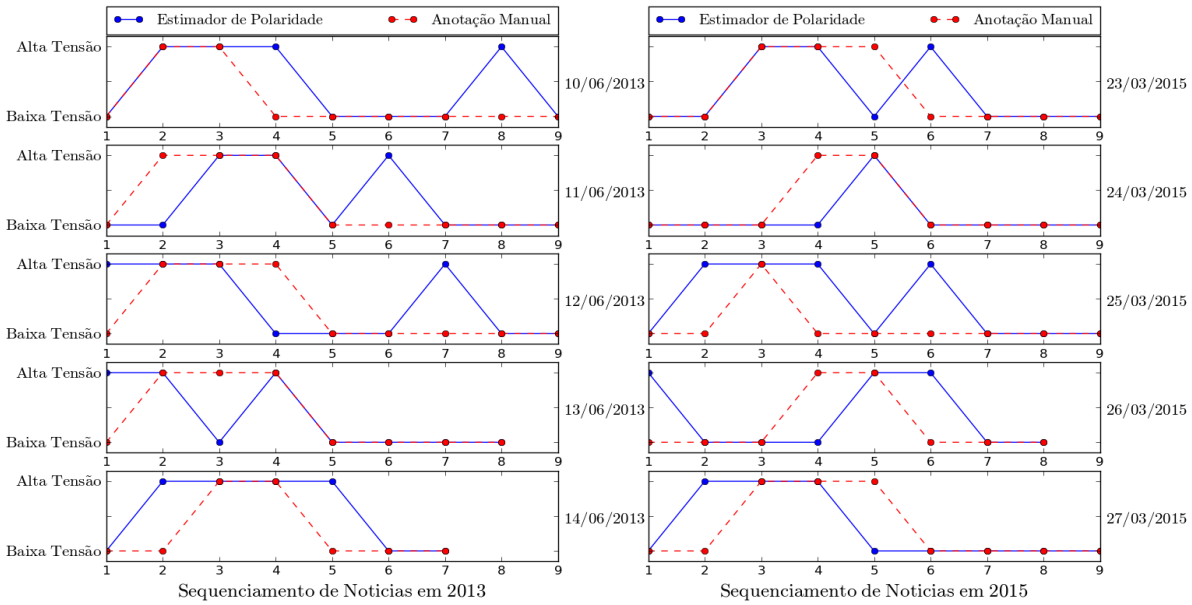


**Figura 4. Distribuição das polaridades tensivas nas notícias do Jornal da Band em 2013 e em 2015.**

organização das notícias desse telejornal.

Para os testes realizados sobre instâncias do Jornal Nacional, foi percebido um padrão diferenciado em relação ao Jornal da Band, mantendo-se o nível de tensão elevado, praticamente, durante toda a escalada, iniciando a exposição de algumas manchetes com conteúdo de alto grau de tensão (crimes ou tragédias). O que chamou atenção nessa análise do Jornal Nacional é o nível de tensão diferenciado dado a uma mesma notícia como, por exemplo, na reportagem sobre as declarações do Papa Francisco incentivando os fiéis da igreja católica a fazerem um planejamento familiar, categorizada como notícia com baixa tensão emocional na análise manual, teve sua respectiva chamada anotada como conteúdo de alta tensão por conta das expressões enérgicas usadas pelos apresentadores, dentre as quais, que o "Papa Francisco diz que os católicos não precisam se reproduzir como coelhos", mostrando-se como uma manobra de produzir efeito de espetáculo sobre uma informação sutil.

## 5. CONCLUSÕES

O estudo interdisciplinar proposto neste trabalho envolve as áreas de Linguística e Ciência da Computação no que se refere, respectivamente, ao embasamento teórico da análise semiodiscursiva de telejornais por meio do sequenciamento das notícias exibidas nos programas em função das polaridades de tensão da enunciação dos fatos. O trabalho proposto mostra-se relevante para enriquecer a área e favorecer o uso de modelos computacionais como ferramenta fundamental para análise de vídeos jornalísticos e afins.

As expressões faciais são formas de comunicação não-verbal amplamente utilizadas em nosso cotidiano. Elas externam nossas manifestações referentes aos estímulos a que somos submetidos e, dessa forma, são elementos que fazem parte da composição midiática dos telejornais, tanto pelas emoções provocadas por eles, quanto pelas emoções que seus apresentadores expressam, inconsciente ou conscientemente, se isso for pautado na estratégia comunicativa da rede de televisão, o que pode fornecer indícios sobre a tensão do discurso gerado pelo notícia e o padrão no sequenciamento dos fatos informados nas instâncias de produção desses objetos informacionais.

Neste contexto, este trabalho propõe um modelo para a obtenção das polaridades de tensão nos vídeos de telejornais e expressar, visualmente, o sequenciamento e o comportamento das notícias que o compõem, permitindo ao analista do discurso perceber e estudar os padrões e a identidade na construção desses programas. Para isso, um sistema computacional foi implementado sobre esse modelo para analisar os recursos audiovisuais desses vídeos automaticamente e identificar a emoção sobre as expressões faciais de um interlocutor no telejornal, bem como as modulações sonoras em seu discurso e o sentimento dos respectivos trechos de fala registrados no *Closed Caption*. Os gráficos obtidos nos experimentos ilustram a validade do modelo proposto, principalmente no que se refere ao levantamento manual realizado que apresenta, realmente, a existência de padrões na construção dos telejornais por meio do sequenciamento de notícias pelos níveis de tensão que elas expressam.

Como trabalhos futuros, pretende-se avaliar o uso de técnicas inovadoras de classificação para cada uma das multimodalidades desses vídeos com o propósito de alcançar maior assertividade na determinação dos níveis de tensão, em termos de maior granularidade, observar as polaridades de tensão emocional ao longo da própria enunciação da notícia e mensurar a contribuição de cada recurso na combinação dos métodos de análise multimodal de sentimentos para subsidiar pesquisas inovadoras na semiótica e análise do discurso de telejornais.

## 6. REFERENCES

[1] Begoña, G. S. M.; Fidalgo, M. R. and Santos, M. C. G.. 2010. Analysing the Development of TV News Programmes: from Information to Dramatization. Revista Latina de Comunicación Social, 65, p. 126-145.

[2] Goffman, E. 1981. The Lecture. Forms of Talk. Pennsylvania: University of Pennsylvania Press, p. 162-195.

[3] Wahl-Jorgensen, K. and Hanitzsch, T.. 2008. The Handbook of Journalism Studies. ICA Handbook Series. Routledge, p. 472.

[4] Pereira, M. H. R.; Pádua, F. L. C.; David-Silva, G. 2015. Multimodal Approach for Automatic Emotion Recognition Applied to the Tension Levels Study in TV Newscasts. Brazilian Journalism Research, v. 11, n. 1. (to appear).

[5] Charaudeau, P. 2006. Discours Journalistique et Positionnements Énonciatifs. Frontières et Dérives. Revue SEMEN, n. 22, Énonciation et responsabilité dans les médias, Presses Universitaires de Franche-Comté, Besançon.

[6] Pereira, M. H. R.; Souza, C. L.; Pádua, F. L. C.; David-Silva, G.; Assis, G. T.; Pereira, A. C. M. 2014. SAPTE: A Multimedia Information System to Support the Discourse Analysis and Information Retrieval of Television Programs. Multimedia Tools and Applications, 74(2): 1-15.

[7] Gutmann, J. F. 2012. What Does Video-Camera Framing Say during the News? A Look at Contemporary Forms of Visual Journalism. Brazilian Journalism Research, v. 8, p. 64-79.

[8] David-Silva, G. 2005. Informação Televisiva: uma Encenação da Realidade (Comparação entre Telejornais Brasileiros e Franceses). Doctoral's Thesis in Linguistic Study - UFMG, Belo Horizonte, Brazil.

[9] Pang, B.; Lee, L. 2008. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval. Hanover, January, v. 2, n. 1-2, p. 1-135.

[10] Reis, J.; Benevenuto, F.; Vaz de Melo, P.; Prates, R.; Kwak, H. and An, J. 2015. Breaking the News: First Impressions Matter on Online News. Proceedings of the 9th International AAAI Conference on Web-Blogs and Social Media. Oxford.

[11] Li, Hong; Cheng, Xiwen; Adson, Kristina; Kirshboim, Tal and Xu, Feiyu. 2012. Annotating Opinions in German Political News. Proceedings of the 8th International Conference on Language Resources and Evaluation. European Language Resources Association, Istanbul.

[12] Ekman, P.; Friesen, W. 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, Calif: Consulting Psychologists Press.

[13] Dornaika, F.; Davoine, F. 2008. Simultaneous Facial Action Tracking and Expression Recognition in the Presence of Head Motion. International Journal of Computer Vision, v. 76, n. 3, p. 257-281.

[14] Florian E.; Weninger, F.; Gross, F. and Schuller, B. 2013. Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. ACM Multimedia (MM), Proceedings of the 21st ACM international conference on Multimedia, Barcelona, p. 835-838.

[15] Littlewort, G.; Bartlett, M. S.; Fasel, Ian; Susskind, J. and Movellan, J. 2004. Dynamics of Facial Expression Extracted Automatically from Video. CVPRW'04. Conference on Computer Vision and Pattern Recognition Workshop, IEEE.

[16] Bartlett, M. S., Littlewort, Gwen; Frank, M.; Lainscsek, C.; Fasel, Ian and Movellan, J. 2006. Fully Automatic Facial Action Recognition in Spontaneous Behavior. 7th International Conference on Automatic Face and Gesture Recognition, IEEE, p. 223-230.

[17] Itseez Corporation. Open Source Computer Vision Library (OpenCV). Reference Manual, Available: http://opencv.org. Access: 2015/04/02.

[18] Bellard, F. and Niedermayer, M. FFmpeg Project. Available: http://www.ffmpeg.org/. Access: 2015/04/04.

[19] Schröeder, M. et al. Emotion Markup Language (EmotionML).1.0. W3C Working Draft, v. 29, p. 3-22, 2010.

[20] Machon, L. M. 2012. Rhythm Structure in News Reading. Brazilian Journalism Research, v. 8, n. 2, p. 8-27.

[21] Pantti, M. 2010. The Value of Emotion: An Examination of Television Journalists' Notions on Emotionality. European Journal of Communication, 25(2): 168-181.

[22] Stegmeier, J. 2012. Toward a computer-aided methodology for discourse analysis. Stellenbosch Papers in Linguistics, v. 41, p. 91-114, 2012.

[23] Van-Dijk, T. A. News analysis. Erlbaum Associates, 1987.

[24] Charaudeau, P. A communicative conception of discourse. Discourse Studies, v. 4, n. 3, p. 301-318, 2002.

[25] Baker, P. Using corpora in discourse analysis. Applied Linguistics, v. 28, n. 2, p. 327–330, 2006.

[26] Cheng, F. Connection between news narrative discourse and ideology based on narrative perspective analysis of news probe. Asian Social Science, v. 8, p. 75-79, 2012.

[27] Lipschultz, J. H. and Hilt, M. L. 2011. Local Television Coverage of a Mall Shooting: Separating Facts From Fiction in Breaking News. Electronic News, v. 5, n. 4, p. 197-214.

[28] Eisenstein, J.; Barzilay, R. and Davis, R. 2008. Discourse Topic and Gestural Form. Proceedings of the 23rd AAAI Conference on Artificial Intelligence. Chicago: ACM DL, v. 2, p. 836-841.

[29] Friedman, H. S.; Dimatteo, M. R. and Mertz, T. I. 1980. Nonverbal Communication on Television News: The Facial Expressions of Broadcasters during Coverage of a Presidential Election Campaign. Personality and Social Psychology Bulletin, v. 6, n. 3, p. 427-435.

[30] Poria, S.; Hussain, A.; Cambria, E. Beyond Text Based Sentiment Analysis: Towards Multi-modal Systems. Springer Cognitive Computation manuscript No. (will be inserted by the editor).

[31] Maynard, D.; Dupplaw, D.; Hare, J. 2013. Multimodal Sentiment Analysis of Social Media. BCS SGAI Workshop on Social Media Analysis. p. 44-55.

[32] Google2SRT. Available: http://google2srt.sourceforge.ne. Access: 2015/05/27.

[33] Araújo, M.; Gonçalves, P.; Cha, M.; Benevenuto, F. iFeel: A System that Compares and Combines Sentiment Analysis Methods. Proceedings of the World Wide Web Conference (WWW'14). Seoul, Korea. April 2014.

[34] Hernandes, N. 2006. A mídia e seus truques: o que jornal, revista, TV, rádio e Internet fazem para captar e manter a atenção do público. 1. ed. São Paulo: Contexto.

[35] Araújo, V. L. S. O processo de legendagem no Brasil. Revista do GELNE, Fortaleza, v. 1/2, n. 1, p. 156-159, 2006.