



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
Pró-Reitoria de Pós-Graduação e Pesquisa
Escola de Ciências Sociais e da Saúde
Programa de Pós-Graduação *Stricto Sensu* em Psicologia

**RELATOS VERBAIS DE CONSUMIDORES EM AVALIAÇÕES *ON-LINE*:
PROSPECÇÃO COMPUTACIONAL E INTERPRETAÇÕES COM BASE NO
*BEHAVIORAL PERSPECTIVE MODEL (BPM)***

Parcilene Fernandes de Brito

Orientador: Prof. Dr. Lauro Eugênio Guimarães Nalini

Goiânia, Goiás, BRASIL
Junho, 2018



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
Pró-Reitoria de Pós-Graduação e Pesquisa
Escola de Ciências Sociais e da Saúde
Programa de Pós-Graduação *Stricto Sensu* em Psicologia

**RELATOS VERBAIS DE CONSUMIDORES EM AVALIAÇÕES *ON-LINE*:
PROSPECÇÃO COMPUTACIONAL E INTERPRETAÇÕES COM BASE NO
*BEHAVIORAL PERSPECTIVE MODEL (BPM)***

Parcilene Fernandes de Brito

Orientador: Prof. Dr. Lauro Eugênio Guimarães Nalini

Tese de Doutorado apresentada ao Programa de Pós-Graduação *Stricto-Sensu* em Psicologia da Pontifícia Universidade Católica de Goiás como requisito parcial para obtenção do título de doutor em Psicologia. Linha de pesquisa: Análise do Comportamento.

Goiânia, Goiás, BRASIL
Junho, 2018

B862r	<p>Brito, Parcilene Fernandes de Relatos verbais de consumidores em avaliação on-line[recurso eletrônico]: prospecção computacional e interpretações com base no Behavioral Perspective Model (BPM)/ Parcilene Fernandes de Brito.-- 2018. 181 f.; il.</p> <p>Texto em português com resumo em inglês Tese(doutorado) - Pontifícia Universidade Católica de Goiás, Programa de Pós-Graduação Stricto Sensu em Psicologia, Goiânia, 2018 Inclui referências f.172-181</p> <p>1. Comportamento do consumidor - Análise. 2. Emoções - Análise. 3. Turismo. 4. Aprendizagem. I.Nalini, Lauro Eugênio Guimarães. II.Pontifícia Universidade Católica de Goiás. III. Título.</p> <p style="text-align: right;">CDU: 159.9:366.1(043)</p>
-------	---

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
Pró-Reitoria de Pós-Graduação e Pesquisa
Escola de Ciências Sociais e da Saúde
Programa de Pós-Graduação *Stricto Sensu* em Psicologia

Brito, P. F. (2018) *Relatos verbais de consumidores em avaliações on-line: prospecção computacional e interpretações com base no Behavioral Perspective Model (BPM)*. **Orientador: Prof. Dr. Lauro Eugênio Guimarães Nalini.**

Esta Tese foi submetida à banca examinadora:

Prof. Dr. Lauro Eugênio Guimarães Nalini

Pontifícia Universidade Católica de Goiás – PUC Goiás
Presidente

Prof. Dr. Jorge Mendes de Oliveira Castro Neto

Universidade de Brasília – UnB
Membro convidado externo

Prof. Dr. André Luiz Barreto Esperidião

Escola Superior Associada de Goiânia – ESUP
Sociedade Brasileira de Cultura e Ensino Superior – SBCES
Membro convidado externo

Prof. Dr. Cristiano Coelho

Pontifícia Universidade Católica de Goiás – PUC Goiás
Membro convidado interno

Prof. Dr. Weber Martins

Pontifícia Universidade Católica de Goiás – PUC Goiás
Membro convidado interno

Prof.^a Dr.^a Helenides Mendonça

Pontifícia Universidade Católica de Goiás – PUC Goiás
Membro suplente

Prof.^a Dr.^a Ana Cristina Resende

Pontifícia Universidade Católica de Goiás – PUC Goiás
Membro suplente

À minha mãe.

Agradecimentos

À minha mãe, meu lar, onde encontro a força e o sentido para levantar todos os dias, mesmo sabendo que será mais um dia sem ela. Pelo esforço da minha mãe (D. Albertina), consegui estudar em um bom colégio e fazer um curso superior, mas, especialmente, tive o ambiente ideal para que minha infância fosse a melhor parte da minha vida. Minha mãezinha ensinou-me que a vida é breve, que é sempre necessário enxergar as pessoas e o mundo muito além dos limites que aparentemente (ou realmente) nos cerca, que é preciso ter tempo para brincar (em qualquer idade), ler, estudar e rir (especialmente de nós mesmos).

Ao meu pai, com quem compartilhei silêncios infinitos e uma estranha necessidade de contar histórias (às vezes, para ninguém).

Aos meus irmãos Edson, Edilson e Pascual, que quando criança pensei que fossem personagens de uma história sobre sacrifícios e perdão. Na história, uma menina tecia casacos de urtiga para salvar a vida dos seus irmãos, então acreditei que fosse essa a minha sina. Só depois descobri que, na verdade, os casacos eram tecidos por eles, e quem precisava se cobrir com as urtigas para sobreviver, era eu.

À minha prima Rosa, que é um esteio em nossa família, sempre presente, especialmente nos momentos mais difíceis. Sem ela, provavelmente, não teria tido forças para finalizar essa tese.

Aos meus sobrinhos e demais familiares, pelo apoio em todos os momentos. Obrigada Douglas por gerenciar as enfadonhas situações do meu mundo real.

Às minhas amigas de infância, Kênia e Irenides, que estiveram comigo no pior momento da minha vida e souberam me dar o espaço e a atenção necessários para que eu suportasse o enorme peso que era permanecer acordada.

Aos meus colegas de trabalho e grandes amigos, Fabiano, Jackson, Cristina, Madia e Fernando, que contribuíram de várias formas no desenvolvimento desta tese, desde o apoio em situações corriqueiras do trabalho e nos estudos conceituais e técnicos da área computação, até o entendimento do quanto determinadas rotinas são importantes para minha sanidade, e por fazerem parte dessa rotina há tanto tempo.

Aos meus orientandos de TCC e de projetos de iniciação científica, que me ajudaram nessa jornada e tornaram meu estudo menos solitário e mais eficaz. Muitos deles tiveram trabalhos relacionados ao projeto da *SentimentALL*, como o Matheus, que contribuiu no modelo de testes da ferramenta, o William, que implementou a primeira versão da *SentimentALL* em seu TCC e, especialmente, o Luan Gomes, que com sua inteligência rara, sua incansável busca pelas melhores saídas computacionais, seu compromisso com a pesquisa científica, buscou em seu TCC e no projeto de iniciação científica caminhos mais elegantes para aprimorar o processo de extração dos dados e alguns dos algoritmos da ferramenta.

Às minhas amigas *excers*, Josi e Ariana, que estiveram presentes virtualmente e, sobretudo, emocionalmente, personalizando a voz amiga que ecoava nos tantos momentos solitários deste estudo, proporcionando pausas regadas às divagações tão necessárias sobre Arquivo X, existência de vida nesse planeta, TrustN01, I-Want-to-Believe, Terma etc.

Ao prof. Jorge Oliveira-Castro, meu professor na disciplina “Questões Conceituais no Behaviorismo”, que me mostrou novos caminhos para a pesquisa multidisciplinar baseada nas áreas de computação e psicologia; sua contribuição na banca de qualificação foi decisiva para a finalização desta tese.

Ao meu orientador, prof. Lauro Nalini, que aceitou o desafio de orientar um trabalho multidisciplinar e, com seu profundo conhecimento na área de Análise Comportamental, foi fundamental para o estabelecimento das variáveis desta tese e para a estruturação dos resultados e discussões. Como não temos controle das inúmeras variáveis da vida, às vezes, o processo de orientação é amplo e complexo, pois vai além de indicar referências, aprimorar questões metodológicas e discussões científicas, assim, as palavras compartilhadas foram também palavras de conforto e empatia. Obrigada prof. Lauro por ser sensível a esses momentos.

“Faço o que posso, percebo que há coisas que estão a fugir-me das mãos e outras que ameaçam fazê-lo, o meu problema é distinguir entre aquelas por que ainda vale a pena lutar e aquelas que devem ser deixadas ir sem pena. Ou com pena. A pena pior, minha filha, não é a que se sente no momento, é a que se vai sentir depois, quando já não houver remédio. Diz-se que o tempo tudo cura. Não vivemos bastante para lhe tirar a prova.”

(José Saramago, *A Caverna*, pág. 156)

Sumário

SUMÁRIO	VIII
LISTA DE QUADROS.....	X
LISTA DE FIGURAS	XI
LISTA DE TABELAS	XIII
APRESENTAÇÃO	18
INTRODUÇÃO	22
Turismo: produto complexo	22
ESTUDO 1: DESENVOLVIMENTO DE FERRAMENTA BASEADA NA TÉCNICA COMPUTACIONAL ANÁLISE DE SENTIMENTOS PARA A GERAÇÃO DE DADOS PROSPECTIVOS DE AVALIAÇÕES <i>ON-LINE</i> DE DESTINOS TURÍSTICOS	28
Análise de Sentimento (AS)	34
Estudos empíricos sobre o uso da Análise de Sentimento (AS) em análises de opiniões sobre produtos	40
OBJETIVO.....	44
MÉTODO.....	45
<i>Ambiente da coleta dos dados</i>	<i>45</i>
<i>Materiais e equipamentos</i>	<i>45</i>
<i>Procedimentos</i>	<i>47</i>
RESULTADOS	58
DISCUSSÃO	75
ESTUDO 2: ANÁLISE DE AVALIAÇÕES <i>ON-LINE</i> DE DESTINOS TURÍSTICOS POR TURISTAS-CONSUMIDORES: INTERPRETAÇÕES A PARTIR DO <i>BEHAVIORAL PERSPECTIVE MODEL</i> (BPM).	86
Avaliação de produtos turísticos em ambientes <i>on-line</i>	89
Análise do comportamento do consumidor: o <i>Behavioral Perspective Model</i> (BPM)	92
OBJETIVOS	103
MÉTODO.....	104

<i>Base de dados</i>	104
<i>Materiais e equipamentos</i>	104
<i>Procedimentos</i>	104
RESULTADOS	105
DISCUSSÃO	158
DISCUSSÃO GERAL	168
REFERÊNCIAS BIBLIOGRÁFICAS	173

Lista de Quadros

Quadro 1: Cesta de produtos e serviços heterogêneos componentes do produto turístico.....	25
Quadro 2: Estrutura de componentes do produto turístico segundo Acerenza (2002).	26

Lista de Figuras

Estudo 1

Figura 1: Fluxograma do algoritmo de correção ortográfica. Adaptado de Norvig (2016) por Araújo (2017) e desse extraído	49
Figura 2: Pseudocódigo da pesquisa por expressões candidatas. Extraído de Araújo (2017) a partir de modificação do algoritmo de Oliveira e Brito (2015).	52
Figura 3: Resultado do teste do PMI, conforme conduzido em Araújo (2017).	55
Figura 4: Fluxograma para identificação da polaridade de uma palavra, extraído de Araújo (2017)...	57
Figura 5: Estrutura da Ferramenta SentimentALL.	59
Figura 6: Modelo relacional parcial descritivo do armazenamento dos dados coletados.....	60
Figura 7: Ilustra o procedimento de normalização dos dados textuais coletados.	61
Figura 8: Relações de dependência em uma frase representadas um Grafo de Dependência.....	64
Figura 9: Ferramenta on-line de análise e anotação manual da relação entre aspectos e polaridades.	68
Figura 10: Análises de aspecto, e aspecto e polaridade de comentários simples feitas pelas duas versões do SentimentALL, conforme os critérios Precision, Recall e F-Measure.	69
Figura 11: Análises de aspecto, e aspecto e polaridade de comentários complexos feitas pelas duas versões do SentimentALL, conforme os critérios Precision, Recall e F-Measure.	69
Figura 12: Modelo relacional parcial para coleta de dados, pré-processamento, análises e totalizações.	70
Figura 13: Dashboard da Ferramenta SentimentALL (Souza, 2017).	71

Estudo 2

Figura 1: Representação esquemática do Behavioral Perspective Model (BPM; adaptada de Foxall, 2010, p. 9; extraída de Nalini, Cardoso, & Cunha, 2013, p. 494).	95
Figura 2: Adaptação, para o contexto de avaliações, da representação esquemática do <i>Behavioral Perspective Model</i> (BPM; adaptada de Foxall, 2010, p. 9; extraída e adaptada de Nalini; Cardoso; Cunha, 2013, p. 494).....	105
Figura 3: <i>Continuum</i> de fechamento-abertura do cenário do comportamento do consumidor-turista baseado na quantidade de avaliações/comentários. (Baseado em Foxall, 1993).	112
Figura 4: Mostra dos valores das proporções gerais de avaliações positivas (AvP) e negativas (AvN) ponderadas em relação às proporções relativas de comentários atribuídos a cada um dos 100 destinos turísticos analisados, conforme posição assumida por estes nos <i>rakings</i> obtidos com a disposição decrescente dos valores. Os expoentes das equações de ajuste (funções potência) indicam a taxa média de decréscimo dos valores da variável a cada posição sucessiva. Curvas correspondentes aos	

<i>rankings</i> na Tabela 3, onde os nomes dos destinos e os valores específicos para cada destino podem ser observados.	116
Figura 5: Mostra os valores das proporções de avaliações positivas (AvP; pontos pretos) e negativas (AvN; pontos brancos) para acomodações (ACO; círculos), restaurantes (RES; triângulos) e atrativos (ATR; quadrados) ponderados em relação às proporções relativas de comentários atribuídos a cada componente avaliado, considerado o conjunto dos 100 destinos turísticos analisados, conforme posição assumida por estes nos <i>rankings</i> por componente obtidos com a disposição decrescente dos valores. Os expoentes das equações de ajuste (funções potência) indicam a taxa média de decréscimo dos valores da variável a cada posição sucessiva. Curvas correspondentes aos <i>rankings</i> na Tabela 4, Tabela 5, Tabela 6, onde os nomes dos destinos e os valores específicos para cada destino podem ser observados.	128
Figura 6: Mostra os valores da proporção relativa ponderada de <i>likes</i> para comentários sobre cada um dos 100 destinos turísticos analisados, conforme posição assumida por estes no <i>ranking</i> obtido com a disposição decrescente dos valores (Tabela 9).....	136
Figura 7: Mostra os valores da proporção relativa ponderada de <i>likes</i> (<i>wpLikes</i>) para comentários sobre os componentes (ACO, RES e ATR) de cada um dos 100 destinos turísticos analisados, conforme posição assumida pelos destinos nos <i>rankings</i> obtidos com a disposição decrescente dos valores (Tabela 10, Tabela 11 e Tabela 12). Os expoentes das equações de ajuste (funções potência) indicam a taxa média de decréscimo dos valores da variável a cada posição sucessiva.	147
Figura 8: Representação do voto útil no <i>TripAdvisor</i> ® (Fonte: <i>Site TripAdvisor</i> ®)	162
Figura 9: Perfil de colaborador com votos úteis (<i>likes</i>) (Fonte: <i>Site TripAdvisor</i> ®)	165
Figura 10: Composição do nível de <i>TripColaborador</i> (Fonte: <i>Site TripAdvisor</i> ®)	165
Figura 11: Selos dos votos úteis do <i>TripAdvisor</i> ® (Fonte: <i>Site TripAdvisor</i> ®)	166
Figura 12: Número total geral de votos úteis (<i>likes</i> ” 🍷) para comentários sobre Acomodações, Restaurantes e Atrações; número total de votos úteis por ACO, RES, ATR feitos por usuários de cada nível.	170

Lista de Tabelas

Estudo 1

Tabela 1: Padrões morfológicos das expressões multipalavras.....	51
Tabela 2: Relações de dependência no Universal Dependencies (UD).....	55
Tabela 3: Exemplo de classificação gramatical no processo de POS-Tagging.	62
Tabela 4: Comentários de turistas com expressões multipalavras.....	63
Tabela 5: Lista relacionada às palavras identificadas como opiniões.	65
Tabela 6: Lista de palavras negativas e adversativas.	65
Tabela 7: Lista relacionada às palavras identificadas como opiniões com polaridade final.....	66
Tabela 8: Palavras identificadas como opinativas.	66
Tabela 9: Variáveis geradas pela SentimentALL v2.	71
Tabela 10: Mostra os 15 aspectos positivos com maior quantidade de incidência nos seis destinos turísticos com maior quantidade de aspectos encontrados nos comentários em língua portuguesa. .	79

Estudo 2

Tabela 1: Mostra o número total geral de comentários (nTC) e o número total e percentuais de comentários por componente avaliado (Acomodações [Aco], Restaurantes [Res] e Atrações [Atr]) para os 100 destinos turísticos mais avaliados no <i>TripAdvisor</i> ® nos meses de fevereiro e março de 2017. Os números totais agregados (Σ) e percentuais para o conjunto completo dos 100 destinos, assim como os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo, são mostrados nas linhas finais da tabela.	107
Tabela 2: Mostra as proporções medianas, médias (com desvios-padrão), mínimas e máximas entre os números de avaliações e os números de comentários por total geral, totais por tipos de avaliação (positivas e negativas) e totais por componentes avaliados (acomodações, restaurantes e atrações)	110
Tabela 3: Mostra as <i>rankings</i> dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) ponderadas em relação às proporções relativas de comentários ($NTC_d/NC_{máx}$) atribuídos a cada um dos 100 destinos turísticos analisados.	113
Tabela 4: Mostra as <i>rankings</i> dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ACOMODAÇÕES (ACO) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_ACO_d/NC_{máx_ACO}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos <i>rankings</i> encontra-se na Figura 5.	119
Tabela 5: Mostra as <i>rankings</i> dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente	

RESTAURANTES (RES) ponderadas em relação às proporções relativas de comentários atribuídos ao componente (NTC_RES _d /NC _{máx} _RES), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos <i>rankings</i> encontra-se na Figura 5.	122
Tabela 6: Mostra os <i>rankings</i> dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ATRATIVOS (ATR) ponderadas em relação às proporções relativas de comentários atribuídos ao componente (NTC_ATR _d /NC _{máx} _ATR), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos <i>rankings</i> encontra-se na Figura 5.	125
Tabela 7: Mostra os coeficientes de correlação ρ de Spearman entre os valores das proporções de avaliações positivas (AvP) para acomodações (ACO), restaurantes (RES) e atrativos (ATR) ponderados em relação às proporções relativas de comentários atribuídos a cada componente avaliado, considerado o conjunto dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%).	129
Tabela 8: Mostra os coeficientes de correlação ρ de Spearman entre os valores das proporções de avaliações negativas (AvN) para acomodações (ACO), restaurantes (RES) e atrativos (ATR) ponderados em relação às proporções relativas de comentários atribuídos a cada componente avaliado, considerado o conjunto dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%).	129
Tabela 9: Mostra, para cada um dos 100 destinos turísticos analisados, o número total de <i>likes</i> (nTLikes) atribuídos aos comentários, o número total de comentários (nTC), a proporção de <i>likes</i> por comentário (nTLikes/nTC), a proporção de comentários em relação ao número máximo de comentários observado (nTC/nC _{máx}) e a proporção de <i>likes</i> ponderada pelo número proporcional de comentários ($wpLikes$). O <i>ranking</i> dos destinos turísticos foi obtido com a disposição decrescente dos valores $wpLikes$. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (nTLikes) e nTC, são mostrados nas linhas finais da tabela.	133
Tabela 10: Mostra, considerado o aspecto ACOMODAÇÕES (ACO) para cada um dos 100 destinos turísticos analisados, o número total de <i>likes</i> atribuídos aos comentários (nTLikes_ACO), o número total de comentários (nTC_ACO), a proporção de <i>likes</i> por comentário (nTLikes_ACO/nTC_ACO), a proporção de comentários em relação ao número máximo de comentários observado (nTC_ACO/nC _{máx}) e a proporção de <i>likes</i> ponderada pelo número proporcional de comentários ($wpLikes_ACO$). O <i>ranking</i> dos destinos turísticos foi obtido com a disposição decrescente dos valores $wpLikes_ACO$. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (nTLikes_ACO) e nTC, são mostrados nas linhas finais da tabela.	138
Tabela 11: Mostra, considerado o aspecto RESTAURANTES (RES) para cada um dos 100 destinos turísticos analisados, o número total de <i>likes</i> atribuídos aos comentários (nTLikes_RES), o número total de comentários (nTC_RES), a proporção de <i>likes</i> por comentário (nTLikes_RES/nTC_RES), a proporção de comentários em relação ao número máximo de comentários observado (nTC_RES/nC _{máx}) e a proporção de <i>likes</i> ponderada pelo número proporcional de comentários ($wpLikes_RES$). O <i>ranking</i> dos destinos turísticos foi obtido com a disposição decrescente dos valores $wpLikes_RES$. Os	

valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (nTLikes_RES) e nTC, são mostrados nas linhas finais da tabela.	141
Tabela 12: Mostra, considerado o aspecto ATRAÇÕES (ATR) para cada um dos 100 destinos turísticos analisados, o número total de <i>likes</i> atribuídos aos comentários (nTLikes_ATR), o número total de comentários (nTC_ATR), a proporção de <i>likes</i> por comentário (nTLikes_ATR /nTC_ATR), a proporção de comentários em relação ao número máximo de comentários observado (nTC_ATR /nC _{máx}) e a proporção de <i>likes</i> ponderada pelo número proporcional de comentários (wpLikes_ATR). O <i>ranking</i> dos destinos turísticos foi obtido com a disposição decrescente dos valores wpLikes_ATR. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (nTLikes_ATR) e nTC, são mostrados nas linhas finais da tabela.	144
Tabela 13: Mostra os coeficientes de correlação <i>rho</i> de Spearman entre os valores das proporções ponderadas de <i>likes</i> (wpLikes) para comentários sobre acomodações (ACO), restaurantes (RES) e atrativos (ATR) emitidos para cada um dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%; bicaudado).....	148
Tabela 14: Mostra os coeficientes de correlação <i>rho</i> de Spearman entre os valores das proporções relativas ponderadas de <i>likes</i> (wpLikes) e as proporções relativas de comentários sobre acomodações (ACO), restaurantes (RES) e atrativos (ATR) emitidos para cada um dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%; bicaudado).....	148
Tabela 15: Mostra as estatísticas <i>U</i> de Mann-Whitney para comparações entre os números totais médios de <i>likes</i> emitidos por <i>Trip</i> Colaboradores classificados nos níveis mínimo (N1 e N2) e máximo (N5 e N6) de experiência como avaliadores, para os componentes acomodações (ACO), restaurantes (RES) e atrações (ATR), e os valores médios comparados (em número de <i>likes</i> e posts), o valor p. observado e os valores de tamanho do efeito (<i>r</i>) para cada comparação (n = 100; ** Nível de confiança = 0,01%; bicaudado).	150
Tabela 16: Mostra <i>rankings</i> dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações positivas (nT_AvP; à esquerda) e do somatório dos números totais de avaliações “Muito bom” (MB) e “Excelente” (EX), nT_(MB+EX) à direita, conforme escala avaliativa do <i>TripAdvisor</i> ®.....	152
Tabela 17: Mostra <i>rankings</i> dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações negativas (nT_AvN; à esquerda) e do somatório dos números totais de avaliações “Horrrível” (HO) e “Ruim” (RU), nT_(HO+RU) à direita, conforme escala avaliativa do <i>TripAdvisor</i> ®.....	155
Tabela 18: Mostra os coeficientes de correlação <i>rho</i> de Spearman entre as variáveis número total de avaliações positivas (nT_AvP), número total de avaliações negativas (nT_AvN), somatório dos números totais de avaliações “Horrrível” e “Ruim” (nT_[HO+RU]) e somatório dos números totais de avaliações “Muito bom” e “Excelente” (nT_[MB+EX]). (n = 100; ** Nível de significância = 0,01%).	157

RESUMO

A grande quantidade de informações disponíveis na *internet* tem viabilizado numerosas investigações de caráter multidisciplinar com o objetivo de entender nuances do comportamento de consumo humano, especialmente identificar as opiniões das pessoas sobre produtos e serviços. A partir do *Behavioral Perspective Model (BPM)*, análises do comportamento do consumidor podem ser realizadas considerando variáveis antecedentes (cenário do comportamento e história de aprendizagem do consumidor) e consequências (reforços e punições, utilitários e informativos) à ocorrência do comportamento. A presente tese investigou o comportamento de consumo no contexto do turismo, com o *BPM* como suporte teórico para interpretações de dados verbais extraídos de comentários disponíveis no *TripAdvisor®*, *website* do setor. Para tanto, analisaram-se as respostas verbais de turistas-consumidores no processo de “opinar *on-line*” sobre componentes de produtos turísticos (especificamente, acomodações – ACO, restaurantes – RES e atrações – ATR). Os participantes da pesquisa foram os indivíduos (desconhecidos) que, entre o início de fevereiro e o final de março de 2017, emitiram, no *TripAdvisor®*, 6.438.497 comentários distribuídos entre os 100 destinos turísticos brasileiros mais avaliados. Descrita em dois estudos (Estudo 1 [E1] e Estudo 2 [E2]), a pesquisa de tese se propôs a: a) extração e análise de informações verbais (comentários) dos turistas com base na técnica computacional Análise de Sentimentos (AS); extração do número de indicações avaliativas dos componentes do produto turístico (ACO, RES e ATR) emitidas por turistas-consumidores com diferentes *status* como *TripColaboradores*; extração do número de votos úteis (*Likes*) nos comentários; b) descrever o responder avaliativo polarizado atribuído aos 100 destinos turísticos avaliados e analisar interpretativamente tal responder a partir do *BPM*. O E1 resultou no desenvolvimento da ferramenta *SentimentALL*, com foco no módulo de AS, e na geração das variáveis primárias exploradas no E2 ($n = 197$). No E2, dados gerados no E1 e medidas derivadas foram explorados (descritos em *rankings* e análises de correlação) e interpretados com recurso ao referencial conceitual do *BPM*. De caráter fundamentalmente exploratório, o esforço interpretativo sugeriu linhas profícuas de pesquisa e a utilidade da integração entre conhecimentos computacionais e psicológicos.

Palavras-Chave: Comportamento do consumidor; *Behavioral Perspective Model*; Análise de Sentimento; Turismo; Avaliações positivas; Avaliações negativas.

ABSTRACT

The vast amount of information available on the Internet have enabled numerous multidisciplinary investigations aimed to understand nuances of human consumption behavior, especially to identify people's opinions about products and services. From Behavioral Perspective Model (BPM), consumer behavior analysis can be conducted focusing on antecedent variables (behavioral setting and consumer learning history) and consequences (reinforcement and punishment, utilitarian and informative) to the occurrence of behavior. The present thesis investigated consumption behavior in the context of tourism, with BPM as theoretical support for interpretations of verbal data extracted from comments available on the TripAdvisor®, a website about tourism. Verbal responses of tourism consumers, engaged in the process of online evaluation of components of tourism products (specifically, Accommodations [ACO], Restaurants [RES] and Attractions [ATR]), were analyzed. Research participants were the unknown individuals who, between the beginning of February and the end of March 2017, emitted 6.438.497 comments distributed among the 100 most evaluated Brazilian touristic destinations at TripAdvisor®. In two studies (Study 1 [E1] and Study 2 [E2]), the thesis research aimed at: a) extraction and analysis of tourist's verbal information (commentaries) through a Sentiment Analysis (SA) computational technique; extraction the number of touristic product component (ACO, RES e ATR) evaluative indications emitted by tourism consumers with different statuses as TripColaborators; extraction of the number of votes (Likes) to the comments; b) describe the polarized evaluative response attributed to the 100 evaluated touristic destinations and interpret such responding considering BPM concepts. E1 resulted in the successful development of the SentimentALL tool, focusing on the AS module, and the generation the primary variables explored in E2 (n = 197). In E2, data generated in E1 and derived measures were explored (described in rankings and correlation analyzes) and interpreted using the BPM conceptual framework. With a fundamental exploratory character, the interpretative effort suggested profitable research lines and utility of the computational and psychological knowledges integration.

Keywords: Consumer behavior; Behavioral Perspective Model; Sentiment Analysis; Tourism; Positive evaluations; Negative evaluations

APRESENTAÇÃO

O sistema de documentos em hipermídia interligados e executados na Internet conhecido como *World Wide Web* (*WWW* ou *Web*) define vasto campo para prospecção e análise de informações sobre produtos, serviços e pessoas. No sistema *Web* encontra-se ampla variedade de páginas de informação dinâmicas e acessíveis, que permitem rápida propagação dos mais diversos tipos de dados. Há grande interesse — por parte de empresários, corporações, detentores de patentes de produtos, indústria do entretenimento, indústria farmacêutica, indústria turística, governos, dentre outros — no entendimento dos fatores que determinam como opiniões acerca de produtos, serviços, marcas, empresas ou pessoas passam a ser compartilhadas por milhares de indivíduos em curto espaço de tempo no ambiente comunicacional definido pelo sistema *Web*. Há, adicionalmente, significativo interesse no conhecimento sobre o que pode ser extraído dessas opiniões que permita descrever e qualificar padrões de comportamento de consumo e, com isso, viabilizar estratégias mercadológicas de influência sobre o consumidor e aperfeiçoamento de processos negociais.

O sistema *Web* mudou (e tem mudado) muito desde que Tim Berners-Lee¹ publicou o primeiro *site on-line*, em 6 de agosto de 1991. Naquele momento, *sites* com conteúdos estáticos, produzidos em sua maioria por empresas e instituições, tinham pouca interatividade com os usuários. Com o surgimento de empresas no ambiente virtual (como a *Amazon*®, o *Yahoo!*®, o *eBay*®, a *Geocities*®, a *Altavista*®, dentre outras), foi iniciada uma revolução no processo de compra e venda e na forma como se compartilham conteúdos digitais, e essas mudanças se refletiram no comportamento das pessoas, na condição de agentes ativos desse processo de interação.

A mudança ocorrida no sistema *Web* pode ser sucintamente apresentada a partir do incremento do alcance de suas funcionalidades. No começo, a *Web* funcionava como uma rede de computadores. Entre os anos de 2002 e 2004, com a

¹ Físico britânico, cientista da computação e professor do *Massachusetts Institute of Technology* (MIT), criador da *World Wide Web* (WWW).

Web 2.0², é iniciado, de forma mais contundente, o processo de interatividade nesse meio, ou seja, o sistema passou a viabilizar maior interação dos usuários com os conteúdos de suas páginas e, assim, permitir maior participação desses, o que marca o início da *Web* social. Mídias sociais (como o *YouTube*®, o *Facebook*®, o *Twitter*®, a *Wikipédia*®, o *Flickr*®) agruparam quantidade imensa de usuários, que terminaram envolvidos em milhões de interações. O potencial mercadológico viabilizado pela *Web* 2.0 foi imediatamente reconhecido pelo mundo corporativo, que não ficou à margem desse processo. No plano das relações de consumo, passou-se a considerar que o que era falado por consumidores nesse ambiente, ou “rastros” de qualquer outro tipo por eles deixados, se analisados de forma adequada, poderiam mudar o processo de gestão nas empresas, a forma como produtos e serviços eram ofertados e vendidos, e o modo de relacionamento entre as empresas e seus consumidores.

Agrega-se a esse movimento de exploração do sistema *Web*, uma vasta gama de interesses, desde aqueles econômicos e antropológicos, até os de cunho psicológico e social. Diante disso, técnicas computacionais têm sido desenvolvidas com o intuito de mapear a *Web* (páginas, portais e redes sociais virtuais na Internet) de forma a especificar os elementos que a estruturam e como esses elementos se relacionam e, assim, buscar compreender os fatores que transformam produtos ou serviços em *hits* de sucesso, campanhas em ações bem-sucedidas ou fracassadas, ou o que provoca a ascensão imediata de alguns indivíduos ou sua decadência, dentre outros. Interesses se colocam, inclusive, no plano do aperfeiçoamento das metodologias de acesso e coleta de dados sobre o comportamento individual de consumidores quando estes interagem com as mais variadas dimensões de produtos ou serviços, tal como essas terminam representadas na *Web*. Tal aperfeiçoamento possibilita aos campos de estudo do consumo trabalhar com dados cujo volume, variedade e poder descritivo potencial de idiosincrasias das relações de consumo têm sido entendidos como substancialmente superiores àqueles obtidos com metodologias tradicionais.

A presente tese buscou extrair e analisar relatos avaliativos dos componentes acomodações (hotéis, pousadas, albergues, etc.), restaurantes e atrações de destinos turísticos feitos, na forma de comentários, por turistas-consumidores no *site TripAdvisor*® (<http://www.tripadvisor.com.br>), considerado o maior *site* de informações

² O termo “Web 2.0” foi criado por Tim O’Reilly, ao apresentar uma evolução da *Web* original (eventualmente denominada “Web 1.0”), e definitivamente popularizado em 2004.

a viajantes dentre todos da indústria turística mundial³, que abriga centenas de milhões de comentários e avaliações sobre componentes de milhares de destinos turísticos mundiais.

A etapa inicial do trabalho, descrita adiante como **Estudo 1**, foi conduzida com recurso a técnicas computacionais. Análise do conteúdo dos comentários sobre os 100 destinos turísticos brasileiros mais avaliados, extraídos do *TripAdvisor*® nos meses de fevereiro e março de 2017, foi realizada com a técnica denominada “Análise de Sentimento” (AS). A AS permitiu, por meio da utilização de uma série de algoritmos, a descrição dos relatos avaliativos dos turistas-consumidores usuários do *website* nos seus componentes opinativos indicadores de sentimentos positivos ou negativos vividos quando da experiência turística nas localidades elegidas para o estudo. Segundo Liu (2010, p. 6), “a análise de sentimento tenta, essencialmente, inferir o sentimento das pessoas baseando-se em suas expressões de linguagem”. Operacionalmente, a AS é feita por meio de Processamento de Linguagem Natural (PLN), com uso de léxicos de sentimentos, aprendizagem de máquina e, por vezes, com definição de contextos com base em ontologias de domínio e na construção de dicionários de termos afins (*thesaurus*) sobre um ambiente delimitado. Além da aplicação da AS ao conteúdo dos comentários, foram extraídas e quantificadas as seguintes informações: **1)** pontuações atribuídas aos componentes dos destinos turísticos pelos turistas-consumidores membros da “Comunidade do *TripAdvisor*®”, de acordo com escala *Likert* utilizada na plataforma (níveis escalares: “*Excelente*”, “*Muito bom*”, “*Razoável*”, “*Ruim*” e “*Horrível*”); **2)** registros de utilidade (*likes*) dos comentários feitos por turistas-consumidores, e **3)** o nível dos turistas-consumidores na condição de avaliadores dos destinos turísticos (denominados “*TripColaboradores*”), conforme taxonomia utilizada na plataforma.

Após a etapa computacional, na segunda etapa descrita adiante como **Estudo 2**, *rankings* baseados nos dados gerados e medidas derivadas (por exemplo, o número total e por componente do produto turístico de relatos indicativos de “sentimentos positivos” e “sentimentos negativos”, razões entre essas medidas ponderadas pela proporção de comentários, etc.) foram obtidos, e análises de correlação entre as distribuições ordenadas foram feitas. Os resultados foram

³ Fonte: comScore Media Metrix para os sites do *TripAdvisor*®, âmbito internacional, julho de 2015.

interpretados com base em conceitos do *Behavioral Perspective Model* (BPM; Foxall, 1996, 2010).

INTRODUÇÃO

Os conceitos “turismo” e “produto turístico” são conteúdos transversais aos dois estudos da presente tese. Portanto, apresentam-se abaixo noções relevantes ao esclarecimento desses conceitos tal como explorados no âmbito do estudo científico do fenômeno turístico, e nos limites do necessário aos estudos desenvolvidos.

Turismo: produto complexo

A pesquisa científica em turismo é relativamente nova, visto que até a segunda metade do século XX, as investigações realizadas nesse contexto se constituíam, em sua maioria, de respostas às demandas empresariais e careciam de instrumentos e metodologias adequados para análises mais consistentes das informações obtidas nos estudos realizados (Candela & Figini, 2012; Ashworth, 1989; Cooper, Fletcher, Fyall, Gilbert, & Wanhill, 2008). Segundo Sinclair e Stabler (1997):

“A investigação no turismo desenvolveu-se apenas a partir da segunda metade da década de 1980 em termos de metodologia científica, com o nascimento de periódicos científicos especializados, departamentos de universidades, centros de pesquisa e de graduação e pós-graduação”. (Sinclair & Stabler, 1997 *apud* Candela & Figini, 2012, p. 1).

O que veio à tona a partir de estudos mais elaborados no âmbito do turismo foi a consolidação do seu caráter interdisciplinar, pois este pode ser compreendido como um sistema formado pela sobreposição e interação de mais de uma área do conhecimento (economia, sociologia, antropologia, geografia, psicologia, etc.). A Organização Mundial do Turismo (OMT; do inglês *United Nations World Tourism Organization* – UNWTO) apresenta, em um de seus manuais técnicos, a seguinte consideração:

“O turismo é considerado uma atividade essencial para a vida das nações por causa de seus efeitos diretos sobre os setores sociais, culturais, educacionais e econômicos das sociedades no âmbito de um país e em suas relações internacionais” (OMT, 1995, p.1)

De acordo com a OMT (2010), é possível constatar a relevância econômica do setor do turismo, pois este representa 9% do Produto Interno Bruto (PIB) e 8% no nível de emprego mundial. Porém, Candela e Figini (2012) ressaltam que mensurar a

contribuição econômica do turismo na economia de um país não é uma tarefa simples, pois existem problemas estatísticos que precisam ser discutidos, e que estão diretamente relacionados com a definição e classificação da oferta e da demanda do setor. Por exemplo, quanto à demanda, o indivíduo que faz refeições e/ou compra um sapato no destino turístico estará realizando atividade específica do turismo? A resposta a essa questão depende do nível e do contexto da análise, já que as referidas atividades poderiam ter sido realizadas também no local em que o turista vive. Contudo, ao ser considerada a categoria “turista-consumidor”, tem-se o entendimento de que determinadas atividades em contexto turístico podem modificar a composição e o valor de produtos e serviços, o que acarreta diretamente em variações do consumo na localidade, caracterizadas por mudanças nas condições para as compras e na quantidade de vendas. Quanto à oferta, há uma série de empresas que fornecem produtos e serviços relevantes para as atividades turísticas. Contudo, tais empresas não se limitam a atender somente à demanda definida pelo contingente de turistas na localidade. Assim, definir os limites para as análises dos fenômenos da indústria turística é tarefa complexa (cf. Candela & Figini, 2012).

A multiplicidade de setores que são afetados pelas atividades relacionadas ao turismo demonstra que os estudos nessa área, mesmo quando restritos a um determinado escopo, têm que considerar amplos conjuntos de variáveis nas análises. Goeldner e Ritchie (2002) apontam que quando o estudo do turismo é iniciado, automaticamente vem à tona uma reflexão sobre as pessoas e suas motivações quando engajadas em atividades turísticas. Ou seja, as pessoas viajam por vários motivos, desde a necessidade de visitar parentes e/ou amigos, ou pelo prazer de passar as férias em lugar acolhedor, e até para realizar negócios ou tratamentos de saúde. Nas viagens, as pessoas podem buscar participar de eventos e/ou atrações proporcionadas pelo local, ou podem simplesmente ficar no quarto do hotel a maior parte do tempo, o que, nesse caso em específico, faz do atendimento e do trabalho de quarto do hotel um fator essencial para um eventual retorno. Além disso, as pessoas podem ser adeptas de esportes e/ou gostar de aventuras, o que torna os passeios programados um fator preponderante para a estadia ser positivamente avaliada. Além desses aspectos, pessoas têm diferentes perfis, que podem ser categorizados a partir de dados sociodemográficos (idade, sexo, escolaridade, religião, renda, etc.) ou podem ser analisados a partir de elementos mais subjetivos, relacionados ao retorno do turista à localidade, à indicação positiva ou negativa da

localidade a um amigo ou com comentários sobre a localidade realizados em *sites* especializados na internet.

Considerando a complexidade do turismo como objeto de estudo, tem-se também o entendimento de que a definição de turismo não é uma tarefa trivial. A Organização Mundial do Turismo (OMT, 1995) afirma que o turismo deve ser concebido como o conjunto das relações entre turistas-consumidores e serviços que resulta da mudança voluntária e temporária de residência motivada por razões outras que não o trabalho, por período não superior a 12 meses. Padilla (1992) apresenta definição que expressa essa concepção:

“O turismo é um fenômeno social que consiste no deslocamento voluntário e temporário de indivíduos ou grupos de pessoas, que, fundamentalmente por motivos de recreação, descanso, cultura ou saúde, saem do seu local de residência habitual para outro, no qual não exercem nenhuma atividade lucrativa nem remunerada, gerando múltiplas inter-relações de importância social, econômica e cultural” (Padilla, 1992, p. 19).

Na literatura geral em turismologia (Acerenza, 2002; Andrade, 2000; Beni, 2001; Arendit, 2002), as definições de turismo são, não raramente, apresentadas a partir de diferentes pontos de vista. Segundo Mathieson e Wall (1982), o turismo é "*a circulação temporária de pessoas para destinos diversos de seus locais habituais de trabalho e de residência; as atividades desenvolvidas durante a sua estadia nesses destinos; e as instalações criadas para atender às suas necessidades.*" (p. 1). Em esforço mais de caracterizar o conceito, Goeldner e Ritchie (2002) afirmam que, para a configuração de uma definição abrangente do turismo, as perspectivas de quatro grupos de atores devem ser consideradas: **1)** os turistas (consumidores): os indivíduos que procuram vivenciar experiências e obter satisfações físicas e psicológicas através da prática do turismo; suas motivações, suas experiências, suas condições financeiras (maiores ou menores restrições orçamentárias), dentre outros aspectos, determinarão as escolhas dos destinos turísticos, os períodos para viajar, as atividades apreciadas, etc.; **2)** as empresas que fornecem produtos e serviços turísticos (ofertantes): os empresários do turismo buscam oportunidades para obter lucro através do oferecimento de produtos e serviços que sejam adequados ao mercado turístico; **3)** os governos: gestores (municipais, estaduais, nacionais, internacionais) das localidades que se constituem destinos turísticos, que podem entender o turismo como um fator de riqueza econômica de suas jurisdições (municípios, estados, países)

ao considerarem, por exemplo, a perspectiva de aumento da renda dos cidadãos locais, com apoio aos investimentos em turismo pelo empresariado da localidade, e **4)** as pessoas que moram nas localidades turísticas e são afetadas pelos fluxos e dinâmicas das atividades turísticas, desde de intercorrências nas suas vidas cotidianas e práticas socioculturais, e possibilidades de benefícios com oportunidades de emprego e ganhos financeiros (cf. Goeldner & Ritchie, 2002; Padilla, 1992).

Segundo Candela e Figini (2012), a atividade turística é expressa a partir da demanda por produtos e serviços no destino turístico. Como os produtos e serviços turísticos são diversos, heterogêneos e nem sempre voltados apenas para a atividade turística, entende-se que tais propriedades configuram o amálgama de préstimos que viabilizam o turismo (o dito “produto turístico”) como um produto complexo, de natureza intangível, difícil de ser mensurado e sistematizado. Para qualquer entendimento aprofundado do turismo, é necessário definir categorias e entender interações e processos que fundamentam a especificação e “tangibilidade possível” do produto turístico. Duas taxonomias propostas com o intuito de organizar as informações nesse domínio são as de Candela e Figini (2012) e Acerenza (2002).

Para Candela e Figini (2012), o produto turístico é uma “cesta” de diferentes produtos e serviços usada para atender as pessoas durante as experiências turísticas que vivem. O Quadro 1 representa esquematicamente o conceito de produto turístico segundo Candela e Figini (2012):

Quadro 1: Cesta de produtos e serviços heterogêneos componentes do produto turístico.

TRANSPORTES	ALOJAMENTO E REFEIÇÕES
Navio	Parques para <i>camping</i>
Bicicleta	Hotéis
Avião	Motéis
Trens	Apartamentos alugados
Ônibus	Restaurantes
etc.	Bares, etc.
ATRAÇÕES	COMPRAS
Atividades recreativas	Lojas de <i>souvenir</i>
Parques temáticos	Lojas de arte
Eventos	Lojas de artesanato
Congressos	Produtos locais
Festivais, etc.	Mercado de produtos de fazenda, etc.
BASE DE DADOS DOS RECURSOS NATURAIS, CULTURAIS E ORGANIZACIONAIS	
Fonte: Candela e Figini (2012, p. 33, adaptação e tradução da autora).	

A cesta de produtos heterogêneos de Candela e Figini (2012) está estruturada em quatro grupos principais: **1)** transportes: meios de locomoção por via rodoviária,

marítima, aérea, etc.; **2)** hospedagem e refeições: desde luxuosos estabelecimentos (hotéis cinco estrelas, restaurante com *chefs* respeitados) até bares pitorescos, barracas em praias; **3)** atrações: desde festivais de música e atividades em parques abertos até condições relacionadas à inovação dos produtos ofertados na localidade (por exemplo, um hotel que ofereça uma piscina no último andar com uma bela vista) — trata-se do grupo que apresenta a maior diversificação de serviços turísticos, tendo relação com as preferências dos consumidores, que vão, e 4) compras: que envolve os vários estabelecimentos que atendem o turista, logo vários tipos de profissionais (como comerciantes varejistas, artesãos, artistas, etc.).

Acerenza (2002), por sua vez, representa o produto turístico com uma estrutura organizada em três componentes principais, conforme pode ser observado no Quadro 2 em representação esquemática elaborada por Cardoso (2012).

Quadro 2: Estrutura de componentes do produto turístico segundo Acerenza (2002).

P R O D U T O T U R Í S T I C O	ATRATIVOS Definem e geram a atratividade do destino turístico	de Lugar	naturais usos e costumes infraestrutura
		de Eventos	feiras e exposições congressos e convenções acontecimentos especiais
	FACILIDADES Viabilizam e permitem a permanência no destino turístico	Alojamento	hotéis, motéis, pousadas albergues <i>campings</i> condomínios
	Alimentação	restaurantes cafés lanchonetes bares	
	Lazer	distrações entretenimentos diversões esportes	
	Complementos	<i>tours</i> locais excursões visitas	
ACESSO Viabilizam e permitem o deslocamento até o destino turístico	Transporte	aéreo terrestre aquático marítimo aquático fluvial	

Fonte: Cardoso (2012)

Os componentes apresentados no Quadro 2 (atrativos, facilidades e acesso) são compostos por um conjunto de classes (lugar, eventos, etc.) que, por sua vez, possuem subclasses (naturais, usos e costumes, etc.).

Conceituações como as de Candela e Figini (2012) e Acerenza (2002), e perspectivas abrangentes de configuração da noção de produto turístico como a de Goeldner e Ritchie (2002) evidenciam as características que fazem o produto turístico ser entendido como um produto complexo, de “estrutura” multidimensional. Assim sendo, a mensuração válida das atividades turísticas, das condições antecedentes que as favorecem, das condições que a viabilizam e suas consequências, se torna tarefa de elevada complexidade. Nesse sentido, quaisquer esforços de desenvolvimento de metodologias que capturem de maneira confiável aspectos relevantes da interação entre turistas-consumidores e dimensões do produto turístico (cf. Cardoso, 2012) são empreendimentos potencialmente benéficos para as várias áreas do conhecimento que estudam o fenômeno turístico, com fortes implicações para processos gerenciais na indústria turística.

Na presente tese (que não visa qualquer questionamento conceitual relevante às configurações do conceito de produto turístico), a configuração do produto turístico apenas assumiu elementos relevantes da estrutura de dimensões já disponíveis no *site TripAdvisor®*, a saber: acomodações, restaurantes e atrações. Conforme Candela e Figini (2012) e Acerenza (2002), respectivamente, aspectos dos componentes “alojamento e refeições” e “facilidades”, no caso das acomodações e dos restaurantes, e “atrações” e “atrativos”, no caso das atrações. Os comentários de turistas-consumidores deixados no *TripAdvisor®* sobre esses aspectos em milhares de avaliações para centenas de localidades, extraídos e classificados em polaridade no Estudo 1, e analisados e interpretados como comportamento de consumo no Estudo 2, definiram a base de dados primária da tese.

Estudo 1: Desenvolvimento de ferramenta baseada na técnica computacional Análise de Sentimentos para a geração de dados prospectivos de avaliações *on-line* de destinos turísticos

O estudo do comportamento do consumidor, assim como a percepção da importância desse estudo, tem sido impactado de forma profunda pelo uso crescente e intensivo de tecnologias da informação por parcelas cada vez maiores da população (Marko, 2015; McAfee & Brynjolfsson, 2012; Song & Liu, 2017). As pessoas, via o uso de *smartphones*, de pulseiras com sensores, e dos mais variados dispositivos móveis de recepção e transmissão de dados, estão, cada vez mais, instrumentalizadas para coletar e transmitir dados (o que, de fato, pode acontecer não necessariamente de modo intencional em várias ocasiões). Conforme Weigend (2013), antigo cientista-chefe da Amazon.com, hoje diretor do *Social Data Lab* da Universidade de Stanford⁴, “... se o século passado foi marcado pela capacidade de observar as interações da matéria física — por exemplo, tecnologias como o raio-x e o radar — este século será definido pela capacidade de observar as pessoas através dos dados que compartilham” (Weigend, 2013 *apud* Regalado, 2013, p. 2). Tais dados, por sua variedade, volume e velocidade de atualização tornam possível a predição de comportamentos e a possibilidade de entrega de respostas cujas perguntas ainda não foram elaboradas.

De acordo com a *International Data Corporation* (IDC; Gantz & Reinsel, 2011), a quantidade de dados digitais criada a cada ano tem crescido exponencialmente: atingiu 2,8 ZB em 2012, o que corresponde a 2,8 trilhões de gigabytes e, segundo cálculos estimativos realizados, o volume de dados alcançará 40 ZB em 2020. De 2005 a 2020, o universo digital crescerá de 0,13 ZB para 40 ZB, ou 40 trilhões de gigabytes, o que corresponde a 5.200 GB para cada pessoa na Terra. A maior parte das informações no meio digital, cerca de 68% em 2012, tem sido gerada por indivíduos que criam e movem arquivos digitais, assistem TV digital, interagem em mídias sociais, enviam imagens e vídeos feitos em dispositivos móveis através da Internet. Segundo Tucker (2013), “... um trabalhador americano típico produz 1,8 milhões de megabytes de dados por ano. Isso é cerca de 5.000 megabytes por dia, incluindo filmes baixados, arquivos do Word, e-mail e os bits gerados pelos computadores quando as informações são movidas através da Internet” (p. 2).

⁴ <https://datalab.stanford.edu/>

Antes da Web 2.0 (a Web Social), os dados utilizados em pesquisas que realizavam Processamento de Linguagem Natural (PLN) eram extraídos, geralmente, de coleções de textos publicados em mídias tradicionais, como jornais, livros ou revistas, ou seja, eram mais limpos, com uma menor incidência de erros gramaticais e editados (Haddi, 2015). Com o desenvolvimento das redes sociais, da interatividade proporcionada por blogs pessoais, ou pela opção de comentários em sites de todos os tipos de organizações (entretenimento, saúde, educação, turismo, varejo), não apenas a quantidade de dados aumentou significativamente, como também esses dados podem ser encontrados nas mais variadas formas e atualizados velozmente. O termo “*Big Data*” é usado para descrever essa expansão colossal da quantidade de dados no universo digital que, segundo Chen, Mao e Liu (2014), definem-se em conjuntos de dados que não podem ser capturados, gerenciados e processados por computadores comuns dentro de um escopo aceitável. O *big data* reflete mudanças nas mais diversas instâncias da sociedade, desde as relações que os indivíduos estabelecem no meio digital, gerando dados através das mensagens e fotos postadas e compartilhadas, até na análise de dados dos mais diversos meios para a criação de novas estratégias de negócios e de vantagem competitiva (cf. Chen, Mao, & Liu, 2014).

Tais mudanças têm sido consideradas decorrentes dos desafios e oportunidades gerados pelo modelo 3Vs (Laney, 2001), configurado, conceitualmente, pelo aumento do volume (V1), da velocidade (V2) e da variedade (V3) de dados. Muito embora, conforme Chen, Mao e Liu (2014), o modelo não tenha sido criado para definir *Big Data* especificamente, é usado por muitas empresas e grupos de pesquisa — por exemplo, a IBM® (Zikopoulos, 2011), a Gartner® (Beyer, 2011) e a Microsoft® (Meijer, 2011) — como base para explicar o fenômeno. Síntese dos três aspectos é apresentada por McAfee e Brynjolfsson (2012):

(V1) Volume: de acordo com James (2016), desde 2011, a população global de pessoas usando a Internet cresceu 60%, passando de 2 bilhões de usuários em 2011 para mais de 3,4 bilhões em 2016. O aumento do volume dos dados é impressionante em vários aspectos, conforme pode ser observado em análises de acesso a dados: *a)* o uso do *Youtube* em 2016 foi três vezes maior do que nos últimos dois anos, com os usuários compartilhando 400 horas de novos vídeos a cada minuto, por dia; *b)* os usuários do Instagram "curtiram" em torno de 2.5 milhões de *posts* e os do *Facebook* compartilharam mais de 216

mil fotos e, segundo o relatório da Zephoria (2016), publicaram 510 mil comentários e atualizaram 293 mil *status* a cada minuto, e c) no *Google Translate* foram traduzidas cerca de 69.500 milhões de palavras.

(V2) Velocidade: as ferramentas de produção e distribuição contribuíram para a geração de informação em tempo real (ou em um tempo bem menor do que era feito anteriormente). Nesse sentido, é possível analisar a opinião das pessoas sobre um episódio de uma série de televisão no momento em que o episódio é exibido via análise dos textos nas postagens realizadas nas redes sociais. Um grupo de pesquisa do *MIT Media Laboratory* usou dados de localização de celulares para inferir a quantidade de pessoas que estava no estacionamento da *Macy's* (uma loja de departamentos de Nova Iorque, EUA) em uma *Black Friday* no início da temporada de compras de Natal, e isso tornou possível estimar as vendas no varejo naquele dia crítico mesmo antes de a própria *Macy's* ter registrado as vendas.

(V3) Variedade: o que torna o *Big Data* um conceito tão atual é a origem e a forma dos dados. Nesse contexto, podem haver dados estruturados tradicionais (em bancos de dados relacionais), mas a grande maioria é composta por dados semiestruturados (*web logs*, *feeds* de mídia social, arquivos coletados de uma fonte de sensor, e-mail, etc.) e não estruturados (áudio, vídeo, *Webpages* e textos). Dados podem ser mensagens e imagens postadas em redes sociais, atualizações realizadas em aplicativos ou outros *softwares*, podem vir de sensores que estejam nos mais diversos ambientes (no pulso de uma pessoa, em uma casa, nas ruas, em uma floresta, etc.), em sinais de GPS de telefones celulares. Como as fontes são relativamente novas — por exemplo, o *Facebook* foi lançado em 2004, o *Twitter* em 2006, um dos smartphones mais usados nos EUA, o *iPhone*, em 2007 —, ainda há um trabalho computacional imenso para lidar com essa variedade de dados. Isso ocorre porque, tradicionalmente, os bancos de dados estão preparados para lidar com informações estruturadas, o que não é o caso, ao menos não em sua maioria, dos dados presentes nos *big data*.

Chen, Mao e Liu (2014) expandiram o modelo dos 3Vs, acrescentando mais um V, que designa o Valor, ou seja, um grande conjunto de dados só pode ser considerado Big Data se dele puder ser extraído valor. Acrescentam ainda que a definição em 4Vs tem sido amplamente reconhecida, pois destaca o significado e a

necessidade do *Big Data*, assim também como o seu problema mais crítico, que é a descoberta de valor em conjuntos de dados de grande escala, de vários tipos e que são gerados rapidamente.

Segundo McAfee e Brynjolfsson (2012), o contexto atual tem os elementos necessários para o *Big Data* ter emergido, o que pode ser notado a partir da mudança em relação à largura da banda da Internet, o barateamento dos meios de armazenamento e a ampliação exponencial desse contexto com o estabelecimento do armazenamento nas nuvens, a proliferação de uma série de tecnologias que permitem a captura e transmissão de dados, como sensores, smartphones, etc., e o crescente número de pesquisas sobre algoritmos mais eficientes para a criação de ferramentas analíticas. Segundo o relatório da IDC 2011, o *Big Data* não pode ser identificado como uma “coisa” única, mas como um conjunto de atividades dinâmicas que permeiam muitos contextos da Tecnologia da Informação (TI). Gantz e Reinsel (2011, p. 6) acrescentam que “... as tecnologias usadas para pesquisas em *Big Data* descrevem uma nova geração de tecnologias e arquiteturas, projetadas para, economicamente, extrair valor de um grande volume de uma vasta variedade de dados, permitindo a captura, descoberta e/ou análise de dados em alta velocidade” (p. 6).

Esse contexto influencia diretamente os estudos relacionados a PLN, pois expande ainda mais as possibilidades de pesquisas nessa área, considerando que os textos iniciais mais organizados agora compõem uma pequena parte desse cenário. Ao considerar esse aspecto e o objetivo desse estudo, a definição de Liddy (2001) para PLN torna-se apropriada a esse trabalho.

Processamento da linguagem natural é uma variedade de técnicas computacionais utilizadas para a análise e representação de textos que ocorrem naturalmente em um ou mais níveis de análise linguística com o objetivo de obter processamento de linguagem semelhante ao humano para uma variedade de tarefas ou aplicações. (Liddy, 2001, para. 2)

Vários elementos desta definição podem ser assim detalhados (Liddy, 2001): a “variedade de técnicas computacionais” está relacionada à multiplicidade de métodos que existem para os variados processos de análise, assim, se for considerada a Análise de Sentimentos (será apresentada mais sobre essa temática na próxima seção), algumas técnicas computacionais que utilizam dicionários de léxicos podem ser mais adequada em determinados contextos. “Textos de ocorrência natural” tem relação com o fato de que os textos podem estar em quaisquer idiomas, podem ser

orais ou escritos, mas devem ser coletados a partir de um ambiente que proporciona seu uso real (não textos construídos para tal fim). Ao apresentar que na PLN o "processamento da linguagem deve ser semelhante ao processo humano", isso a insere como uma disciplina da Inteligência Artificial (IA), ainda que pela sua concepção, ela dependa de várias outras disciplinas. A expressão "para uma série de tarefas ou aplicações" enfatiza que muitas áreas usam a PLN para alcançar um determinado fim, o que a torna, em várias situações, um meio para realizar uma tarefa específica, e isso ficará evidente nas seções a seguir que tratam da Análise de Sentimentos e dos procedimentos necessários ao desenvolvimento da ferramenta utilizada nesta tese.

Por último, a noção de "níveis de análise linguística", que faz referência ao fato de que existem vários tipos de processamento de linguagem, pois cada nível transmite diferentes tipos de significado. Geralmente, os sistemas de PLN utilizam alguns desses níveis (e suas combinações) enquanto o ser humano demonstra ser capaz de usar todos os níveis. Em síntese, os níveis de PLN podem ser assim sistematizados (Liddy, 2001): a) Fonologia: aborda o estudo dos sons que compõem as palavras a partir de regras fonéticas, fonêmicas e prosódicas; b) Morfologia: estuda os componentes que compõem as palavras, denominados morfemas, assim, uma palavra pode ser analisada nesse nível considerando três tipos de morfemas (prefixo, raiz, sufixo); c) lexical: interpreta o significado das palavras individualmente. Geralmente nesse nível pode haver o uso de um léxico de palavras que se adequa a abordagem adotada pelo sistema PLN, por exemplo, se for um sistema de Análise de Sentimentos, o léxico pode ser de palavras opinativas e, também, pode ter a informação a respeito da polaridade da palavra; d) Sintático: analisa as palavras em uma frase de forma a descobrir a estrutura gramatical da sentença, logo é necessária uma gramática e um analisador para obter como saída uma representação da sentença com as dependências estruturais entre as palavras; e) Semântico: é um dos níveis que contribuem para o entendimento do significado das palavras, mas não o único, todos os níveis auxiliam nesse processo; esse nível de processamento geralmente associa o significado a uma estrutura sintática, tendo como foco as interações entre os significados no nível da palavra na sentença; é útil para eliminar a ambiguidade ao associar palavras a um contexto; f) Discursivo: a análise nesse nível concentra-se nas propriedades do texto como um todo, enquanto os níveis anteriores trabalham com unidades de palavra e/ou sua relação em uma sentença, ou seja, as

diferenças entre as análises é mais uma distinção em termos de granularidade; g) Pragmático: verifica como a interpretação de uma frase pode ser alterada a partir do seu uso em diferentes situações, assim o significado está associado a um ou mais contextos. Esse contexto (cf. Briscoe, 2013) inclui o domínio linguístico e situacional de um enunciado, por exemplo, nem todos os comandos são imperativos, assim “Você pode passar a salada?” em uma análise gramatical pode ser considerada uma pergunta, mas provavelmente será interpretada como um comando ou pedido na maioria das situações. O conhecimento pragmático pode ser usado, também, no preenchimento de informações (elípticas), por exemplo, “O Jalapão aumentou o número de turistas depois de aparecer como cenário em uma novela. Palmas também.”.

Os sistemas que utilizam em seu processo de análise a linguagem natural (que analisam comentários de sites, *posts* em redes sociais, etc.) geralmente usam alguns dos níveis de análise apresentados acima. Na pesquisa desenvolvida neste trabalho, o foco está na análise dos dados relacionados às avaliações produzidas por consumidores turistas no site *TripAdvisor*®, que são compostas por dados semiestruturados (informações sobre o perfil do usuário, marcações relacionadas à escala *Likert* utilizada na plataforma) e dados não estruturados (os comentários redigidos pelos usuários). Para análise dos dados não estruturados, foram utilizados os níveis morfológico, lexical e sintático da PLN, que serão detalhados na seção de desenvolvimento da ferramenta.

Esse cenário representa uma parte do que pode ser considerado o Big Data do setor do turismo, mais especificamente o contexto relacionado ao processo de avaliação do consumo. Ao considerar os quatro Vs apresentados anteriormente, essa pesquisa trabalha com um volume significativo de dados para o tipo de análise proposta, a partir de uma variedade de tipos (desde dados semiestruturados e não estruturados até dicionário de termos relacionados a léxicos de sentimentos) e com análise de valor dos resultados interpretados a partir de conceitos do *Behavioral Perspective Model* (BPM; Foxall, 1990, 1996, 2010), que será apresentada no Estudo 2 dessa tese.

Para que essa análise de valor fosse realizada, anteriormente foi necessária a aplicação da técnica “Análise de Sentimentos”, que usa a PLN como meio para atingir seus objetivos, pois a análise é realizada em textos não estruturados, que precisam ser sistematizados em sentenças, ter suas palavras etiquetadas de acordo com sua

representação na frase, para que as opiniões expressas no texto possam ser identificadas e analisadas como sentimentos positivos ou negativos. A “Análise de Sentimentos”, nesse contexto, é usada para extrair os termos relacionados às características descritas pelos usuários acerca dos destinos turísticos nos componentes avaliados (acomodações, restaurantes e atrações) e, principalmente, para verificar a associação dessas características às palavras opinativas presentes nas sentenças. Com isso, foi possível inferir, por exemplo, a quantidade de características positivas e negativas avaliadas pelos usuários em seus comentários. Essa técnica é apresentada na próxima seção.

Análise de Sentimento (AS)

Segundo Liu (2010), “A *Análise de Sentimento* (AS; do inglês, “*Sentiment Analysis*”) ou *Mineração de Opinião* (MO; do inglês, “*Opinion Mining*”) é o estudo computacional de opiniões, sentimentos e emoções expressas em textos” (p. 629). “Traduzir” computacionalmente sentimentos e emoções não é algo trivial, envolvendo processos analíticos com elevado grau de complexidade, onde se consideram “malhas” de probabilidades de acertos e erros, e não somente a aplicação de critérios de verdade ou falsidade (cf. Liu, 2010).

Para Albornoz, Plaza, Gervás e Díaz (2011), a análise de sentimento, em síntese, pode ser compreendida a partir de três estados sequenciais: 1) a detecção de subjetividade, 2) o reconhecimento da polaridade e 3) a classificação da intensidade da inferência. Na detecção de subjetividade, tem-se como objetivo a descoberta dos termos, expressões ou frases que contém uma opinião (Wiebe, Bruce, & O'hara, 1999; Pang & Lee, 2004; Kim & Hovy, 2004). “Opinião” é entendida na AS como o sentimento do indivíduo acerca de um dado objeto (ou propriedade, atributo de um objeto). Considerando que os sentimentos e percepções sobre algo são produtos sociais, então, de certa forma, como afirma Skinner (1945/1984), é a comunidade humana, essencialmente verbal, que ensina o indivíduo a nomear aquilo que sente, percebe ou pensa. A detecção da subjetividade nesse contexto refere-se, então, às nomeações de alguns sentimentos extraídas do discurso verbal (avaliações) do sujeito. As nomeações verificadas, por sua vez, no âmbito da técnica, consolidam-se em processo caracterizado por confrontações, filtragens e seleções em relação a léxicos que definem o conjunto de emissões de comunidades verbais (Skinner, 1957)

consideradas suas frequências de ocorrência (como mostrado adiante). Nesse sentido, cabe especificar que o “sentimento” detectado pela técnica se configura, antes, como padrão de subjetividade da comunidade verbal tal como expresso pelas nomeações emocionais e opinativas características dessa comunidade, e não necessariamente como subjetividade no nível dos indivíduos que, numa coleta de dados definida, emitiram as nomeações. No reconhecimento da polaridade, o texto é classificado como positivo ou negativo (Pang, Lee, & Vaithyanathan, 2002; Turney, 2002; Esuli & Sebastiani, 2006). Na classificação da intensidade da inferência, tem-se a identificação de n graus de positividade ou negatividade, de modo que uma determinada característica apresentada em um texto pode estar polarizada como fracamente ou fortemente positiva; há, portanto, uma escala para as polaridades (Esuli & Sebastiani, 2006; Brooke, 2009; Wilson, Wiebe, & Hoffman, 2009; Albornoz, Plaza, & Gervás, 2010).

De forma geral, em ambientes virtuais, opiniões podem ser expressas sobre qualquer coisa: um artista, um programa de televisão ou até sobre produtos, destinos turísticos e acontecimentos. Normalmente, as opiniões vêm embutidas em um conjunto de expressões que podem representar os sentimentos e as emoções da pessoa a respeito de objeto(s) ou temática(s). Nesse arranjo, analisar sentimentos é entender que tipo de palavra ou conjunto de palavras pode, em um dado contexto, indicar opinião boa, ruim ou indiferente sobre determinado objeto ou temática (o foco da discussão). Na leitura do comentário apresentado abaixo sobre um hotel:

“[1] Estive em agosto neste hotel. [2] É muito agradável. [3] O atendimento é muito bom. [4] O café da manhã é saboroso. [5] Mas, a piscina estava suja. [6] Minha mãe ficou com raiva de mim, pois viajei sem falar com ela. [7] E ela achou o preço do hotel muito caro.”

O que é possível extrair desse comentário? Primeiro, ele é composto por um conjunto de sentenças e nem todas são opiniões. A sentença [1] expressa um fato. As demais sentenças expressam opiniões. As sentenças [2], [3] e [4] expressam opiniões positivas, e as sentenças [5] e [7] expressam opiniões negativas. A opinião da sentença [2] refere-se ao hotel como um todo. Já as opiniões das sentenças [3], [4] e [5] tem relação com os aspectos “atendimento”, “café da manhã” e “piscina”. A opinião da sentença [7] é sobre o preço do hotel, mas a opinião/emoção da sentença [6] é referente ao “mim” (no caso, a opinião da mãe foi direcionada ao autor do comentário),

não a algo relacionado ao hotel. Em um dado contexto, o usuário pode querer saber sobre determinadas características do objeto, por exemplo, quais hotéis tiveram o atendimento avaliado como positivo. Nem todos os objetos que estão avaliados no comentário interessam. Assim, muito provavelmente, a opinião da mãe sobre o(a) filha(a) não interferirá na opinião sobre o hotel como um todo, mas a opinião dela sobre o preço do hotel é um aspecto que pode ter relevância. Nesse comentário, tem-se, nas sentenças [2], [3], [4] e [5], o autor do comentário como fonte da opinião; já nas sentenças [6] e [7], a fonte da opinião é outra pessoa, citada pelo autor.

Considerando o exemplo acima, tem-se que um objeto (ou temática) pode ser composto por um conjunto de componentes e ter uma série de atributos. Liu (2010, pp. 629 - 631) apresenta as diretrizes analíticas a partir das quais um objeto pode ser formalmente decomposto com base na relação das suas partes:

- **Objeto:** um objeto **O** é uma entidade que pode ser produto, pessoa, evento, organização, tópico, etc. Esse objeto é associado a um par **O: (T, A)**, onde “**T**” é a hierarquia de componentes (partes), subcomponentes, e assim por diante, e o “**A**” é o conjunto de atributos do objeto. Com base nessa definição, um objeto pode ser representado hierarquicamente, como uma árvore, sendo que sua raiz é o próprio objeto. Exemplo: Um determinado hotel é um objeto. O hotel tem um conjunto de componentes (por exemplo, quartos, restaurante, área de piscina), e também um conjunto de atributos (por exemplo, qualidade dos serviços, localização, atendimento). O componente restaurante também pode ter um conjunto de atributos (por exemplo, qualidade da comida, atendimento). Considerando a analogia da árvore, é possível, ainda, horizontalizar a árvore e omitir a hierarquia, e usar as características (ou aspectos) para representar ambos os componentes e atributos. Liu (2010, p. 630) diz que “... *nesta simplificação, o próprio objeto pode também ser visto como uma característica (mas uma característica especial), que é a raiz da árvore original*”. Quando o comentário opinativo é feito sobre o objeto como um todo, isso é chamado de uma opinião geral sobre o objeto (por exemplo, “*O hotel é muito agradável*”). Já quando este é feito sobre uma característica específica, denomina-se um parecer específico sobre uma característica do objeto (por exemplo, “*O café da manhã do hotel é saboroso*”).
- **Trecho de opinião sobre uma característica:** um trecho de texto referente a um parecer sobre uma característica **f** de um objeto **O** avaliado em um

documento d é um grupo de sentenças seguidas que expressam uma opinião positiva ou negativa sobre f . Para Liu (2010, p. 630), “... é possível que uma sequência de sentenças (pelo menos uma) em um documento opinativo expresse uma opinião sobre um objeto ou uma característica do objeto, mas é também possível que uma única frase expresse opinião sobre mais de uma característica (aspecto)”; por exemplo, “O atendimento do hotel é bom, mas o café da manhã é péssimo”.

- **Característica explícita e implícita:** se alguma característica f ou algum de seus sinônimos aparecerem na sentença s , f é chamada de característica explícita. Se nem a característica f e nenhum de seus sinônimos aparecem em s , mas f está implícito, então f é chamada de característica implícita em s . Por exemplo, na sentença “O hotel é legal, mas é tão longe do shopping”, a característica implícita é a localização do hotel.

- **Titular/autor da opinião:** o titular/autor da opinião, que pode ser uma pessoa ou uma empresa que expressa suas opiniões, também é conhecido como fonte de opinião.

- **Opinião:** uma opinião sobre uma característica f é uma atitude, visão, emoção ou avaliação positiva ou negativa sobre uma característica f expressa pelo autor da opinião.

- **Orientação da opinião:** a orientação de uma opinião sobre uma característica f indica o quanto essa opinião é positiva, negativa ou neutra. A orientação da opinião também é conhecida como orientação do sentimento, polaridade da opinião ou orientação semântica.

- **Opinião direta:** é uma opinião expressa na forma de uma quintupla $(o_j, f_{jk}, oo_{ijkl}, h_i, t_l)$, ou seja, é expressa em uma característica/aspecto f_{jk} de um objeto o_j , apresentando uma polaridade h_i , expressa por um titular oo_{ijkl} em um momento específico no tempo t_l . Por exemplo, é o que ocorre no comentário: “O Hotel Pilar tem um atendimento excelente”, emitido por João da Silva em 02/11/2015. Assim, tem-se: a característica/aspecto f_{jk} = atendimento, com a polaridade h_i = positiva, referente ao objeto o_j = Hotel Pilar, expressa pelo autor oo_{ijkl} = João da Silva, na data t_l = 02/11/2015).

- **Opinião comparativa:** quando a opinião expressa uma relação de igualdade ou de diferença entre um ou mais objetos. Comumente, são opiniões expressas usando frases superlativas ou comparativas. Por exemplo: “*O Hotel Pilar é o melhor da região*”; “*O atendimento do Hotel Pilar é melhor que o do Hotel Vitória*”.

Um objeto, para ser computacionalmente descrito, deve ser modelado de forma sistêmica, ou seja, é necessário entender o todo, decompor as partes, ter clareza dos objetivos que se quer alcançar com a decomposição. De certa forma, como esse problema envolve sentido e significado (a semântica), é ainda mais necessário estabelecer contextos, para que o nível de complexidade computacional possa diminuir e tornar a implementação do algoritmo possível. Segundo Chen (2011, pp. 6, 7), a pesquisa sobre a análise de sentimento se concentra em: **1)** como incorporar léxicos para melhorar a precisão da classificação do sentimento, **2)** como identificar automaticamente novos aspectos para a análise de sentimento, e **3)** como identificar pontos de vista das pessoas quando as opiniões dessas não podem ser facilmente julgadas como positivas ou negativas. Enquanto, conforme Chen (2011), o primeiro item tem gerado várias pesquisas sobre técnicas computacionais relacionadas à estrutura da linguagem, o segundo e o terceiro itens levantam questões acerca do sentido das palavras e expressões em um dado universo, de forma que estas possam ser mapeadas em sentimentos expressos nas opiniões das pessoas. Assim, cabe registrar a importância, para dar sentido a aplicações da análise de sentimento, da consideração, em caráter multidisciplinar, aos conhecimentos na intersecção de áreas que trabalham a linguagem, o comportamento e a computação. De outro modo, resultados almejados com aplicações da análise de sentimento não terão embasamento conceitual que os sustentem e nem técnica que os tornem passíveis de modelagem computacional.

A Análise de Sentimento (AS), considerando que tem como base o Processamento de Linguagem Natural (PLN), pode ser realizada em diferentes níveis de granularidade, desde a análise de um documento como um todo, até a análise das sentenças de forma individual ou dos seus aspectos. Esses níveis são explicados a seguir a partir de uma síntese apresentada em Haddi (2015):

- a) **Nível do DOCUMENTO:** para essa análise, considera-se que o documento expressa um sentimento único (cf. Pang, Lee, & Vaithyanathan, 2002; Maas *et al.*, 2011), mesmo que essa suposição

tenha críticas, pois um texto pode conter mais de uma opinião e, portanto, uma análise global apontando um único sentimento poderia ser muito falha (Liu, 2012). Há trabalhos que consideram essa análise satisfatória, pois a indicação da polaridade de um texto, tendo como premissa a maior quantidade de polaridade encontrada em sua análise é suficiente para os seus propósitos. No *TripAdvisor*®, o usuário pode fornecer uma nota geral a avaliação (em uma escala de 'horível' a 'excelente'), que pode corroborar ou refutar uma possível nota geral inferida de um comentário (que é um documento).

- b) **Nível da SENTENÇA:** é uma análise que busca definir a polaridade das sentenças que compõem um texto (cf. Yu & Hatzivassiloglou, 2003, Täckström & McDonald, 2011). Ou seja, é admitido que cada sentença possa expressar um único sentimento. Em Albornoz, Plaza, Gervás e Díaz (2011) é apresentado um método que pode fornecer a orientação de polaridade da sentença a partir de um Vetor de Intensidade de Características (mais detalhes no trabalho relacionado a seguir). Um dos pontos que merece atenção ao fazer uma análise de sentença é a verificação se esta é subjetiva, pois quando a frase apresenta apenas fatos, isso não agrega valor ao classificador, já que não mostra as opiniões do autor da sentença (Kim & Hovy, 2005; Tang, Tan, & Cheng, 2009). Portanto, algoritmos que conseguem classificar uma frase, indicando se nela ocorre a presença de uma opinião, pode melhorar o desempenho dos classificadores do sentimento no nível da sentença, já que evita análises desnecessárias (Pang & Lee, 2004; Yu & Hatzivassiloglou, 2003; Kim & Novy, 2004).
- c) **Nível do ASPECTO:** os sistemas de Análise de Sentimentos baseados em aspecto, no qual se encaixa esse trabalho, recebem como entrada um conjunto de textos (comentários do *TripAdvisor*® que representam avaliações do usuário-turista) relacionados a uma entidade específica (um destino turístico). Assim, esses sistemas tentam detectar os principais aspectos (recursos) da entidade (por exemplo, atendimento, localidade, comida) e sua polaridade (negativa ou positiva). Em síntese, é o processo de extrair os aspectos relevantes na avaliação de um

determinado produto e determinar o sentimento (opinião) relacionado a ele (Popescu & Etzioni, 2005; Chen *et al*, 2013).

Ao considerar o “aspecto” como foco da análise, é necessário fazer as etapas de PLN relacionadas ao pré-processamento, que são as etapas de normalização textual, correção ortográfica, *pos-tagging* e análise de dependência sintática. Desta forma, o uso prático das etapas relacionadas aos níveis morfológicos, léxicos e sintáticos serão apresentados na seção que trata do desenvolvimento da ferramenta. Os tópicos a seguir apresentam estudos de caso que utilizam a Análise de Sentimentos na avaliação de produtos, bem como os métodos que podem ser utilizados para efetivar tal análise.

Estudos empíricos sobre o uso da Análise de Sentimento (AS) em análises de opiniões sobre produtos

Trabalho de mineração de opinião sobre características de produtos e sua polaridade (positiva ou negativa) foi desenvolvido por Hu e Liu (2004), que usaram *reviews* de cinco produtos (duas câmeras digitais, um DVD *player*, um *mp3* e um telefone celular) postados nos *sites* de duas lojas virtuais (*Amazon.com*⁵ e *Cnet.com*⁶). Hu e Liu (2004) resumiram as críticas dos clientes sobre os produtos e, com a abordagem, desenvolveram estudo diferente da sumarização tradicional de textos, visto que foram consideradas as características específicas dos produtos presentes nos comentários, bem como as opiniões dos clientes em relação a essas características. Cada comentário tinha um título e a revisão do autor. As informações adicionais sobre data, hora, nome do autor, localização e avaliações do comentário como um todo (a partir da atribuição de estrelas) não foram consideradas. Para cada produto, foram coletados e analisados 100 comentários. Em seguida, processo de limpeza dos documentos foi realizado, quando foram extraídas as *tags* HTML e usado o *NLProcessor*⁷ para gerar o *POS-Tagging*, que resulta na categorização das palavras pela classe gramatical a que pertencem. Com isso, foram extraídas as

⁵ <http://www.amazon.com>

⁶ <http://www.cnet.com/>

⁷ *NLProcessor – Text Analysis Toolkit*, 2000. <http://www.infogistics.com/textanalysis.html>

características presentes nos comentários e, por fim, determinada a polaridade de cada uma delas.

Hu e Liu (2004) também procuraram encontrar as características que mais interessavam às pessoas dentro do universo pesquisado. Para isso, usaram o algoritmo *Apriori* (Agrawal & Srikant, 1994) para verificar, por *reviews*, a frequência de itens associados presentes na base de características. Nesse contexto, um conjunto de itens é um conjunto de palavras ou uma frase que ocorre em um *review*. O algoritmo *Apriori* funciona em duas etapas: **1ª)** encontra todos os conjuntos de itens frequentes, e **2ª)** gera as regras dos conjuntos de itens frequentes descobertos. No estudo de Hu e Liu (2004), apenas o primeiro passo foi realizado, ou seja, encontrar os conjuntos de itens mais frequentes, ou, mais especificamente, as características candidatas a ser as preferidas dos autores dos *reviews* sobre os produtos. Após essa classificação, foi realizado um processo de “poda”, eliminando as características não relevantes ao processo ou redundantes. Somente após esse processo, foi feita a extração das palavras opinativas e, por último, determinada a polaridade. Para tanto, os autores usaram a técnica *Bootstrapping* e a base de informações léxicas *WordNet* (Miller *et al.*, 1990). As técnicas estatísticas *Recall* e *Precision* foram aplicadas na verificação do desempenho dos algoritmos usados, que se mostraram eficazes.

Em Albornoz, Plaza, Gervás e Díaz (2011), foram analisados 1.000 comentários sobre hotéis, obtidos no *site* de viagens *Booking.com*⁸, para desenvolvimento de um sistema que identifica e quantifica as características importantes para os consumidores ao avaliar um determinado produto ou serviço. Assim, depois que foi verificada a polaridade e a força da opinião expressa em cada frase, foi calculada uma pontuação para cada característica, tendo como base a polaridade das frases associadas a ela. Os valores que quantificam as opiniões foram inseridos em um Vetor de Intensidade de Características (VFI; do inglês, *Vector of Feature Intensities*). O VFI é o dado de entrada para o algoritmo de aprendizagem de máquina que prevê uma classificação para a revisão. Considerada a pertinência do estudo de Albornoz *et al.* (2011) para o presente estudo, cabe aqui detalhamento um pouco maior do mesmo.

Albornoz *et al.* (2011) coletaram 25 comentários por hotel, em um total de 60 hotéis (1500 avaliações) no *site Booking.com*. Cada avaliação continha as seguintes

⁸ <http://www.booking.com/>

informações: **1**) cidade (do hotel), nacionalidade (do revisor), a data (da avaliação) e o tipo de usuário (em um conjunto de sete categorias, como viajante solitário, grupo, etc.); **2**) uma pontuação entre 0 e 10, calculada automaticamente pelo sistema do *booking.com*, a partir das taxas atribuídas em 5 aspectos (funcionários, serviços/instalações, limpeza do quarto, preço e localização), usada no *site* para descrever a opinião geral do revisor, e, por último, **3**) a revisão em texto livre, separadamente, do que o usuário havia gostado e não gostado durante sua estadia. Albornoz *et al.* concluíram, nas primeiras análises, que a pontuação global atribuída a uma avaliação não tinha, geralmente, relação com tudo o que estava descrito no texto opinativo do usuário. Assim, duas revisões com quase a mesma pontuação poderiam refletir opiniões bem diversas. Para atenuar esse problema, foi usada a técnica de anotação manual para cada comentário a partir de dois conjuntos de categorias: **1**) [excelente, bom, regular, ruim, muito ruim] e **2**) [bom, regular, ruim]. Algumas revisões contraditórias foram removidas e, depois, selecionadas aleatoriamente 1.000 *reviews*.

Na etapa de quantificação das opiniões (o cerne do estudo), Albornoz *et al.* (2011), usaram o sistema de reconhecimento de polaridade apresentado por Albornoz, Plaza e Gervás (2010), que em síntese, tal como descrito em Albornoz *et al.* (2011):

“... tem como objetivo a extração em cada frase dos conceitos *WordNet* que implicam em um significado emocional, atribuindo-lhes uma emoção a partir de um conjunto de categorias disponível em um léxico afetivo e, depois, utilizar essas informações como entrada em um modelo de regressão logística para prever a polaridade da sentença e a probabilidade da polaridade” (Albornoz *et al.*, 2011, p. 60).

Albornoz *et al.* (2011) experimentaram duas estratégias para atribuição de valores no Vetor de Intensidade de Características (VFI). A primeira é a *Binary Polarity*, em que, para cada frase do comentário, o valor atribuído numa dada posição do vetor é incrementado ou decrementado por 1 a partir do valor predito para a sentença (positivo ou negativo). A segunda técnica é a *Probability of Polarity*, em que a característica tem sua probabilidade aumentada ou diminuída a partir da polaridade atribuída à frase pelo classificador de probabilidade. A partir do uso do VFI como entrada, os autores testaram três classificadores *Weka*⁹ (*Logistic*, *LibSVM*, *FT*) para três classes inicialmente (bom, regular, ruim), utilizando três conjuntos de

⁹ *Weka* é uma coleção de algoritmos de aprendizagem de máquina e *data mining*, disponível em <http://www.cs.waikato.ac.nz/ml/weka/>

características. O melhor desempenho (71,7%) foi obtido no classificador *Logistic*. Outros experimentos foram realizados e os resultados foram favoráveis (ao *Logistic*) em comparação com os obtidos em outros métodos similares, como Pang *et al.* (2002) e Albornoz *et al.* (2010).

Para a detecção da opinião (cf. Wiebe, Bruce, & O'hara, 1999; Pang & Lee, 2004; Kim & Hovy, 2004), a compreensão semântica de termos (cf. Chen, 2011) e a identificação dos objetos alvo do processo de AS (cf. LIU, 2010), os trabalhos apresentados em Hu e Liu (2004) e Albornoz *et al.* (2011) usaram algumas técnicas de pré-processamento de texto e AS para a extração das opiniões expressas nas avaliações (do inglês “*reviews*”) dos produtos e nos comentários de consumidores turistas e utilizaram dicionários de léxicos de sentimentos para o reconhecimento da polaridade dos termos. Contudo, apenas o segundo trabalho apresentou as etapas usadas para a classificação da intensidade da polaridade, que relacionou a polaridade da sentença com a de cada aspecto. Já a identificação de novos aspectos para análise de sentimentos tem relação com a possibilidade de definição de dicionários de termos, não apenas de léxicos de sentimentos, mas de características relacionadas a determinados contextos, por exemplo, as características mais avaliadas no setor do turismo, mas esse fator não foi apresentado nos trabalhos relacionados.

No trabalho em questão, algumas das etapas citadas nesses estudos de caso e os elementos de AS apresentados nas definições do conceito foram integradas na concepção do modelo de AS aplicado, por exemplo, etapas técnicas mais gerais relacionadas ao pré-processamento de textos dos comentários e o uso de dicionários de léxicos de sentimentos. Também foi feita a detecção da opinião a partir dos aspectos e suas polaridades extraídos nas sentenças que compõem os comentários. No entanto, a etapa relacionada à verificação da intensidade da polaridade não foi considerada, pois o foco desse trabalho é o aspecto e sua polaridade (negativa ou positiva).

De uma forma geral, dois pontos foram evidenciados nesse estudo, a questão do grande volume de dados disponíveis na internet que podem ser usados para análises e prospecções nas mais diversas áreas, especialmente na área do turismo, considerando a quantidade de comentários sobre avaliações de produtos turísticos que há em sites de grandes organizações desse setor. As técnicas computacionais que podem ser utilizadas quando a perspectiva é gerar informações relevantes para o entendimento das opiniões dos turistas, considerando os textos escritos por esses

nos comentários. Para tanto, foi explorada a técnica de Análise de Sentimentos, que para ser realizada necessita, anteriormente, que se realize etapa de tratamento de texto relacionado aos métodos da PLN.

Essas constatações evidenciam o que Benckendorff, Sheldon e Fesenmaier (2014) apontaram em seus estudos, que a Tecnologia da Informação (TI) tem sido amplamente utilizada na indústria do turismo desde o advento da internet. Assim, ao considerar o *Big Data* dessa indústria, tem-se, além dos dados que já constam nas bases de dados das empresas, os dados presentes em *sites* especializados em viagens turísticas (*Booking.com*, *Tripadvisor.com*, etc.), nas mídias sociais, na telefonia móvel, além de outras fontes de dados, como aqueles do transporte público, de câmeras urbanas, de cartões de crédito, etc. Segundo Heerschap, Ortega, Priem e Offermans (2014), o *Big Data* afetará o cenário da produção das estatísticas sobre a atividade turística, e isso implicará na mudança da forma como essas estatísticas são produzidas e, especialmente, na criação de meios para agregar dados que não são coletados, nem analisados, da forma tradicional. Esses dados, no presente estudo, têm relação com o processo de pós-compra, mais especificamente na avaliação de produtos turísticos. Nesse contexto, o uso e desenvolvimento de algumas técnicas/ferramentas computacionais viabilizaram as análises pretendidas para o Estudo 2, e, portanto, definiram o objetivo do presente Estudo, descrito a seguir.

OBJETIVO

Considerado o conjunto de conhecimentos acima descritos, referentes às possibilidades decorrentes do uso de técnicas computacionais de mineração de dados e Análise de Sentimento para encontrar padrões de informação representativos de comportamento de consumo humano, o presente estudo teve como objetivo:

Conceber, desenvolver e aplicar sistema computacional informatizado voltado à: **a)** extração de informações verbais (comentários) de turistas-consumidores sobre destinos turísticos localizados na plataforma *TripAdvisor®*, dedicada ao gerenciamento de informações e *e-commerce* turísticos; **b)** prospecção do conteúdo dos comentários extraídos com técnica computacional baseada em algoritmo de análise de sentimento para a quantificação de “sentimentos positivos” e “sentimentos negativos” sobre as acomodações (ACO; infraestrutura para hospedagem), restaurantes (RES; infraestrutura para

alimentação) e atrações (ATR; infraestrutura de atrativos) dos destinos avaliados; **c)** extração do número de indicações avaliativas dos componentes do produto turístico (ACO, RES e ATR) feitas por turistas-consumidores com diferentes *status* como *TripColaboradores*, com a escala de satisfação usada pelo *TripAdvisor*® (*Likert* com cinco [5] níveis: “*Horrível*”, “*Ruim*”, “*Razoável*”, “*Muito bom*” e “*Excelente*”) para cada destino, e **d)** extração do número de votos úteis (“*likes*” ↩) feitas por turistas-consumidores para os comentários sobre os componentes do produto turístico (ACO, RES e ATR) para cada destino.

MÉTODO

Ambiente da coleta dos dados

O ambiente da coleta de dados foi o *website TripAdvisor*® (www.tripadvisor.com.br), fundado em fevereiro de 2000 e sediado em Needham, Massachusetts (EUA). O *TripAdvisor*® possui mais de 32 milhões de membros cadastrados, fluxo de 375 milhões de visitantes únicos mensais (segundo o *Google Analytics*®¹⁰) e abriga mais de 140 milhões de comentários sobre componentes de produtos turísticos (sobretudo, hospedagem, alimentação e atrativos). O *TripAdvisor*®, cujo posicionamento é mundial, é dedicado à propagação de informações turísticas e mediação de negócios no setor turístico. O total de 6.438.497 comentários distribuídos entre os 100 destinos turísticos brasileiros mais avaliados entre início de fevereiro e final de março de 2017, definiram a base de dados primários do estudo, onde incidiram as técnicas de desenvolvimento do sistema.

Materiais e equipamentos

Dois computadores com 16GB de memória RAM, Processador Intel Core^(TM) i7-4790S CPU 3.20GHz, notebook pessoal (Intel(R) Core(TM) i7-3537U CPU @ 2.00 GHz, Memória RAM 16 GB, Sistema Operacional de 64 bits) foram os equipamentos

¹⁰ Fonte: *Google Analytics*®; média mensal de usuários únicos, 2º trimestre de 2015.

utilizados. Os materiais e ferramentas criadas ou adotadas no desenvolvimento do sistema foram:

1) Robôs *WebCrawler* foram criados usando ferramenta desenvolvida por Araújo (2017), com aplicação específica para coleta de dados em páginas HTML do *TripAdvisor*[®]. **2)** Os dados resultantes das extrações foram armazenados em banco de dados *Microsoft*[®] *SQL Server*[®] 2014¹¹. **3)** Para processamento de linguagem natural (no caso, o português) e viabilizar tratamento de elementos textuais avaliativos dos componentes dos produtos turísticos, foi utilizada a biblioteca *NLTK*^{®12} (*Natural Language ToolKit*), que é plataforma para Processamento de Linguagem Natural (PLN) vinculada à linguagem de programação *Python*^{®13}. Adicionalmente, utilizou-se a) o *SentiLex – PT*, v2.0¹⁴ (Silva, Carvalho, & Sarmiento, 2012), que consiste em léxico de palavras do português descritivas de sentimentos, composto de lemas e formas flexionadas, com indicativo da polaridade (negativa, neutra ou positiva) de cada item lexical; b) o Dicionário LIWC, em português do Brasil (Balage-Filho, Pardo, & Aluísio, 2013), que é um dos léxicos de sentimentos do *software Linguistic Inquiry and Word Count (LIWC)*¹⁵; c) o *OpLexicon* v3.0¹⁶ (Souza, Vieira, Buseti, & Chishman, 2012), que é um léxico em português que possui 4 categorias de palavras: verbos, adjetivos, *hashtag* e *emoticons*; e, ainda, d) o sistema para análise de dependência baseada em dados, o analisador sintático *MaltParser*¹⁷ (Nivre, Hall, & Nilsson, 2006), para o reconhecimento das relações sintáticas entre as palavras de uma sentença e o *corpus* anotado sintaticamente com, aproximadamente, 100 idiomas, e e) o *Universal Dependencies (UD)*¹⁸, para a realização de treinamentos e testes do processamento.

¹¹ <http://www.microsoft.com/SQL>

¹² <http://www.nltk.org/>

¹³ <http://www.python.org/>

¹⁴ http://dmir.inesc-id.pt/project/SentiLex-PT_02

¹⁵ <http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>

¹⁶ <http://ontolp.inf.pucrs.br/Recursos/downloads-OpLexicon.php>

¹⁷ <http://www.maltparser.org>

¹⁸ <http://universaldependencies.org/>

Procedimentos

A criação do sistema baseado em Análise de Sentimento (AS) derivou da condução de método de desenvolvimento computacional que envolveu materiais, ferramentas e raciocínios de implementação adotados ou desenvolvidos em várias etapas. Para o processo de implementação e análise de desempenho da ferramenta denominada *SentimentALL* concebida para a tese, alguns artigos, trabalhos de estágio, de conclusão de curso e iniciação científica foram desenvolvidos e orientados pela presente autora¹⁹. As etapas do desenvolvimento são apresentadas a seguir:

a) Extração dos dados

O processo para extração original de informações do *TripAdvisor*® utilizou *framework* para *Python* denominado *Scrapy*, implementado em (Araújo, 2017), que permite a criação de programas (*spiders*) que navegam por páginas *web* e delas extraem dados. Para o desenvolvimento do processo de **extração dos dados**, foram consideradas três categorias (tecnicamente, “objetos”) presentes no *TripAdvisor*®: Acomodações (ACO), Restaurantes (RES) e Atrações (ATR). A coleta de dados foi sistematizada em duas etapas, e para cada uma das etapas foram utilizados três *spiders*, um para cada objeto. Na primeira etapa, os *spiders* executavam a coleta da quantidade de avaliações de cada objeto (ACO, RES e ATR) em todas as cidades do Brasil com aparição no *TripAdvisor*®. Na segunda etapa, foi feita a extração dos dados das 100 cidades com maior quantidade de comentários e gerado um arquivo *Comma Separated Values* (CSV) para cada coleta realizada. Os arquivos CSV resultantes

¹⁹ Araújo, L.G. A. (2017). *SENTIMENTALL VERSÃO 2: Desenvolvimento de Análise de Sentimentos em Python* (Trabalho de conclusão de curso). Centro Universitário Luterano de Palmas - CEULP/ULBRA, Palmas, TO, Brasil.

Araújo, L. G. A., Marinho, D. S., & Brito, P. F. (2017). Otimizando Maltparser para Análise de Dependências na Língua Portuguesa. In *XVII Jornada de Iniciação Científica do CEULP/ULBRA*, Palmas-TO.

Leal, M.R., Oliveira, W. C. C., & Brito, P. F. (2017). Avaliação de Desempenho de uma Ferramenta de Análise de Sentimentos baseada em Aspectos. In *Congresso de Computação e Tecnologias da Informação*, 19., Palmas, TO.

Oliveira, W. C. C. (2015). *SentimentALL: Módulo para Análise de Sentimentos em Português*. 86 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Centro Universitário Luterano de Palmas, Palmas.

Oliveira, W. C. C., & Brito, P. F. (2015). Utilização do Pointwise Mutual Information na Identificação de Expressões Multipalavras. In *XVII ENCOINFORM – Congresso de Computação e Sistemas de Informação*, 17, Palmas, 25-34.

Souza, F. R. M. (2017). *Implementação de um Dashboard para a ferramenta SentimentALL* (Trabalho de Conclusão de Curso). Centro Universitário Luterano de Palmas - CEULP/ULBRA, Palmas, TO, Brasil.

foram armazenados na base de dados e continham os dados relevantes para a prospecção computacional (por exemplo, comentários e avaliações dos destinos turísticos, informações sobre os autores dos comentários, etc.) relacionados a cada um dos objetos (ACO, RES e ATR).

b) **Pré-processamento textual**

A base lexical utilizada foi a biblioteca NLTK® do Python. A partir do uso do módulo de pré-processamento implementado em Araújo (2017), foi realizado tratamento do texto de cada comentário, procedimento esse composto por cinco etapas: **1)** normalização, **2)** correção ortográfica, **3)** *POS-Tagging*, **4)** identificação de expressões multipalavras e **5)** análise de dependência sintática. Essas etapas são apresentadas detalhadamente a seguir:

Etapa 1: Normalização textual. Para a normalização textual foram utilizadas expressões regulares apresentadas em Oliveira (2015). O critério de substituição dos *emoticons* por palavra da língua portuguesa considerou a lista apresentada em Araújo, Gonçalves e Benevenuto (2013).

Etapa 2: Correção ortográfica. Após a normalização do texto, foi utilizado um algoritmo para a execução do processo de correção da grafia das palavras. Baseado em Norvig (2016), o algoritmo foi aplicado com adequações relacionadas à estrutura morfossintática da língua portuguesa. O algoritmo selecionado funciona adequadamente caso utilize *corpus* com lista de palavras conhecidas e suas respectivas frequências. Como em Norvig (2016) o *corpus* é composto por palavras da língua inglesa, foi necessária a utilização de dois *spiders*, disponíveis em Araújo (2017), para a coleta de notícias nos sites da BBC Brasil²⁰ e TNH1²¹. Além do uso do *corpus*, uma lista com quase um milhão de palavras encontradas no *site* do Project Natura²² foi utilizada para aumento significativo no conjunto de palavras conhecidas.

Em linhas gerais, o algoritmo executa o seguinte processo: recebe uma palavra (entrada), se a palavra estiver na lista de palavras conhecidas, então não precisa ser corrigida. Caso não esteja, transformações podem ser realizadas na palavra de forma a obter um conjunto de palavras candidatas. A palavra candidata que tiver a maior frequência é a palavra resultante no processo de correção (saída). Mas, se o algoritmo não gerar nenhuma palavra candidata conhecida, a palavra permanece na sua versão

²⁰ www.bbc.com/portuguese

²¹ www.tnh1.com.br

²² <http://natura.di.uminho.pt/download/sources/Dictionaries/wordlists/LATEST/>

original. A Figura 1, adaptação de Norvig (2016) feita por Araújo (2017) e desse extraída, ilustra síntese do processo na forma de fluxograma.

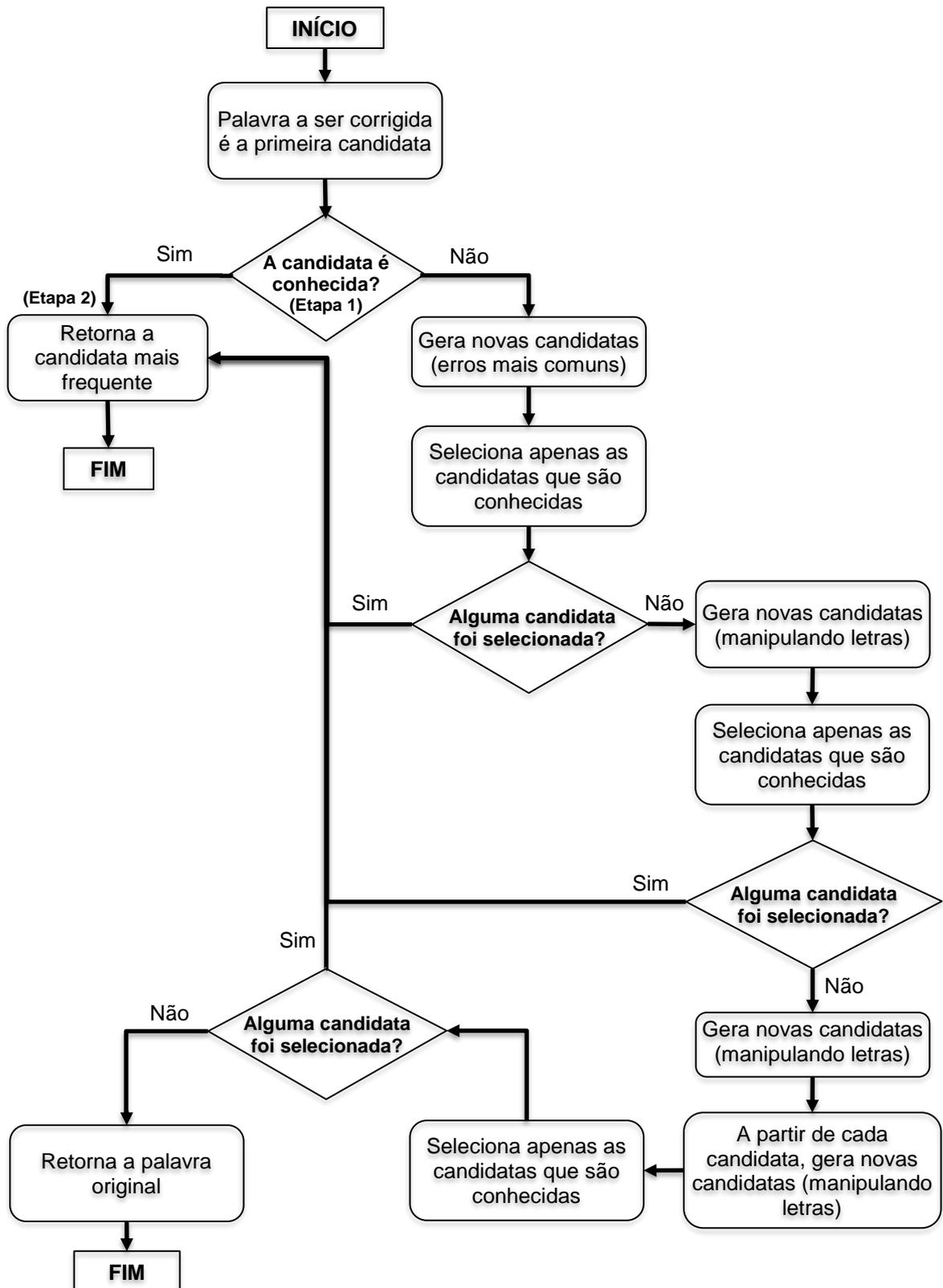


Figura 1: Fluxograma do algoritmo de correção ortográfica. Adaptado de Norvig (2016) por Araújo (2017) e desse extraído

Seguindo o fluxograma apresentado na Figura 1, tem-se que o processo pode ser realizado a partir de três transformações: a primeira transformação, adicionada ao algoritmo de Norvig (2016) e apresentada com detalhes em Araújo (2017), está relacionada aos erros ortográficos mais comumente encontrados, referentes às semelhanças grafêmicas ou a falta de acentuação. Por exemplo, seguindo a execução do algoritmo, a palavra “cafe”, que é incorreta, gera novas palavras candidatas a partir da verificação dos erros mais comuns, que considera alguns tipos de agrupamentos (por exemplo, “eêê”, que gera as palavras “cafe”, café e “cafê”). A partir disso, são selecionadas apenas as palavras candidatas conhecidas (Etapa 1) e verificado se algumas palavras candidatas foram selecionadas. Caso sim, é executada a etapa que retorna a palavra candidata mais frequente (Etapa 2). As outras transformações têm relação com manipulação de letras a partir do processo de eliminação, adição, transposição e substituição.

Etapa 3: POS-Tagging. Para a execução dessa etapa, utilizou-se o *corpus* Mac-Morpho, disponibilizado pela biblioteca NLTK®, *corpus* esse que possui mais de um milhão de palavras em idioma português classificadas de acordo as classes gramaticais do idioma. O Mac-Morpho é fornecido aos classificadores (denominados *taggers*), disponíveis no módulo *POS-Tagging* da NLTK, como conjunto de treino para que o módulo seja capaz de classificar corretamente os termos dos comentários. Para a sua utilização, é necessário passar, como entrada, o conjunto de *tokens* que será classificado e a base de dados de referência (nesse caso, o *corpus* Mac-Morpho).

Etapa 4: Identificação de expressões multipalavras. Geralmente, as palavras opinativas são adjetivos ou substantivos adjetivados, e os aspectos são substantivos ou expressões multipalavras, ou seja, palavras que separadas perdem o sentido dentro de um determinado contexto (por exemplo, “café da manhã”, “cara feia” [quando representando mau humor, impaciência]). A “Identificação de Expressões Multipalavras” é realizada em duas etapas. A primeira etapa busca encontrar possíveis expressões candidatas a ser expressões multipalavras, e a segunda tem como objetivo filtrar as expressões a partir da utilização de uma fórmula apresentada em Church e Hanks (1990) e generalizada por Cruys (2011). O primeiro passo faz a busca nos dados procurando por padrões morfológicos que identificam expressões candidatas. No estudo, o algoritmo implementado em Oliveira e Brito (2015) e em Araújo (2017) considerou os padrões morfológicos apresentados em Justeson e Katz (1995) para multipalavras, a partir da junção de termos substantivos (N, do inglês

nouns), adjetivos (A, do inglês *adjectives*) e preposições (P, do inglês *prepositions*), conforme pode ser visualizado na Tabela 1, a partir de sistematização proposta em Boos, Prestes e Villavicencio (2014):

Tabela 1: Padrões morfológicos das expressões multipalavras.

PADRÃO MORFOLÓGICO	EXEMPLO
NA	Governo Federal
NN	Nações Unidas
NNA	Fundo Monetário Internacional
NAA	Produto Interno Bruto
NPN	Bolsa de Valores

Fonte: Boos, Prestes e Villavicencio (2014)

No trabalho de Araújo (2017), foi realizada pequena mudança no algoritmo proposto por Oliveira e Brito (2015), relacionada com a quantidade de *tokens* considerada nas expressões analisadas. No caso, foi modificado de um máximo de quatro *tokens* para apenas três, pois as contrações (por exemplo, “das” – “de” + “as”) não foram desmembradas.

A primeira etapa do algoritmo apresentado na Figura 2 é iniciada a partir da busca por um substantivo nos *tokens* que compõem as sentenças. Quando o encontra, verifica se esse termo e mais dois *tokens* a frente formam alguns dos padrões morfológicos (NNA, NAA ou NPN da Tabela 1). Caso um padrão seja reconhecido, os três *tokens* passam a compor um conjunto de expressões candidatas. Se os três termos não formarem um dos padrões, o algoritmo verifica se o próximo *token* é um substantivo ou adjetivo e, se for, esses *tokens* comporão uma expressão candidata, pois se enquadram no modelo NA ou NN da Tabela 1. O processo se repete até que todas expressões candidatas sejam encontradas.

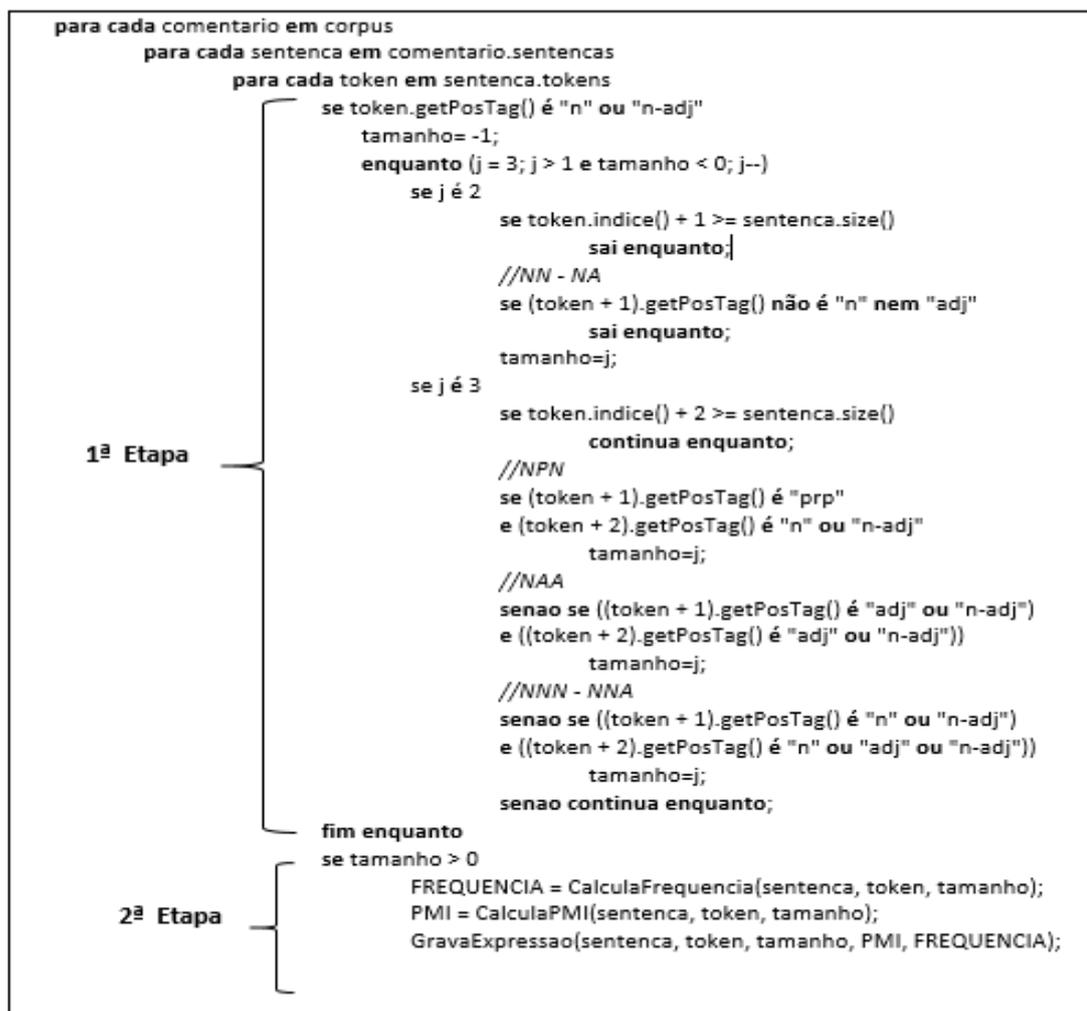


Figura 2: Pseudocódigo da pesquisa por expressões candidatas. Extraído de Araújo (2017) a partir de modificação do algoritmo de Oliveira e Brito (2015).

No tratamento das multipalavras foi utilizado, adicionalmente, o método de Informação Mútua Ponto-a-Ponto (PMI; do inglês *Pointwise Mutual Information*), apresentado em Church e Hanks (1990). O PMI compara a probabilidade de observação das palavras x e y juntas (probabilidade de ocorrência conjunta) com as probabilidades de ocorrência independente de x e y . Se existe uma verdadeira relação entre x e y , então a probabilidade de ocorrência conjunta, $P(x, y)$, será muito maior do que a probabilidade de ocorrência independente, $P(x)$ e $P(y)$. Assim, $I(x, y) > 0$ (Church & Hanks, 1990, p. 23).

Assim, dada duas palavras x e y , o PMI busca determinar a co-ocorrência dessas palavras em determinado domínio. Logo, a fórmula é a função da probabilidade de ocorrência em conjunto das duas palavras (x e y), em relação à

probabilidade de ocorrência individual de cada palavra no domínio, conforme Equação 1:

Equação 1:

$$I(x, y) = \log_2 \frac{P(x, y)}{P(x) P(y)}$$

Vista a possibilidade de ocorrência de multipalavras com mais de duas palavras (*tokens*) em relatos tais como os analisados no presente estudo, foi usada a generalização da medida PMI proposta em Cruys (2011; Equação 2):

Equação 2:

$$SI_{(x_1, x_2, \dots, x_n)} = \log_2 \frac{P(x_1, x_2, \dots, x_n)}{\prod_{i=1}^n P(x_i)}$$

Onde:

$P(x_1, x_2, \dots, x_n)$ representa a probabilidade de ocorrência da expressão em sua totalidade,

e

$\prod_{i=1}^n P(x_i)$ representa a multiplicação das probabilidades de ocorrência de cada uma das palavras (*tokens*) componentes da expressão.

Na base analisada, foram encontradas por volta de três milhões de expressões candidatas. Utilizando as expressões candidatas, iniciava-se a segunda etapa (Figura 2), que consistiu na filtragem das expressões. A filtragem é realizada tendo como base um intervalo de valor do PMI. Para a determinação desse intervalo, foram utilizados os testes feitos em Araújo (2017), que utilizou diversos valores de intervalos em um conjunto de cinquenta frases. Nesse conjunto, sessenta e quatro expressões multipalavras foram destacadas. Na definição do melhor intervalo, foram utilizadas as medidas para avaliação de classificadores *Precision* e *Recall*, além da *F-Measure*, que faz uma média harmônica entre os valores encontrados nessas medidas. Os dados relacionados às frases analisadas foram organizados em uma “Matriz de Confusão”, que consiste em uma tabela de cruzamento de informações que apresenta visualmente o número de acertos (Verdadeiros-Positivos ou V_p), erros (Falsos-

Positivos ou F_p), perdas (Falsos-Negativos ou F_n) e rejeições corretas (Verdadeiros-Negativos ou V_n) (Baker, 2008). As equações para obtenção das medidas são as seguintes:

Equação 3:

$$Precision = \frac{V_p}{V_p + F_p}$$

Equação 4:

$$Recall = \frac{V_p}{V_p + F_n}$$

e,

Equação 5:

$$FMeasure = \frac{2 * Precision * Recall}{Precision + Recall}$$

onde V_p é a quantidade de expressões reconhecidas que estão corretas; F_p é quantidade de expressões reconhecidas que estão incorretas, e F_n é a quantidade de expressões que deveriam ter sido reconhecidas, mas não foram.

O teste foi feito variando o valor de x de 0 a 9, com intervalo de 1, da seguinte equação: $x < PMI$. Com isso, são consideradas expressões multipalavras as expressões com PMI maior do que x . A Figura 3 mostra o gráfico com os valores de *Precision*, *Recall* e *F-Measure*, para cada valor de x .

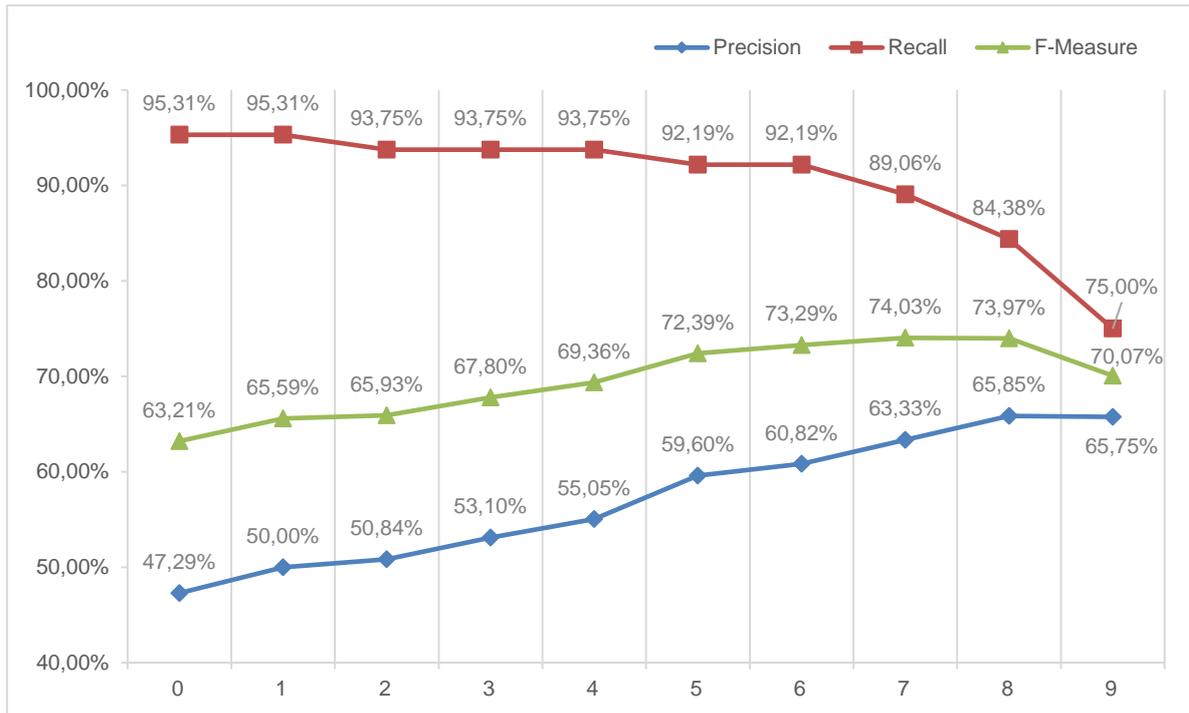


Figura 3: Resultado do teste do PMI, conforme conduzido em Araújo (2017).

Como pode ser observado na Figura 3, o valor de x que obteve a melhor porcentagem de *F-Measure* é 7 (74,03%). Com a eliminação de expressões com pouca frequência (menor do que 25), pode-se verificar que o resultado melhorava. A *Precision* alcançou 73,33%, a *Recall* passou a 85,94% e a *F-Measure* a 79,14%.

Etapa 5: Análise de dependência sintática. Para a realização desse processo, foi usado o sistema *MaltParser* que, a partir de modelo definido em seu algoritmo, verifica e define os tipos de dependências das palavras de uma sentença (Nivre, Hall, & Nilsson, 2006). Para tanto, esse classificador utiliza o *Universal Dependencies* (UD), um *corpus* em diversas línguas (inclusive o idioma Português) com anotações que identificam as relações de dependência entre termos (Tabela 2).

Tabela 2: Relações de dependência no Universal Dependencies (UD).

DEPENDÊNCIAS	DESCRIÇÃO
advmod	Modificador do advérbio
amod	Modificador do adjetivo
cc	Conjunção coordenativa
conj	Palavras conectadas por conjunção

det	Determinante
nmod	Modificador nominal
nsubj	Sujeito nominal
xcomp:adj	Complemento

Na utilização do analisador sintático *MaltParser* para a análise de dependência, o treinamento realizado utilizou os dados de treinamento do UD, que contém 9.262 sentenças. Como o *MaltParser* utiliza nove modelos de algoritmos determinísticos de análise sintática, foi utilizado um teste feito em Araújo (2017) para identificar qual algoritmo teria resultado mais eficiente no contexto do presente estudo. Para tanto, realizou-se a comparação entre o arquivo de teste e o de saída, com o objetivo de avaliar o quão bem os algoritmos estavam definindo as relações sintáticas entre os termos. Assim, foi comparado, para cada termo presente nas sentenças do arquivo de teste, o seu equivalente no arquivo de saída, de forma a verificar se as relações (e seus tipos) identificadas pelos algoritmos estavam coerentes. Nos testes, segundo Araújo (2017), o algoritmo que obteve o melhor resultado foi o *Stack swap-lazy*, com acerto de aproximadamente 84% na relação entre termos e, aproximadamente, 88% de acerto no tipo de relação. Outros testes foram realizados em Araújo, Marinho e Brito (2017) que também apontaram o *Stack swap-lazy* como o mais adequado para uso no estudo.

c) **Desenvolvimento do módulo de Análise de Sentimento (AS)**

Foram duas versões da ferramenta *SentimentALL* concebida para essa tese, a primeira foi implementada em Oliveira (2015) e a segunda implementada em Araújo (2017), onde está o módulo de AS que gerou os resultados usados neste trabalho. Na Figura 4, o fluxograma definido em Araújo (2017) é apresentado com a representação gráfica do processo de verificação das palavras opinativas presentes nas sentenças resultantes da análise de dependência da etapa anterior, bem como a sua polaridade.

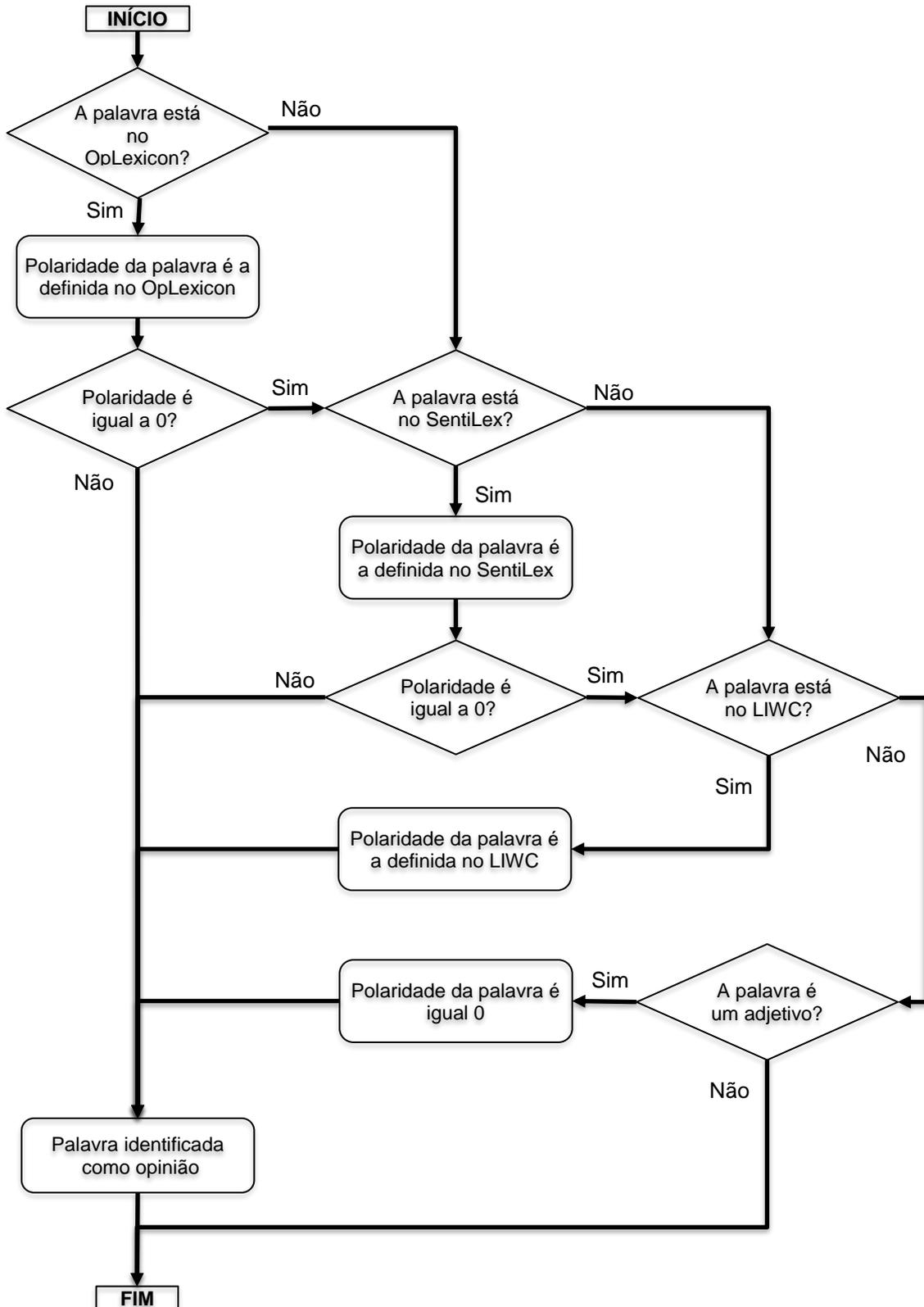


Figura 4: Fluxograma para identificação da polaridade de uma palavra, extraído de Araújo (2017).

Em síntese, o fluxograma apresentado na Figura 4 indica que a verificação de cada palavra da sentença era feita considerando duas situações: se a palavra estava

presente em um dos léxicos de sentimento (*OpLexicon*, *SentiLex* e *LIWC*) ou, caso contrário, se ainda assim a palavra era um adjetivo (identificada como tal no processo anterior [análise de dependência]). Desta forma, caso a palavra estivesse em um dos léxicos, ela assumia a sua polaridade. Mas, se a polaridade fosse neutra, então era verificado um próximo léxico até que polaridade diferente da neutra fosse encontrada, ou até que fosse feita a verificação em todos os léxicos. Se a palavra não estivesse em nenhum dos léxicos e fosse um adjetivo, então assumia a polaridade neutra. Ao final, a palavra podia assumir um de dois estados: ser uma opinião ou não ser uma opinião.

A próxima etapa de AS realizada foi a identificação dos aspectos. A identificação é feita a partir de uma série de percursos implementados em Araújo (2017). Os percursos têm como objetivo encontrar um aspecto, uma palavra etiquetada com a *tag* "N" (do inglês, *Noun*), que está, geralmente, próxima a uma palavra identificada como opinião, que pode ser uma das três *tags*: adjetivo (ADJ), advérbio (ADV) ou verbo (V). Com isso, um conjunto de caminhos foi criado para cada *tag*. Os percursos foram definidos a partir da análise de frases presentes na base de dados (sentenças que compõem os comentários).

d) **Totalização das variáveis primárias do estudo**

Uma série de consultas foi realizada na base de dados da *SentimentALL* para o preenchimento dos totalizadores relacionados às variáveis primárias do Estudo 2, que armazenam os quantitativos por destino turístico, por características e avaliações dos autores dos comentários e, principalmente, por avaliações positivas e negativas extraídas do processo de AS.

RESULTADOS

A ferramenta *SentimentALL*, concebida para gerar os dados necessários ao Estudo 2, com sua segunda versão implementada em Araújo (2017), utilizou uma série de algoritmos, apresentados na seção anterior, com o intuito de extrair informações verbais de avaliações de turistas-consumidores. Dessas informações, além dos dados relacionados ao turista e os esquemas avaliativos do próprio *site*, foram identificados nos comentários textuais dos turistas os aspectos (características) relacionados aos destinos turísticos, bem como as suas polaridades (negativa ou positiva). Em uma

síntese, a Figura 5 apresenta a estrutura da ferramenta, composta por seis módulos: 1) módulo de extração de dados direcionado à plataforma *TripAdvisor*®; 2) módulo de tratamento textual, necessário ao entendimento das partes que compõem os textos dos comentários, identificando substantivos, adjetivos etc. e suas relações em uma sentença; 3) módulo de Análise de Sentimentos, que identifica os aspectos e suas polaridades; 4) módulo de testes que verificou o quão eficiente a ferramenta foi na identificação dos aspectos e destes vinculados a uma polaridade; 5) módulo Totalizador, com o somatório relacionado a cada variável primária da tese; e, por último, 6) módulo de visualização de dados para a apresentação dos resultados de uma forma dinâmica.

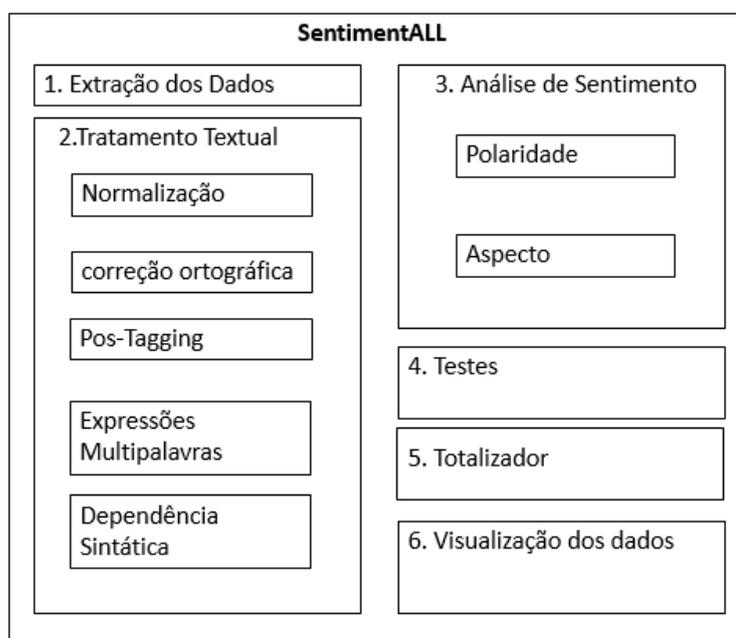


Figura 5: Estrutura da Ferramenta SentimentALL.

O processo realizado no módulo de **extração de dados** resultou no armazenamento de dados a partir da seguinte categorização: **1) Objeto:** nome (por exemplo, *Pousada dos Girassóis*), código, tipo (atração, acomodação, restaurante); **2) Destino:** cidade (código, nome), estado (nome), país (nome); **3) Autor:** nome, código, cidade, nível de colaboração (tem relação à pontuação dos *TripColaboradores* em uma escala de 1 a 6, mas para abranger todos os usuários, estabeleceu-se o nível 0, quando o usuário tem uma quantidade de pontos menor que a mínima considerada, ou seja, menos de 300 pontos; 1 = maior ou igual a 300 pontos e menor que 500 pontos; 2 = maior ou igual a 500 pontos e menor que 1.000; 3 = maior ou igual a 1.000

pontos e menor que 2.500; 4 = maior ou igual a 2.500 pontos e menor que 5.000; 5 = maior ou igual a 5.000 pontos e menor que 10.000; e 6 = mais de 10.000 pontos), ano de cadastro no *site*; **4) Avaliação:** título, nota (de 1 a 5, atribuída pelo turista-consumidor, organizada da seguinte forma: 1 = Horrível [HO]; 2 = Ruim [RU]; 3 [RZ] = Razoável; 4 = Muito bom [MB], e 5 = Excelente [EX]), data de postagem, comentário, código da avaliação, quantidade de pessoas que curtiram o comentário (relevância). As tabelas apresentadas no modelo relacional representado na Figura 6 ilustram como os dados extraídos no processo de coleta foram armazenados. As siglas PK (do inglês, *Primary Key*) e FK (do inglês, *Foreign Key*) presentes na Figura 6 indicam, respectivamente, quais os campos são identificadores únicos (chave primária) e quais são referências a uma chave primária de uma outra tabela (chave estrangeira).

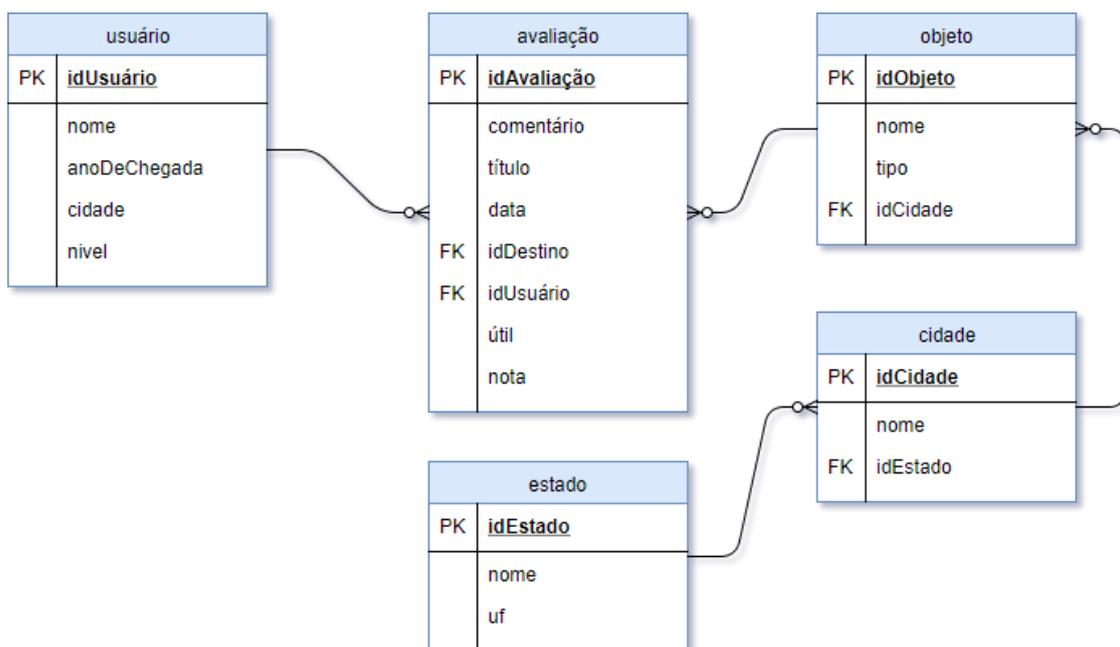


Figura 6: Modelo relacional parcial descritivo do armazenamento dos dados coletados.

O módulo de **tratamento textual**, composto por uma série de etapas, teve como resultado inicial a **normalização** do texto. Para exemplificar as modificações realizadas no texto nessa primeira etapa, tem-se o comentário apresentado na Figura 7. Pode-se verificar no campo A da Figura 7, os itens submetidos à etapa de normalização em destaque. No campo B, o texto apresenta o resultado da execução do processo. Por exemplo, a substituição das palavras com caracteres repetidos para sua versão normal, remoção dos *links* e datas que não interessam a análise,

substituição de expressões em “internetês” pelas equivalentes em português (por exemplo, “rs rs rs” por “risos”) e substituição dos *emoticons* por palavra da língua portuguesa descritiva da emoção indicada.

<p>AAAMMMEIII :) o atendimento desse hotel, apesar do preço caro R\$ 2.500,00 :(Ateh o café da manhã é maravilhoso rs rs rs instagram.com/fulanadetal/ estive por lá em <u>10/10/2015</u>.</p>	A
<p>Amei (feliz) o atendimento desse hotel, Apesar do preço caro (valor) (triste) Até o café da manhã é maravilhoso (risos) estive por lá em (data).</p>	B

Figura 7: Ilustra o procedimento de normalização dos dados textuais coletados.

Para que as palavras fossem identificadas de forma mais precisa, utilizou-se a etapa de **correção ortográfica**, com o propósito de evitar problemas como a não identificação de uma palavra opinativa (escrita incorretamente em um comentário) em um dicionário de léxico. Em Araújo (2017) foram realizados testes no algoritmo utilizado para esse processo com a utilização de 100 palavras corretas e 100 incorretas. Na identificação das palavras corretas, o algoritmo obteve 100% de acerto, de modo que não houve tentativa de correção desnecessária. Na identificação das palavras incorretas, a porcentagem diminuiu para 88% de acerto, o que já era esperado dada as inúmeras variações que resultam das etapas de transformação e do fato de a palavra ser selecionada a partir de um único critério: a frequência que uma palavra candidata ocorre no conjunto de termos. No entanto, os erros detectados nem sempre causam problemas no processo de análise de sentimentos, pois muitas vezes os erros têm relação com a conjugação do verbo em uma frase, o tempo verbal, etc. Com a finalização dessa etapa, os comentários foram armazenados no formato de sentenças normalizadas, corrigidas ortograficamente e *tokenizadas*. Esse último é o processo que divide as sentenças em conjuntos de *tokens*, quando cada elemento é apresentado como uma unidade, que pode ser uma palavra, um número ou uma marca de pontuação (por exemplo, amei, atendimento, hotel).

Na etapa de **POS-Tagging** foi realizada a categorização das palavras por classe gramatical (se adjetivo, se substantivo, etc.). Em Araújo e Brito (2018) foram apresentados os dois problemas verificados nessa etapa e suas respectivas soluções. Primeiro, com a utilização do Mac-Morpho, o procedimento de *POS-Tagging* não conseguiu identificar as contrações e combinações de preposições, por exemplo, nesse corpus a contração “da” é escrita na sua forma original, proposição “de” + o artigo “a”, ou seja, “de a”. Assim, foi utilizado o método implementado em Araújo (2017), que faz uma varredura no *corpus* buscando a etiqueta PREP relacionada às preposições e a sua próxima palavra (por exemplo etiquetada com ART, de artigo), e realizado a contração (PREP[de] + ART[a] = da). O segundo problema teve relação com as palavras não encontradas no Mac-Morpho, que eram etiquetadas como DEFAULT. Para tanto, utilizou-se uma lista de adjetivos e verbos disponíveis no site *Linguateca*²³. Assim, quando os *taggers* etiquetavam uma palavra como DEFAULT, essa palavra era verificada na lista de adjetivos e verbos, e etiquetada como ADJ ou V, conforme o caso. Se a palavra não fosse localizada nessas listas, era feita uma última análise, a verificação, por meio de expressão regular, se o termo era um número e, caso positivo, era etiquetada com NUM. Se, ao final, nenhuma verificação fosse satisfeita, as palavras eram etiquetadas como substantivo (*tag* N). A Tabela 3 apresenta exemplo do processo de *POS-Tagging* ao realizar a classificação gramatical dos *tokens* da sentença “O Parque do Jalapão é simplesmente maravilhoso”.

Tabela 3: Exemplo de classificação gramatical no processo de POS-Tagging.

ID	TOKENS	CLASSE GRAMATICAL
1	O	artigo
2	Parque	substantivo
3	do	preposição
4	Jalapão	substantivo
5	é	verbo
6	simplesmente	advérbio
7	maravilhoso	adjetivo

²³ <https://www.linguateca.pt>

A detecção de **expressões multipalavras** descritoras de componentes do produto turístico foi imprescindível para análise mais coerente dos relatos avaliativos presentes nos comentários, sendo mesmo controle metodológico a aprimorar substancialmente a confiabilidade do procedimento e validade interna da pesquisa. Por exemplo, a expressão “cachorro quente” sem o tratamento via PMI seria tratada como duas palavras distintas: o substantivo “cachorro” (o aspecto) e o adjetivo “quente” (a palavra opinativa), o que não teria qualquer sentido no contexto. A Tabela 4, apresentada a seguir, mostra alguns dos comentários dos turistas que continham expressões multipalavras, e como o tratamento dessas expressões torna possível sua detecção (ao menos, grande parte delas, conforme descrito na seção Método).

Tabela 4: Comentários de turistas com expressões multipalavras.

COMENTÁRIOS	EXPRESSÕES MULTIPALAVRAS
As crianças adoraram a chegada do papai noel de lancha.	papai noel
A ceia de natal foi sensacional.	ceia de natal
Recomendação especial para os pratos de inverno, tipo fondue e raclette, que são deliciosos!	pratos de inverno
A especialidade é a carne vermelha, que realmente é muito saborosa.	carne vermelha
Pedimos uma espumante (miolo 80 reais) servida com um copo de vidro de boteco.	copo de vidro
Mesmo assim, o quality está em meio ao centro histórico de são luís.	centro histórico
Utilizei o transporte público e não foi traumático.	transporte público
Fica ao lado de uma estrada de ferro, de uma avenida extremamente movimentada a noite e do aeroporto.	estrada ferro
Eu estava viajando de lua de mel e precisávamos descansar um pouco e o 'fast sleep' foi a solução.	lua de mel
A piscina térmica e externa são maravilhosas.	piscina térmica
O hotel disponibiliza de guarda sol e toalhas para a praia que é em frente ao hotel.	guarda sol
No sábado, dia 5, às 14,30 hs, foi que consertaram nosso ar condicionado.	ar condicionado

O Processamento de Linguagem Natural (PLN) é composto por diversas etapas de análise, cada uma delas permitindo abordagem diferenciada e, por vezes, aprofundamento no processo de descrição dos dados. Na etapa de análise sintática, utilizou-se técnica para a identificação de **dependências sintáticas**, que segundo

Nivre (2005, p. 3), “... é baseada na ideia de que a estrutura sintática de uma frase consiste de relações binárias assimétricas entre palavras da frase”. Para a exemplificação do quão relevante foi a análise de dependência sintática no processo de Análise de Sentimento (AS), pode-se considerar a seguinte frase: “O atendimento é muito bom e rápido”, apresentada na forma de um grafo de dependência na Figura 8.

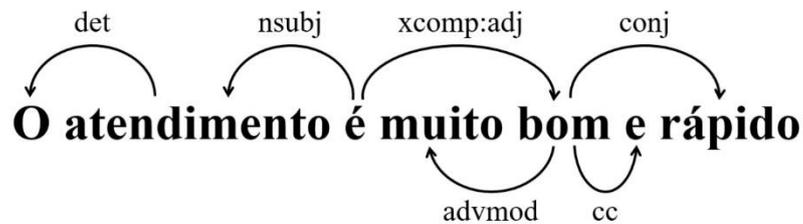


Figura 8: Relações de dependência em uma frase representadas um Grafo de Dependência.

Como pode ser observado na Figura 8, cada termo da frase tem uma relação de dependência. Com isso, é possível inferir que “atendimento” está relacionado a duas palavras opinativas, “bom” e “rápido”. Isso acontece porque “atendimento” relaciona-se ao verbo “é” como sujeito nominal deste, e o verbo tem como elemento mais próximo, o adjetivo “bom”, que ao unir-se por conjunção com o adjetivo “rápido”, relaciona-o também ao verbo e, conseqüentemente, liga-o ao sujeito nominal “atendimento”. Além disso, ainda tem a relação do advérbio “muito” modificando a intensidade do adjetivo “bom”. Com a execução desse procedimento, é finalizada a etapa de tratamento textual e gerado o arquivo utilizado na etapa de AS.

O módulo de **Análise de Sentimento** utiliza a base de palavras presente nos léxicos de sentimentos *SentiLex*, *OpLexicon* e *LIWC*; e o banco de dados que contém as sentenças devidamente tratadas nas etapas de PLN. Com a utilização do algoritmo para a identificação das palavras opinativas (descrito na seção Método), foi criada uma lista relacionada às palavras opinativas em cada sentença. Por exemplo, na sentença “O₍₁₎ restaurante₍₂₎ é₍₃₎ bom₍₄₎ e₍₅₎ aconchegante₍₆₎”, foram armazenadas na lista as seguintes informações: a) identificadores da opinião [4, 6]; b) *Head*, parâmetro que representa o *Id* da palavra que é pai da palavra em questão (opinião) na árvore de dependências [3, 4]; c) a relação estabelecida entre os dependidos [xcomp:adj, CONJ] (o primeiro é a relação entre “bom” e o verbo “é”, referindo-se ao fato do verbo ter como complemento, do tipo adjetivo, a palavra opinativa; e o segundo é a relação

entre as opiniões que estão ligadas pela conjunção “e”); *d*) a etiqueta da palavra que expressa a opinião [ADJ, ADJ] (as duas palavras são do tipo adjetivo); *e*) a polaridade [1,1] (polaridade positiva), e *f*) as palavras opinativas [bom, aconchegante].

Na sequência da etapa de AS é determinada a polaridade final para os casos em que as palavras opinativas na sentença vêm acompanhadas de palavras negativas ou adversativas. Assim, a lista é percorrida para verificar se há a ocorrência dessas palavras. Supondo as sentenças S1: “*O*₍₁₎ *restaurante*₍₂₎ *não*₍₃₎ *é*₍₄₎ *bom*₍₅₎” e S2: “*O*₍₁₎ *quarto*₍₂₎ *é*₍₃₎ *pequeno*₍₄₎, *porém*₍₅₎ *confortável*₍₆₎” tem-se a situação exposta na Tabela 5:

Tabela 5: Lista relacionada às palavras identificadas como opiniões.

LISTA DE INFORMAÇÕES	S1	S2
<i>ID da Opinião</i>	5	4, 6
<i>Head</i>	4	3, 4
<i>Relação</i>	Entre 4 e 5: xcomp:adj	Entre 3 e 4: xcomp:adj Entre 4 e 6: conjunção adversativa
<i>Tag</i>	ADJ	ADJ, ADJ
<i>Polaridade</i>	1	0, 1
<i>Opinião</i>	bom	pequeno, confortável

Para a determinação da polaridade final, o algoritmo implementado em Araújo (2017) realizou busca na lista relacionada às opiniões para verificar se na sentença há a incidência de palavras negativas ou adversativas como dependentes da opinião, conforme ilustrado na Tabela 6.

Tabela 6: Lista de palavras negativas e adversativas.

PALAVRAS NEGATIVAS	PALAVRAS ADVERSATIVAS
<i>não; tampouco; nem; nunca; jamais; mal</i>	<i>mas; porém; contudo; todavia; entretanto; embora; apesar</i>

Há três formas de negação identificadas a partir do algoritmo. A primeira ocorre quando a palavra opinativa depende diretamente de alguma palavra negativa, como na frase “*Não gostei*”. A identificação se dá quando é verificado que a palavra opinativa está relacionada a um advérbio e que este advérbio está na lista de palavras negativas. Quando isso acontece, a polaridade da palavra opinativa é modificada (no

caso, de 1 para -1). A outra situação ocorre quando o verbo está relacionado com duas palavras, uma opinião e um advérbio, como acontece na sentença S1 da Tabela 5. Nesse caso, a opinião será modificada pelo advérbio caso esteja na relação de palavras negativas. Por fim, a última forma está relacionada às conjunções adversativas, que são identificadas a partir de um verbo que tenha como complemento dois adjetivos e um deles apresenta uma conjunção adversativa como dependente, como é o caso do exemplo apresentado na sentença S2 da Tabela 5. Nesse caso, o algoritmo verifica a palavra opinativa que tem dependência com uma conjunção adversativa (no exemplo, “confortável”), multiplica o valor da polaridade dessa opinião por -1 e usa o resultado dessa operação para modificar a polaridade da outra palavra opinativa, quando esta tiver a polaridade neutra (igual a zero [0]). Assim, resultado da Tabela 5 modificado é apresentado na Tabela 7.

Tabela 7: Lista relacionada às palavras identificadas como opiniões com polaridade final.

LISTA DE INFORMAÇÕES	S1	S2
<i>ID da Opinião</i>	5	4, 6
<i>Head</i>	4	3, 4
<i>Relação</i>	Entre 4 e 5: xcomp:adj	Entre 3 e 4: xcomp:adj Entre 4 e 6: conj adversativa
<i>Tag</i>	ADJ	ADJ, ADJ
<i>Polaridade</i>	-1	-1, 1
<i>Opinião</i>	bom	pequeno, confortável

Contudo, mesmo utilizando três (3) dicionários de léxicos de sentimentos (Figura 4, seção Método) e realizadas as transformações relacionadas às palavras negativas ou adversativas, algumas palavras opinativas ficaram com o valor de polaridade neutra, conforme pode ser observado na Tabela 8.

Tabela 8: Palavras identificadas como opinativas.

66%	14.416.952	Palavras positivas (itens duplicados)
10%	2.143.245	Palavras negativas (itens duplicados)
24%	5.294.939	Palavras neutras (itens duplicados)

Assim, no final dessas etapas, com os testes realizados em Araújo (2017) e de acordo com as conclusões de Hatzivassiloglou e McKeown (1997), todas as opiniões

que estavam com a polaridade neutra passaram a ser positiva. Em Hatzivassiloglou e McKeown (1997), foi verificado que as palavras positivas são as mais frequentes e na primeira versão da *SentimentALL* (Brito, Oliveira, & Souza, 2015), verificou-se que 83% das palavras que estavam com o *status* de neutra eram, de fato, positivas.

Na etapa de identificação dos aspectos, o fato das palavras estarem relacionadas por meio das dependências sintáticas contribuiu para a associação mais precisa dos substantivos (ou conjunto de substantivos) às suas palavras opinativas. Dessa forma, quando o aspecto é identificado, também é possível detectar se este está relacionado a outro aspecto, como no exemplo apresentado na sentença “*O atendimento, o restaurante e o quarto são maravilhosos*”. De forma análoga ao processo da palavra opinativa, a dependência é detectada pela conjunção “e”. Assim, é feita uma mesclagem dos substantivos (nesse exemplo, *atendimento, restaurante e quarto*), relacionando-os à mesma palavra opinativa (nesse caso, *maravilhosos*), logo à mesma polaridade (positiva).

Para verificar o quão os resultados da *SentimentALL* foram coerentes no processo de determinação dos aspectos das sentenças e das suas polaridades, alguns **testes** foram realizados. Para tanto, foram usados, novamente, os critérios estatísticos *Precision*, *Recall* e *F-Measure* para avaliação de processos de classificação de texto. Para viabilizar a análise dos testes, foi feito trabalho de anotação manual da relação dos aspectos com suas polaridades em 100 comentários-piloto, trabalho este realizado com a utilização da ferramenta *on-line* implementada em Oliveira (2015) e apresentada na Figura 9. Os 100 comentários foram divididos em duas categorias: Simples e Complexo. Considerou-se comentários Simples aqueles que não deram margem a ambiguidades, por exemplo: “*O atendimento é excelente, porém o preço da comida é caro*”. Os comentários Complexos foram aqueles com alguma informação implícita, com duplo sentido ou que sobre o qual não havia certeza sobre as relações entre aspectos e palavras opinativas na sentença, por exemplo: “*Muito boa mesmo, considero a melhor da cidade, amei*”. Três colaboradores realizaram as anotações manuais dos aspectos (características) de cada sentença, indicando as polaridades (se negativa, destaque em vermelho; se positiva, destaque em azul).

Marque as características identificadas no comentário. As características positivas devem ser marcadas de azul e as negativas de vermelho.

Excelente [localização](#) e [atendimento](#) Fica na esquina com o shopping azul. O [café da manhã](#) o [almoço](#) e o [jantar](#) são saborosos e diversificados. Mas [sauna](#) [piscina](#) estão em manutenção. Em compensação, o hotel oferece uma boa [programação cultural](#)

Detalhando:

"café da manhã" por ser um termo composto foi selecionada por completo, incluindo os espaços. O mesmo ocorre com o termo "programação cultural". Por outro lado, piscina e sauna devem ser selecionadas separadamente pois são características distintas.

Destino: São Paulo - SP

Objeto: Hotel Transamérica São Paulo

Qualidade do sono:4,5;Localização:4;Quartos:4,5;Atendimento:4,5;Custo-benefício:3,5;Limpeza:4,5;

A

Excelente Hotel

O hotel é excelente. Fiquei impressionado com as acomodações e com a atenção dos profissionais. Meu pai havia feito uma cirurgia e recebeu um tratamento muito especial. Eu indico e voltarei outras vezes.

Figura 9: Ferramenta on-line de análise e anotação manual da relação entre aspectos e polaridades.

Em Leal, Oliveira e Brito (2017) foi feita a análise da primeira versão do *SentimentALL* e em Araújo (2017) a análise da segunda versão. Para cada tipo de comentário (simples e complexo), duas análises foram feitas. A primeira considerou apenas a identificação correta dos aspectos na sentença, e a segunda considerou a identificação correta dos aspectos com a polaridade. Como pode ser observado nos gráficos da Figura 10, houve melhora no desempenho da *SentimentALL* v2 nas duas situações (o acerto na identificação dos aspectos e o acerto na identificação dos aspectos e da polaridade). Na primeira análise, o *F-Measure* passou de 61,96% para 76,50% (ganho de 14,54%). Na segunda análise, o aumento foi ainda mais considerável, saindo de 55,77% na *SentimentALL* v1 para 72,15 na *SentimentALL* v2 (ganho de 16,38%). Na análise dos comentários complexos, apresentada na Figura 11, o desempenho do sistema para aspecto aumentou de 44,27% na *SentimentALL* v1 para 61,18% na *SentimentALL* v2 (ganho de 16,91%). Na análise que considerou aspecto e polaridade verificou-se o maior ganho, que foi de 18,17%, visto o aumento de 37,82% na *SentimentALL* v1 para 55,99% na *SentimentALL* v2.

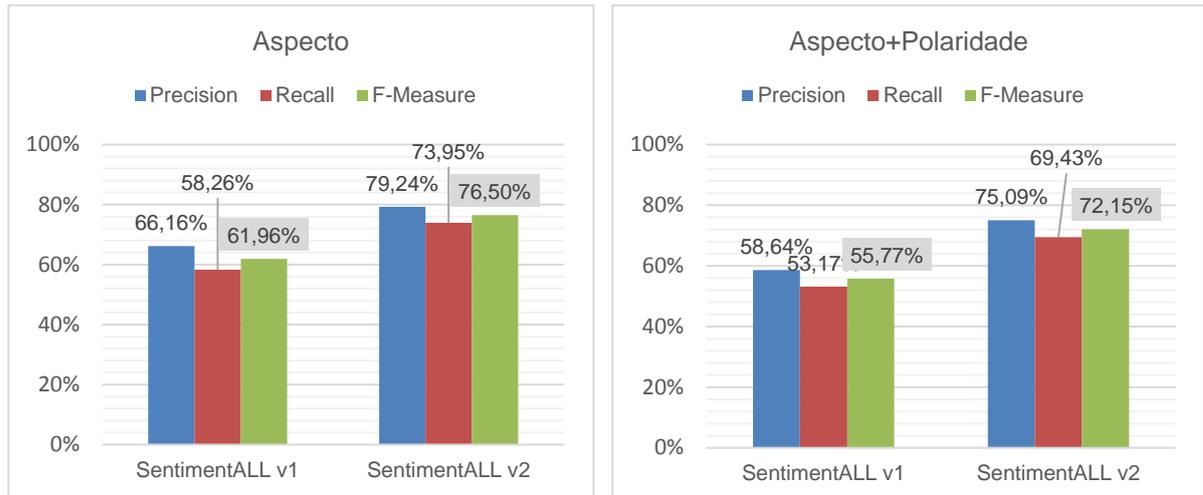


Figura 10: Análises de aspecto, e aspecto e polaridade de comentários simples feitas pelas duas versões do SentimentALL, conforme os critérios Precision, Recall e F-Measure.

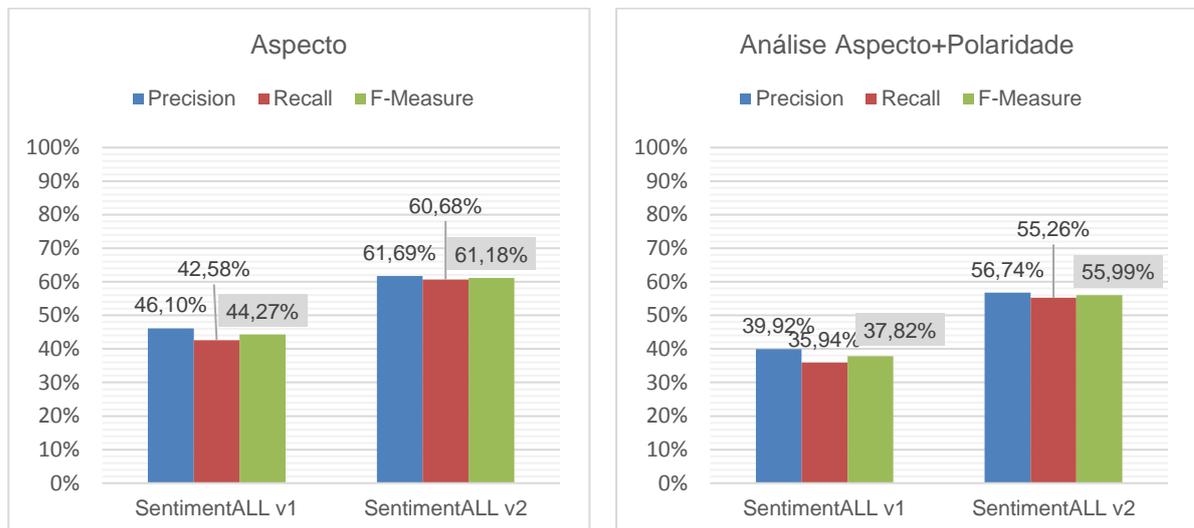


Figura 11: Análises de aspecto, e aspecto e polaridade de comentários complexos feitas pelas duas versões do SentimentALL, conforme os critérios Precision, Recall e F-Measure.

Após o processo de coleta, processamento, análise de sentimento e testes, foi realizada a **totalização** das 197 variáveis primárias desta tese, relacionadas com os dados extraídos e/ou inferidos nas etapas anteriores. Na Figura 12, que apresenta o modelo relacional do banco de dados da ferramenta *SentimentALL v2*, tem-se a tabela “Totalizador”, responsável por armazenar as acumulações. Essa tabela possui os seguinte campos: *id* (identificador de cada registro na tabela), Variável (Descritor da variável armazenada, por exemplo, V006 - Número total geral de “SENTIMENTOS POSITIVOS” para ACOMODAÇÕES), Valor (valor obtido através da consulta realizada na base), Descrição (descrição das solicitações realizadas), Versão (identificador da data em que o armazenamento foi realizado), *idCidade* (identificador

da cidade, campo que possui uma chave estrangeira [FK] indicativa de que este campo aponta para uma chave primária [PK] de outra tabela, no caso, da tabela Cidade).

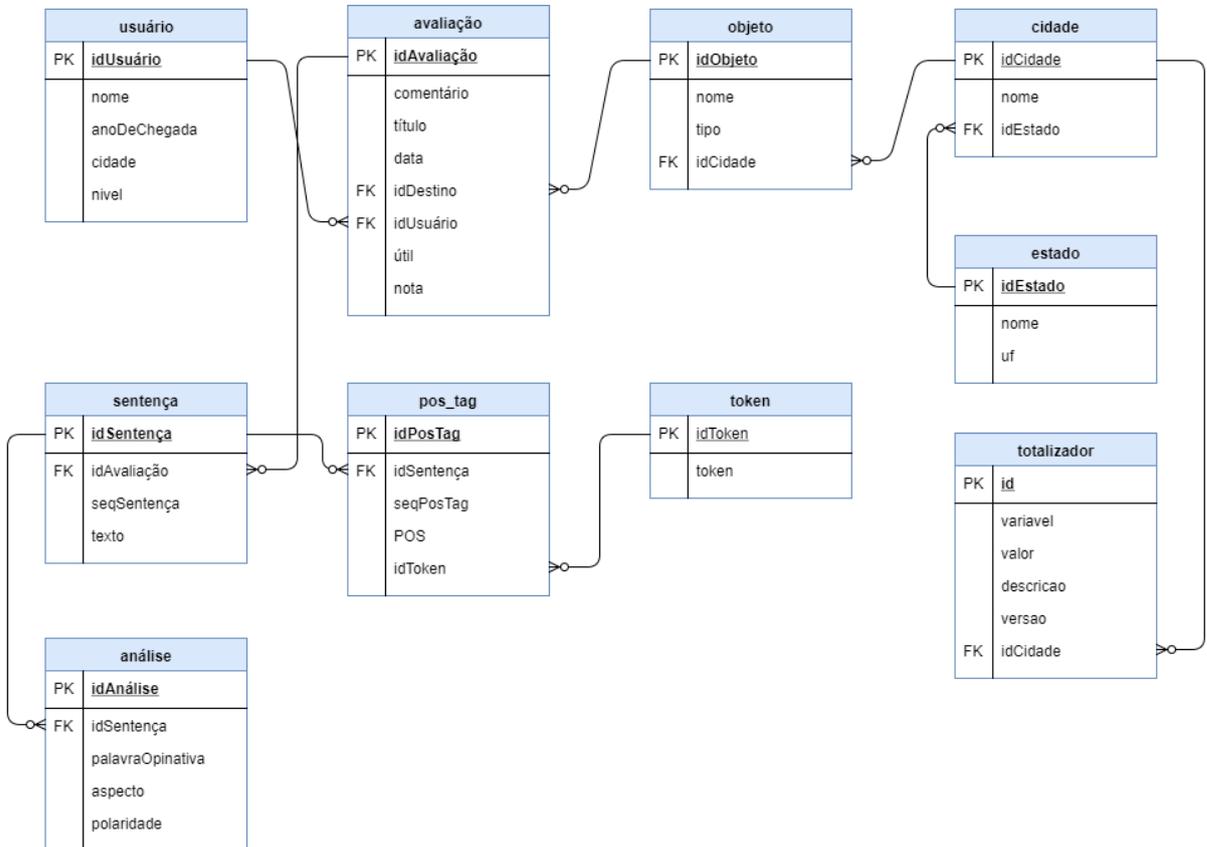


Figura 12: Modelo relacional parcial para coleta de dados, pré-processamento, análises e totalizações.

Os somatórios armazenados na tabela Totalizador foram utilizados na etapa de **visualização** em forma de um *dashboard*, conforme pode ser observado na Figura 13.

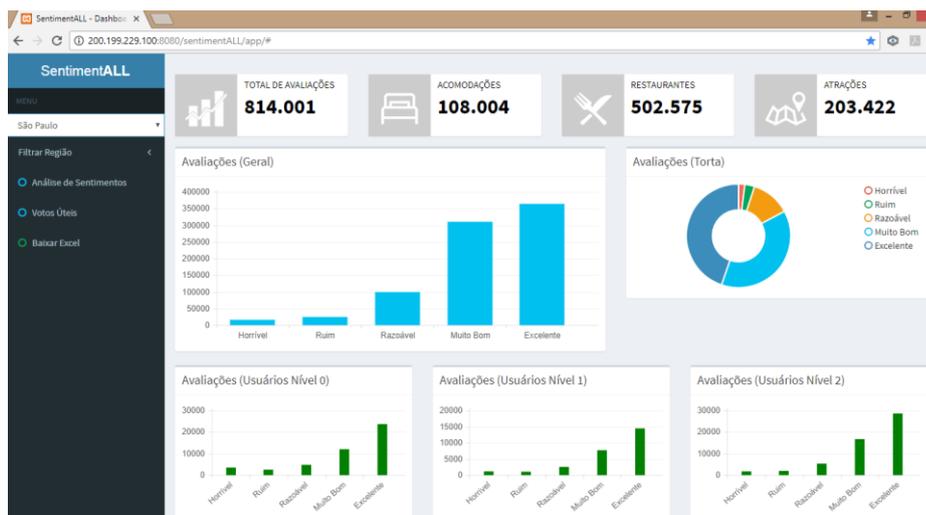


Figura 13: Dashboard da Ferramenta SentimentALL (Souza, 2017).

Com a verificação da eficiência da ferramenta em relação ao processo de análise de sentimentos e a consequente extração dos aspectos com suas respectivas polaridades, as 197 variáveis armazenadas na tabela Totalizador (apresentada na Figura 12) tornaram-se a fonte primária das análises comportamentais do Estudo 2. Essas variáveis (apresentadas na Tabela 9) têm relação com o processo de AS que identificou os aspectos positivos e negativos em cada um dos três componentes do produto turístico (ACO, RES, ATR) e, também, com os dados extraídos sobre os colaboradores e suas avaliações, por exemplo, seu nível de colaborador, sua avaliação considerando a escala *Likert* do *TripAdvisor*® e, por fim, as totalizações de votos úteis nos comentários por cada um dos três componentes (ACO, RES, ATR).

Tabela 9: Variáveis geradas pela SentimentALL v2.

V001	Nome dos destinos analisados
V002	Número total geral de comentários sobre o destino, usado na análise de sentimentos
V003	Número total geral de comentários sobre ACOMODAÇÕES no destino, usado na análise de sentimentos
V004	Número total geral de comentários sobre RESTAURANTES no destino, usado na análise de sentimentos
V005	Número total geral de comentários sobre ATRAÇÕES no destino, usado na análise de sentimentos
V006	Número total geral de avaliações ("sentimentos") POSITIVAS para ACOMODAÇÕES
V007	Número total geral de avaliações ("sentimentos") NEGATIVAS para ACOMODAÇÕES
V008	Número total geral de avaliações ("sentimentos") POSITIVAS para RESTAURANTES
V009	Número total geral de avaliações ("sentimentos") NEGATIVAS para RESTAURANTES
V010	Número total geral de avaliações ("sentimentos") POSITIVAS para ATRAÇÕES
V011	Número total geral de avaliações ("sentimentos") NEGATIVAS para ATRAÇÕES
V012	Número total geral de avaliações HORRÍVEL para ACOMODAÇÕES
V013	Número total geral de avaliações RUIM para ACOMODAÇÕES
V014	Número total geral de avaliações RAZOÁVEL para ACOMODAÇÕES
V015	Número total geral de avaliações MUITO BOM para ACOMODAÇÕES
V016	Número total geral de avaliações EXCELENTE para ACOMODAÇÕES
V017	Número total geral de avaliações HORRÍVEL para RESTAURANTES
V018	Número total geral de avaliações RUIM para RESTAURANTES

Tabela 9: (Continuação) Variáveis geradas pela SentimentALL v2.

V019	Número total geral de avaliações RAZOÁVEL para RESTAURANTES
V020	Número total geral de avaliações MUITO BOM para RESTAURANTES
V021	Número total geral de avaliações EXCELENTE para RESTAURANTES
V022	Número total geral de avaliações HORRÍVEL para ATRAÇÕES
V023	Número total geral de avaliações RUIM para ATRAÇÕES
V024	Número total geral de avaliações RAZOÁVEL para ATRAÇÕES
V025	Número total geral de avaliações MUITO BOM para ATRAÇÕES
V026	Número total geral de avaliações EXCELENTE para ATRAÇÕES
V027	Número total de avaliações (“sentimentos”) POSITIVAS para ACOMODAÇÕES pelos CN_0
V028	Número total de avaliações (“sentimentos”) NEGATIVAS para ACOMODAÇÕES pelos CN_0
V029	Número total de avaliações (“sentimentos”) POSITIVAS para RESTAURANTES pelos CN_0
V030	Número total de avaliações (“sentimentos”) NEGATIVAS para RESTAURANTES pelos CN_0
V031	Número total de avaliações (“sentimentos”) POSITIVAS para ATRAÇÕES pelos CN_0
V032	Número total de avaliações (“sentimentos”) NEGATIVAS para ATRAÇÕES pelos CN_0
V033	Número total de avaliações HORRÍVEL para ACOMODAÇÕES pelos CN_0
V034	Número total de avaliações RUIM para ACOMODAÇÕES pelos CN_0
V035	Número total de avaliações RAZOÁVEL para ACOMODAÇÕES pelos CN_0
V036	Número total de avaliações MUITO BOM para ACOMODAÇÕES pelos CN_0
V037	Número total de avaliações EXCELENTE para ACOMODAÇÕES pelos CN_0
V038	Número total de avaliações HORRÍVEL para RESTAURANTES pelos CN_0
V039	Número total de avaliações RUIM para RESTAURANTES pelos CN_0
V040	Número total de avaliações RAZOÁVEL para RESTAURANTES pelos CN_0
V041	Número total de avaliações MUITO BOM para RESTAURANTES pelos CN_0
V042	Número total de avaliações EXCELENTE para RESTAURANTES pelos CN_0
V042	Número total de avaliações HORRÍVEL para ATRAÇÕES pelos CN_0
V044	Número total de avaliações RUIM para ATRAÇÕES pelos CN_0
V045	Número total de avaliações RAZOÁVEL para ATRAÇÕES pelos CN_0
V046	Número total de avaliações MUITO BOM para ATRAÇÕES pelos CN_0
V047	Número total de avaliações EXCELENTE para ATRAÇÕES pelos CN_0
V048	Número total de avaliações (“sentimentos”) POSITIVAS para ACOMODAÇÕES pelos CN_1
V049	Número total de avaliações (“sentimentos”) NEGATIVAS para ACOMODAÇÕES pelos CN_1
V050	Número total de avaliações (“sentimentos”) POSITIVAS para RESTAURANTES pelos CN_1
V051	Número total de avaliações (“sentimentos”) NEGATIVAS para RESTAURANTES pelos CN_1
V052	Número total de avaliações (“sentimentos”) POSITIVAS para ATRAÇÕES pelos CN_1
V053	Número total de avaliações (“sentimentos”) NEGATIVAS para ATRAÇÕES pelos CN_1
V054	Número total de avaliações HORRÍVEL para ACOMODAÇÕES pelos CN_1
V055	Número total de avaliações RUIM para ACOMODAÇÕES pelos CN_1
V056	Número total de avaliações RAZOÁVEL para ACOMODAÇÕES pelos CN_1
V057	Número total de avaliações MUITO BOM para ACOMODAÇÕES pelos CN_1
V058	Número total de avaliações EXCELENTE para ACOMODAÇÕES pelos CN_1
V059	Número total de avaliações HORRÍVEL para RESTAURANTES pelos CN_1
V060	Número total de avaliações RUIM para RESTAURANTES pelos CN_1
V061	Número total de avaliações RAZOÁVEL para RESTAURANTES pelos CN_1
V062	Número total de avaliações MUITO BOM para RESTAURANTES pelos CN_1
V063	Número total de avaliações EXCELENTE para RESTAURANTES pelos CN_1
V064	Número total de avaliações HORRÍVEL para ATRAÇÕES pelos CN_1
V065	Número total de avaliações RUIM para ATRAÇÕES pelos CN_1
V066	Número total de avaliações RAZOÁVEL para ATRAÇÕES pelos CN_1
V067	Número total de avaliações MUITO BOM para ATRAÇÕES pelos CN_1
V068	Número total de avaliações EXCELENTE para ATRAÇÕES pelos CN_1
V069	Número total de avaliações (“sentimentos”) POSITIVAS para ACOMODAÇÕES pelos CN_2
V070	Número total de avaliações (“sentimentos”) NEGATIVAS para ACOMODAÇÕES pelos CN_2
V071	Número total de avaliações (“sentimentos”) POSITIVAS para RESTAURANTES pelos CN_2
V072	Número total de avaliações (“sentimentos”) NEGATIVAS para RESTAURANTES pelos CN_2
V073	Número total de avaliações (“sentimentos”) POSITIVAS para ATRAÇÕES pelos CN_2
V074	Número total de avaliações (“sentimentos”) NEGATIVAS para ATRAÇÕES pelos CN_2
V075	Número total de avaliações HORRÍVEL para ACOMODAÇÕES pelos CN_2

Tabela 9: (Continuação) Variáveis geradas pela SentimentALL v2.

V076	Número total de avaliações RUIM para ACOMODAÇÕES pelos CN_2
V077	Número total de avaliações RAZOÁVEL para ACOMODAÇÕES pelos CN_2
V078	Número total de avaliações MUITO BOM para ACOMODAÇÕES pelos CN_2
V079	Número total de avaliações EXCELENTE para ACOMODAÇÕES pelos CN_2
V080	Número total de avaliações HORRÍVEL para RESTAURANTES pelos CN_2
V081	Número total de avaliações RUIM para RESTAURANTES pelos CN_2
V082	Número total de avaliações RAZOÁVEL para RESTAURANTES pelos CN_2
V083	Número total de avaliações MUITO BOM para RESTAURANTES pelos CN_2
V084	Número total de avaliações EXCELENTE para RESTAURANTES pelos CN_2
V085	Número total de avaliações HORRÍVEL para ATRAÇÕES pelos CN_2
V086	Número total de avaliações RUIM para ATRAÇÕES pelos CN_2
V087	Número total de avaliações RAZOÁVEL para ATRAÇÕES pelos CN_2
V088	Número total de avaliações MUITO BOM para ATRAÇÕES pelos CN_2
V089	Número total de avaliações EXCELENTE para ATRAÇÕES pelos CN_2
V090	Número total de avaliações (“sentimentos”) POSITIVAS para ACOMODAÇÕES pelos CN_3
V091	Número total de avaliações (“sentimentos”) NEGATIVAS para ACOMODAÇÕES pelos CN_3
V092	Número total de avaliações (“sentimentos”) POSITIVAS para RESTAURANTES pelos CN_3
V093	Número total de avaliações (“sentimentos”) NEGATIVAS para RESTAURANTES pelos CN_3
V094	Número total de avaliações (“sentimentos”) POSITIVAS para ATRAÇÕES pelos CN_3
V095	Número total de avaliações (“sentimentos”) NEGATIVAS para ATRAÇÕES pelos CN_3
V096	Número total de avaliações HORRÍVEL para ACOMODAÇÕES pelos CN_3
V097	Número total de avaliações RUIM para ACOMODAÇÕES pelos CN_3
V098	Número total de avaliações RAZOÁVEL para ACOMODAÇÕES pelos CN_3
V099	Número total de avaliações MUITO BOM para ACOMODAÇÕES pelos CN_3
V100	Número total de avaliações EXCELENTE para ACOMODAÇÕES pelos CN_3
V101	Número total de avaliações HORRÍVEL para RESTAURANTES pelos CN_3
V102	Número total de avaliações RUIM para RESTAURANTES pelos CN_3
V103	Número total de avaliações RAZOÁVEL para RESTAURANTES pelos CN_3
V104	Número total de avaliações MUITO BOM para RESTAURANTES pelos CN_3
V105	Número total de avaliações EXCELENTE para RESTAURANTES pelos CN_3
V106	Número total de avaliações HORRÍVEL para ATRAÇÕES pelos CN_3
V107	Número total de avaliações RUIM para ATRAÇÕES pelos CN_3
V108	Número total de avaliações RAZOÁVEL para ATRAÇÕES pelos CN_3
V109	Número total de avaliações MUITO BOM para ATRAÇÕES pelos CN_3
V110	Número total de avaliações EXCELENTE para ATRAÇÕES pelos CN_3
V111	Número total de avaliações (“sentimentos”) POSITIVAS para ACOMODAÇÕES pelos CN_4
V112	Número total de avaliações (“sentimentos”) NEGATIVAS para ACOMODAÇÕES pelos CN_4
V113	Número total de avaliações (“sentimentos”) POSITIVAS para RESTAURANTES pelos CN_4
V114	Número total de avaliações (“sentimentos”) NEGATIVAS para RESTAURANTES pelos CN_4
V115	Número total de avaliações (“sentimentos”) POSITIVAS para ATRAÇÕES pelos CN_4
V116	Número total de avaliações (“sentimentos”) NEGATIVAS para ATRAÇÕES pelos CN_4
V117	Número total de avaliações HORRÍVEL para ACOMODAÇÕES pelos CN_4
V118	Número total de avaliações RUIM para ACOMODAÇÕES pelos CN_4
V119	Número total de avaliações RAZOÁVEL para ACOMODAÇÕES pelos CN_4
V120	Número total de avaliações MUITO BOM para ACOMODAÇÕES pelos CN_4
V121	Número total de avaliações EXCELENTE para ACOMODAÇÕES pelos CN_4
V122	Número total de avaliações HORRÍVEL para RESTAURANTES pelos CN_4
V123	Número total de avaliações RUIM para RESTAURANTES pelos CN_4
V124	Número total de avaliações RAZOÁVEL para RESTAURANTES pelos CN_4
V125	Número total de avaliações MUITO BOM para RESTAURANTES pelos CN_4
V126	Número total de avaliações EXCELENTE para RESTAURANTES pelos CN_4
V127	Número total de avaliações HORRÍVEL para ATRAÇÕES pelos CN_4
V128	Número total de avaliações RUIM para ATRAÇÕES pelos CN_4
V129	Número total de avaliações RAZOÁVEL para ATRAÇÕES pelos CN_4
V130	Número total de avaliações MUITO BOM para ATRAÇÕES pelos CN_4
V131	Número total de avaliações EXCELENTE para ATRAÇÕES pelos CN_4
V132	Número total de avaliações (“sentimentos”) POSITIVAS para ACOMODAÇÕES pelos CN_5

Tabela 9:(Continuação) Variáveis geradas pela SentimentALL v2.

V133	Número total de avaliações ("sentimentos") NEGATIVAS para ACOMODAÇÕES pelos CN_5
V134	Número total de avaliações ("sentimentos") POSITIVAS para RESTAURANTES pelos CN_5
V135	Número total de avaliações ("sentimentos") NEGATIVAS para RESTAURANTES pelos CN_5
V136	Número total de avaliações ("sentimentos") POSITIVAS para ATRAÇÕES pelos CN_5
V137	Número total de avaliações ("sentimentos") NEGATIVAS para ATRAÇÕES pelos CN_5
V138	Número total de avaliações HORRÍVEL para ACOMODAÇÕES pelos CN_5
V139	Número total de avaliações RUIM para ACOMODAÇÕES pelos CN_5
V140	Número total de avaliações RAZOÁVEL para ACOMODAÇÕES pelos CN_5
V141	Número total de avaliações MUITO BOM para ACOMODAÇÕES pelos CN_5
V142	Número total de avaliações EXCELENTE para ACOMODAÇÕES pelos CN_5
V143	Número total de avaliações HORRÍVEL para RESTAURANTES pelos CN_5
V144	Número total de avaliações RUIM para RESTAURANTES pelos CN_5
V145	Número total de avaliações RAZOÁVEL para RESTAURANTES pelos CN_5
V146	Número total de avaliações MUITO BOM para RESTAURANTES pelos CN_5
V147	Número total de avaliações EXCELENTE para RESTAURANTES pelos CN_5
V148	Número total de avaliações HORRÍVEL para ATRAÇÕES pelos CN_5
V149	Número total de avaliações RUIM para ATRAÇÕES pelos CN_5
V150	Número total de avaliações RAZOÁVEL para ATRAÇÕES pelos CN_5
V151	Número total de avaliações MUITO BOM para ATRAÇÕES pelos CN_5
V152	Número total de avaliações EXCELENTE para ATRAÇÕES pelos CN_5
V153	Número total de avaliações ("sentimentos") POSITIVAS para ACOMODAÇÕES pelos CN_6
V154	Número total de avaliações ("sentimentos") NEGATIVAS para ACOMODAÇÕES pelos CN_6
V155	Número total de avaliações ("sentimentos") POSITIVAS para RESTAURANTES pelos CN_6
V156	Número total de avaliações ("sentimentos") NEGATIVAS para RESTAURANTES pelos CN_6
V157	Número total de avaliações ("sentimentos") POSITIVAS para ATRAÇÕES pelos CN_6
V158	Número total de avaliações ("sentimentos") NEGATIVAS para ATRAÇÕES pelos CN_6
V159	Número total de avaliações HORRÍVEL para ACOMODAÇÕES pelos CN_6
V160	Número total de avaliações RUIM para ACOMODAÇÕES pelos CN_6
V161	Número total de avaliações RAZOÁVEL para ACOMODAÇÕES pelos CN_6
V162	Número total de avaliações MUITO BOM para ACOMODAÇÕES pelos CN_6
V163	Número total de avaliações EXCELENTE para ACOMODAÇÕES pelos CN_6
V164	Número total de avaliações HORRÍVEL para RESTAURANTES pelos CN_6
V165	Número total de avaliações RUIM para RESTAURANTES pelos CN_6
V166	Número total de avaliações RAZOÁVEL para RESTAURANTES pelos CN_6
V167	Número total de avaliações MUITO BOM para RESTAURANTES pelos CN_6
V168	Número total de avaliações EXCELENTE para RESTAURANTES pelos CN_6
V169	Número total de avaliações HORRÍVEL para ATRAÇÕES pelos CN_6
V170	Número total de avaliações RUIM para ATRAÇÕES pelos CN_6
V171	Número total de avaliações RAZOÁVEL para ATRAÇÕES pelos CN_6
V172	Número total de avaliações MUITO BOM para ATRAÇÕES pelos CN_6
V173	Número total de avaliações EXCELENTE para ATRAÇÕES pelos CN_6
V174	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES
V175	Número total geral de likes (votos úteis) para comentários sobre RESTAURANTES
V176	Número total geral de likes (votos úteis) para comentários sobre ATRAÇÕES
V177	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES feitos pelos CN_0
V178	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES feitos pelos CN_1
V179	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES pelos CN_2
V180	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES pelos CN_3
V181	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES pelos CN_4
V182	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES pelos CN_5
V183	Número total geral de likes (votos úteis) para comentários sobre ACOMODAÇÕES pelos CN_6
V184	Número total geral de likes (votos úteis) para comentários sobre RESTAURANTES feitos pelos CN_0
V185	Número total geral de likes (votos úteis) para comentários sobre RESTAURANTES feitos pelos CN_1
V186	Número total geral de likes (votos úteis) para comentários sobre RESTAURANTES pelos CN_2
V187	Número total geral de likes (votos úteis) para comentários sobre RESTAURANTES pelos CN_3
V188	Número total geral de likes (votos úteis) para comentários sobre RESTAURANTES pelos CN_4
V189	Número total geral de likes (votos úteis) para comentários sobre RESTAURANTES pelos CN_5

Tabela 9: (Continuação) Variáveis geradas pela *SentimentALL* v2.

V190	Número total geral de <i>likes</i> (votos úteis) para comentários sobre RESTAURANTES pelos CN_6
V191	Número total geral de <i>likes</i> (votos úteis) para comentários sobre ATRAÇÕES feitos pelos CN_0
V192	Número total geral de <i>likes</i> (votos úteis) para comentários sobre ATRAÇÕES feitos pelos CN_1
V193	Número total geral de <i>likes</i> (votos úteis) para comentários sobre ATRAÇÕES pelos CN_2
V194	Número total geral de <i>likes</i> (votos úteis) para comentários sobre ATRAÇÕES pelos CN_3
V195	Número total geral de <i>likes</i> (votos úteis) para comentários sobre ATRAÇÕES pelos CN_4
V196	Número total geral de <i>likes</i> (votos úteis) para comentários sobre ATRAÇÕES pelos CN_5
V197	Número total geral de <i>likes</i> (votos úteis) para comentários sobre ATRAÇÕES pelos CN_6

Os dados da tabela Totalizador (Figura 12) foram exportados para um arquivo .xlsx (*Microsoft Excel®*) com 19.700 entradas. Essa foi a última etapa do Estudo 1 e a fonte primária de onde as variáveis secundárias, exploradas nas análises do Estudo 2, foram geradas. Na próxima seção serão realizadas as discussões acerca do Estudo 1, com ênfase nas técnicas utilizadas, na importância de análises de grandes conjuntos de dados para o setor do turismo e nas possibilidades de estudos futuros que possam ampliar o alcance da *SentimentALL*.

DISCUSSÃO

A concepção da ferramenta *SentimentALL*, objetivo do Estudo 1, foi condição indispensável ao alcance dos objetivos da Tese como um todo (Estudo 1 e Estudo 2), visto que os dados pretendidos para análise e interpretação no Estudo 2 estão baseados nas respostas emitidas por consumidores-turistas na forma de comentários avaliativos destes na condição de usuários do site *TripAdvisor®*. Assim, projetou-se uma ferramenta capaz de extrair os comentários do site e criar meios de recuperar informações relevantes, ou seja, as opiniões dos consumidores-turistas sobre os diversos aspectos que compõem um destino turístico. O desenvolvimento da ferramenta foi viabilizado a partir de uma série de algoritmos e implementações apresentadas em Norvig (2016), Nivre, Hall, & Nilsson (2006), Araújo (2017), Oliveira (2015), dentre outros. Considerando as variáveis relevantes ao Estudo 2, também foi necessário extrair algumas informações referentes aos usuários que fizeram as avaliações, conforme dados apresentados na seção anterior.

Há várias ferramentas (pagas ou gratuitas) no mercado que possibilitam a extração de dados de páginas *web*. No entanto, o presente estudo concentrou-se em informações pontuais sobre os destinos turísticos e sobre os turistas que efetuaram avaliações desses destinos em um site específico, o *TripAdvisor®*. Assim, depois de

utilizar uma ferramenta mais genérica na primeira versão da *SentimentALL* para fazer esse processo (a ferramenta *import.io*²⁴, de extração de dados da *web*), optou-se por criar um módulo na *SentimentALL* para tal fim, que não tinha o aspecto abrangente da *import.io* (ou seja, nesse módulo os *spiders* foram implementados para as páginas específicas do *TripAdvisor*®), mas permitiu uma extração mais pontual dos dados necessários para essa pesquisa. Para que o Módulo de Extração de Dados da *SentimentALL*, no futuro, suporte a extração de dados de outras páginas web, será necessário implementar *spiders* para o novo contexto, ou, pode ser iniciado um projeto que tenha como foco a definição de um modelo de extração de dados que funcione de maneira genérica, o que não é um trabalho trivial.

Há uma complexidade na Análise de Sentimentos que está relacionada ao fato de que os algoritmos, em sua maioria, lidam com a linguagem no nível sintático e léxico, assim, ao considerar analisar a linguagem em seus aspectos funcionais, por exemplo, isso elevaria o processo de análise ao nível pragmático (cf. Liddy, 2001; Briscoe, 2013). Contudo, mesmo que os diversos níveis da análise linguística não tenham sido usados no processo de análise realizado nesse estudo, para o propósito dessa ferramenta, o nível sintático forneceu as saídas necessárias para o trabalho desenvolvido na Tese. Para tanto, o módulo de tratamento textual considerou a análise do aspecto a partir da sua representação na sentença ao utilizar o *Maltparser* (Nivre, Hall, & Nilsson, 2006) para o estabelecimento da dependência sintática entre os termos. Esse algoritmo, juntamente com a anterior identificação das classes gramaticais das palavras com a utilização da biblioteca NLTK® no desenvolvimento do processo de *Pos Tagging* (no trabalho de Hu e Liu (2004) foi utilizado o *NLProcessor*), foi essencial para a realização de uma análise mais precisa das relações entre os termos das sentenças, logo dos aspectos presentes no texto. Assim, foi possível, dentre outras situações, realizar a junção de uma série de aspectos através do elemento “conjunção” relacionando-os a uma mesma polaridade, modificar a polaridade de léxicos que são, normalmente, apontados como positivos em dicionários de sentimentos (por exemplo, o verbo gostar), quando estes estavam acompanhados de uma palavra negativa ou identificar a polaridade de uma palavra opinativa neutra quando havia alguma relação adversativa na sentença.

²⁴ <https://www.import.io/>

Na análise do desempenho da ferramenta foi fundamental separar os comentários em simples e complexos, pois os textos escritos pelos usuários muitas vezes eram de difícil entendimento para pessoas fluentes na língua portuguesa, pois os autores dos comentários, em alguns casos, escreviam sobre características implícitas, misturavam fatos com opiniões, faziam encadeamentos de ideias sem um vínculo entre os termos, etc. Considerando isso, a margem de acerto/erro apresentado na *SentimentALL V2*, cujos testes realizados em relação ao nível de acerto na detecção do aspecto+polaridade nos comentários simples foi superior a 72% e nos comentários complexos foi superior a 56%, viabilizou as análises realizadas (no Estudo 2, apresentado a seguir), que considera a totalidade de aspectos positivos e negativos em cada um dos componentes do produto turístico.

Para uma melhoria nos índices de acerto da ferramenta na identificação dos dois tipos de comentários, há dois trabalhos futuros que podem ser desenvolvidos: um que investigue mais profundamente o sistema *MaltParser*, pois há nove algoritmos de classificação e dois de aprendizagem de máquina disponíveis nesse sistema. Os algoritmos de classificação foram testados em trabalhos relacionados a *SentimentALL*, mas ainda há espaço para mais testes, por exemplo nos algoritmos de aprendizagem de máquina, como o LIBSVM. Como o *MaltParser* utiliza o *corpus* do UD, poderia ser feita uma pesquisa que tratasse do aperfeiçoamento desse *corpus* para a língua portuguesa, pois este ainda carece de algumas informações. Um segundo trabalho pode ser voltado para a parte específica do processo de AS, na definição de léxicos de sentimentos mais condizentes com o contexto do turismo a partir do auxílio de um especialista no domínio. Para a ampliação da ferramenta *SentimentALL*, que permita seu uso na análise de outros produtos ou serviços em outras áreas, como na educação, saúde, etc., um estudo mais abrangente sobre léxicos de sentimentos também se faz necessário, inclusive buscar formas mais dinâmicas de expandir os léxicos, como algumas técnicas discutidas em Qiu, Liu, Bu e Chen (2009, 2011).

No Estudo 1, foram executados dois dos três estados da AS apresentados em Albornoz, Plaza e Gervás, que são a detecção da opinião do autor do comentário (cf. Wiebe, Bruce, & O'hara, 1999; Pang & Lee, 2004; Kim & Hovy, 2004) e o estabelecimento da polaridade (Pang, Lee, & Vaithyanathan, 2002; Turney, 2002; Esuli & Sebastiani, 2006), com o uso dos dicionários de léxicos e das modificações que surgem com a detecção de palavras negativas e adversativas na sentença. A

classificação da intensidade (Esuli & Sebastiani, 2006; Brooke, 2009; Wilson, Wiebe, & Hoffman, 2009), que é o terceiro estado apontado por Albornoz *et al* (2011), não foi tratada nesse estudo, pois as variáveis primárias necessárias ao Estudo 2 tinham relação com quantitativos de opiniões positivas e negativas acerca dos destinos turísticos, não sobre o quão essas opiniões eram mais ou menos positivas ou negativas. Mas, em estudos futuros, a partir da análise de dependência já realizada pela ferramenta será possível identificar, por exemplo, os advérbios presentes no texto (*tag advmod*) e verificar se estes são de intensidade (tendo como base uma relação de advérbios de intensidade, p.ex.: muito, pouco, bastante, extremamente, tudo, nada, quase) e estabelecer, com isso, níveis de intensidade dos adjetivos usados na identificação da polaridade.

Como o modelo de dados da *SentimentALL* permite relacionar os aspectos às sentenças e essas ao comentário, em uma ampliação das potencialidades da ferramenta, será possível explorar a Análise de Sentimentos no nível do documento, como a realizada em Albornoz, Plaza e Gervás (2010). Em um primeiro momento, essa análise pode ser estatística, por exemplo, primeiro fazer a atribuição de pesos aos aspectos presentes na base considerando a proporção entre a quantidade de incidência do aspecto e a quantidade geral de aspectos, armazenando esse resultado em uma tabela denominada “Peso”. Depois, criar uma tabela “Aspecto da Avaliação” para armazenar os aspectos/polaridades de cada avaliação a partir de uma consulta às tabelas Avaliação, Sentença e Análise (apresentadas na Figura 11 da seção anterior). E, por fim, a criação de uma tabela para armazenar a polaridade da avaliação referente a cada comentário, ou seja, uma tabela que armazenará, para cada comentário, o somatório do produto do peso de cada aspecto e sua polaridade. Se o resultado desse somatório for positivo ou negativo, é possível verificar uma tendência da polaridade do comentário. Considerando que na Tabela Avaliação (Figura 11) tem armazenada a nota dada pelo usuário ao destino turístico como um todo por meio da escala de satisfação usada pelo *TripAdvisor*®, é possível verificar o quão a análise global dada pelo usuário é similar à tendência de polaridade do comentário escrito por ele. No entanto, nesse nível de análise, vários testes devem ser realizados para verificar se o modelo de análise estatística proposto estará coerente com o resultado de uma análise manual da polaridade geral em uma amostra de comentários.

A *SentimentALL* fornece um amplo conjunto de dados sobre avaliações do setor do turismo. Ainda que as especificidades em relação aos aspectos presentes nos comentários dos turistas não tenham sido objetos desse estudo, estas permitem uma análise interessante sobre as possibilidades da ferramenta. Por exemplo, os dados apresentados na Tabela 10 correspondem aos 15 aspectos positivos (considerando os três tipos de objetos avaliados: acomodações, restaurantes e atrações) que apareceram em maior quantidade nos comentários sobre os destinos turísticos. Para essa exemplificação, foram considerados apenas os seis destinos turísticos com a maior quantidade de comentários em língua portuguesa no *TripAdvisor*®. Um dos pontos que merece destaque nessa tabela é a ocorrência das palavras em *negrito* (local, localização e lugar) que podem estar relacionadas a um mesmo significado. A ferramenta *SentimentALL* não está preparada para verificar esse tipo de associação. Em Agarwal, Mittal, Bansal e Garg (2015), é proposto um novo modelo de análise de sentimentos baseado em conhecimento extraído de uma ontologia (vide Gruber, 1995; Noy & Mcguinness, 2001; Staab & Studer, 2004; Alalwan, Zedan, & Siewe, 2009; Telnarova, 2010 para uma contextualização sobre o significado, uso e aplicação de ontologia). Para tanto, os autores basearam a ontologia na *ConceptNet*²⁵ para determinar os conceitos, produzir os recursos importantes do domínio e extrair as polaridades dos conceitos usando um léxico de polaridade contextual desenvolvido pela equipe. Considerando o contexto do turismo, criar uma ontologia do domínio que possa resolver situações como essa (a questão de palavras semelhantes terem ou não um mesmo significado) é válida para trazer informações mais precisas sobre os aspectos extraídos.

Tabela 10: Mostra os 15 aspectos positivos com maior quantidade de incidência nos seis destinos turísticos com maior quantidade de aspectos encontrados nos comentários em língua portuguesa.

São Paulo		Rio de Janeiro		Brasília		Fortaleza	
atendimento	122068	atendimento	71678	atendimento	34857	atendimento	32486
ambiente	115684	comida	71657	comida	34017	ambiente	28087
comida	110376	lugar	67463	ambiente	33310	comida	25603
lugar	87838	ambiente	60630	lugar	27313	lugar	22356
restaurante(s)	80220	vista	44127	restaurante(s)	22221	praia	15576
local	42441	restaurante(s)	47017	local	15729	hotel	14008
preço	40151	opção(ões)	31884	hotel	12826	localização	13903
localização	38073	local	29055	preço	12030	local	13696
pratos	37974	preço	28643	pratos	10235	restaurante(s)	17910

²⁵ <http://conceptnet.io/>

Tabela 10: (Continuação) Mostra os 15 aspectos positivos com maior quantidade de incidência nos seis destinos turísticos com maior quantidade de aspectos encontrados nos comentários em língua portuguesa.

hotel	33046	hotel	25352	vista	9429	preço	11903
opção(ões)	46599	localização	24717	localização	9274	café da manhã	10179
qualidade	25270	praia	23047	opção(ões)	13858	funcionários	7469
café da manhã	23665	passeio	20004	café da manhã	8123	opção(ões)	10538
funcionários	17393	pratos	19034	qualidade	6845	pratos	6325
quartos	15728	café da manhã	17554	quartos	5645	qualidade	6318
Belo Horizonte		Natal		Curitiba		Salvador	
atendimento	31526	atendimento	27794	lugar	25821	comida	23493
ambiente	31032	comida	26459	atendimento	17769	atendimento	22585
comida	30134	restaurante(s)	24238	hotel	12868	ambiente	18471
lugar	23723	lugar	20656	ambiente	12226	lugar	18186
restaurante(s)	18016	ambiente	20351	local	11790	restaurante(s)	15137
local	13318	hotel	17615	localização	11335	vista	12260
hotel	11559	praia	15437	comida	11268	hotel	10947
localização	11478	localização	10941	restaurante(s)	10897	local	10408
preço	11166	preço	9574	café da manhã	9799	localização	8768
opção(ões)	13505	café da manhã	9359	opção(ões)	8180	praia	8674
pratos	8023	vista	9210	passeio	7517	preço	8061
café da manhã	7619	funcionários	9185	preço	6532	opção(ões)	7891
qualidade	6719	local	9121	funcionários	5929	café da manhã	7022
funcionários	5357	passeio	8064	quartos	5604	pratos	5138
quartos	4583	pratos	7523	qualidade	3773	funcionários	4777

Para a definição de uma ontologia, conforme Noy e McGuinness (2001, p. 3), é necessário a realização de uma descrição formal de conceitos em um domínio de discurso, ou seja, quais os conceitos que compõe o universo do turismo (por exemplo, acomodação, restaurante), as características e atributos que devem ser descritas por meio das propriedades de cada conceito (atendimento, localização, etc.) e, especialmente, verificar as restrições no domínio (Hotel tem quartos, Restaurantes tem comida, Hotel tem Restaurantes, etc.). Assim, é possível fazer a Análise de Sentimentos de avaliações de Hotéis (por exemplo) por meio de uma ontologia que descreva as várias notações dos conceitos (entidades-classes) e suas propriedades (recursos), para que a máquina possa “entender” que os aspectos/polaridades extraídos dos comentários se referem a certas classes e a algumas das suas propriedades, aumentando, assim, as possibilidades de inferências de informações relevantes ao setor.

Ao considerar os 15 aspectos com maior incidência nesses destinos turísticos com maior quantidade de avaliações (Tabela 10), é possível utilizar uma ontologia para integrar os aspectos/polaridades com maior incidência aos componentes do produto turístico (ACO, RES, ATR), e fazer, ainda, uma relação com as características extraídas do perfil do usuário-turista disponíveis nos sites e outras informações relacionados a localidade (p.ex.: região, estado). Com um propósito semelhante, Yaakub, Li. e Feng (2011) desenvolveram uma ontologia para alguns produtos cujos *reviews* estavam sendo analisados, um deles, mais detalhado no artigo, um celular, foi sistematizado a partir de suas características, as mais gerais (tamanho, tela, bateria, etc.) e as mais técnicas (conectividade, acessórios, aplicativos, etc.). Com essa abordagem, primeiro são identificadas as entidades (aspectos) e opiniões presentes nos *reviews* dos clientes relacionados especificamente a esse produto (celular), depois essas informações são transformadas em uma tabela de fatos dos comentários. No caso do contexto do turismo, é possível fazer uma ontologia bem mais abrangente, considerando as estruturas de componentes do produto turístico presentes na literatura da área (cf., Candela & Figini, 2012; Acerenza, 2002).

Os comentários das 100 cidades analisadas nessa pesquisa correspondem a 71% do total geral de avaliações em língua portuguesa do *TripAdvisor*® no período da extração dos dados (entre os meses de fevereiro a março de 2017). São mais de 6 milhões de comentários analisados, acumulando mais de 22 milhões de aspectos (considerando itens duplicados, ou seja, um mesmo aspecto pode aparecer em vários comentários). As informações contidas no *TripAdvisor*® representam uma parcela dos dados que estão na internet sobre avaliação de destinos turísticos e compõem parte do Big Data desse setor, ou seja, dados que podem ser extraídos e analisados na busca por indicadores do comportamento do consumidor em serviços vinculados ao turismo. Por exemplo, o trabalho apresentado em Miah, Vu, Gammack e McGrath (2016), relaciona informações de uma rede social de compartilhamento de fotos, especialmente de fotos georreferenciadas de uma cidade turística para realizar análises relevantes para a melhoria dos serviços relacionados ao turismo do local, especialmente no que concerne aos passeios e rotas turísticas que são preferidos pelos visitantes.

Para Davenport (2013), existem vários benefícios potenciais no uso do *Big Data* na indústria do turismo. No entanto, isso exige mudanças nas metodologias de análise e nas tecnologias empregadas, por exemplo, qualquer análise que considere

comentários de turistas ou publicações de imagens e vídeos, precisa de um conjunto de ferramentas adequadas para a manipulação desses elementos e, geralmente, essas ferramentas não estão disponíveis gratuitamente no mercado, nem são de fácil uso. Segundo Heerschap et. al. (2014), a prática na indústria do turismo tem mostrado que os métodos de processamento, interpretação e visualização de *Big Data* não são fáceis de implementar, havendo grandes desafios, muito embora os primeiros resultados mostrem-se muito promissores. Uma direção possível é o estreitamento entre academia e mercado, no sentido de realizar pesquisas aplicadas, com investimento em infraestrutura e bolsas de estudo para o desenvolvimento de ferramentas de extração e análise que se adaptam às necessidades do setor do turismo e suas diversas ramificações, considerando que é uma indústria com múltiplas atividades e, por isso, mesmo complexa (Acerenza, 2002; Candela & Figini, 2012). Alguns benefícios potenciais incluem (Davenport, 2013; Heerschap et al, 2014):

- a) melhor apoio à tomada de decisão gerencial, pois haverá informações mais rápidas e mais centradas nas características dos turistas-consumidores, obtidas a partir de ampla variedade de bases de dados.
- b) criação de novos produtos e serviços para turistas-consumidores pelos ofertantes no mercado turístico — por exemplo, os fornecedores de serviços de gerenciamento de viagens poderão desenvolver novos produtos e serviços baseados em dados para indivíduos e corporações, e as agências de viagem poderão criar sistemas de busca e apresentação de produtos e serviços turísticos baseados em contextos, o que virá em significativo auxílio aos turistas-consumidores na experiência de pesquisa prévia e organização da viagem turística.

Com a ferramenta desenvolvida neste trabalho, pode-se obter informações para produzir melhorias nos serviços que são realizados nesse setor, por exemplo, ao verificar através dos comentários sobre os destinos turísticos quais os aspectos foram mais relevantes, tanto positivamente, quanto negativamente, para o turista. Com isso, é possível criar estratégias que torne esses aspectos mais potencializados no local, por exemplo, através de direcionamento de campanhas de *marketing*. Para gestores de serviços de turismo, ter uma síntese de algumas dessas situações, como os dados apresentados na Tabela 10, que mostram que aspectos como “atendimento”, “comida”, “ambiente”, “localização”, “restaurante” estão presentes (em grande

quantidade) nos comentários de todas as 6 cidades com a maior quantidade de avaliações no site, pode confirmar algumas hipóteses já levantadas, ou, com a junção de outros dados, pode tornar possível a inferência de novas conclusões sobre alguns serviços desse setor.

Para Akerkar (2012), os gerentes das empresas da indústria do turismo precisam considerar o que está acontecendo no setor a partir das possibilidades colocadas pelo Big Data e como empresas neste setor podem participar desse processo. Cabe verificar a real necessidade de mais dados na tomada de decisão da empresa, e quais dentre os dados que estão sendo ignorados podem melhorar os processos internos e, especialmente, melhorar a relação com os consumidores. Segundo Akerkar, ao identificar o comportamento de turistas a partir desse variado e gigantesco conjunto de dados e canais, será possível produzir melhorias substanciais nos processos da indústria turística, além de criar expressivo diferencial competitivo para as empresas do setor. As ações dos turistas durante o transcurso de uma viagem, como também suas ações antes da escolha do destino turístico e sua avaliação ao final da viagem, podem fornecer ao setor um diferencial nas estratégias negociais, na inovação tecnológica e na visualização mais personalizada da prospecção econômica nesse setor (cf. Fuchs, Höpken, & Lexhagen, 2014; Song & Liu, 2017, Irudeen & Samaraweera, 2013).

Há algumas vantagens no uso de ferramentas e métodos de análise inovadores voltados para o *Big Data* no setor do turismo em relação às metodologias tradicionais, conforme apresentam Song e Liu (2017): a) confiabilidade, pois o *Big Data* tem como foco as ações reais dos usuários ao invés de questionários de pesquisas feitos por especialistas; b) mais conhecimento sobre o mercado alvo da indústria produzido pelos próprios consumidores (cf. Leng *et al*, 2016); c) possibilidade de obtenção de referências cruzadas com dados de variadas fontes, e d) transmissão de dados em tempo real, o que torna possível o acesso a informações relevantes sobre determinados contextos antes mesmo que as fontes de dados oficiais estejam disponíveis (cf. Song e Liu, 2017). Segundo Shafiee e Ghatari (2016), milhões de registro de dados são produzidos diariamente na indústria turística, processo esse que gera conjunto de dados complexo e desestruturado, que vão desde as transações diárias até as preferências dos clientes dispostos em redes sociais, em sites gerenciadores de turismo, etc. Para Davenport (2013, p. 12), "... como as relações com os clientes historicamente foram fragmentadas em uma variedade de sistemas e

bancos de dados, a agregação de dados deve criar melhores relacionamentos, mais receitas e produtos e serviços melhor orientados” (p. 12).

A junção de tecnologia da informação, computação e dados abundantes sobre as ações de indivíduos proporciona nova visão sobre os interesses e os comportamentos de consumidores, qualquer que seja o setor de atividade econômica. No turismo, na medida em que viagens são moldadas de acordo com o interesse pessoal, os produtos que as empresas direta ou indiretamente vierem a apresentar ao consumidor tenderão a ser melhor recebidos. Para Shafiee e Ghatari (2016, p. 3), nesse sentido “... os dados de um determinado contexto podem fornecer uma visão sobre os padrões comportamentais dos clientes e desenvolver um processo que anteriormente era muito difícil ou impossível de ser feito” (p. 3). Tal processo, especialistas admitem, tem relação com a personalização do atendimento ao consumidor e antecipação dos seus desejos (cf. Akerkar, 2012; Shafiee & Ghatari, 2016, Pries & Dunnigan, 2015).

Song & Liu (2017, p. 16) acrescentam que “... estudos anteriores sobre o turismo se basearam principalmente em pesquisas ou pontos de vista de especialistas, o que significa que eles tomaram amostras da população como um todo, mas não tinham dados reais sobre todos os turistas” (p. 16). Nesse sentido, uma situação é entender o gosto de uma população de consumidores a partir de um questionário respondido por uma amostra de alguns deles, outra é capturar informações dessa população (ou número significativamente maior de indivíduos do que é possível amostrar) nos comentários disponibilizados em sites de viagem, nas fotos e vídeos que postam em redes sociais, etc. Segundo especialistas, dados turísticos assim reunidos darão, potencialmente, relevância maior às inferências acerca dos diferentes perfis de turistas, e desse modo será possível indicar, com grau maior de confiança, viagens que se adequem a cada perfil. Davenport (2013, p. 12) acrescenta ainda que, através de análises preditivas, destinos mais favorecidos, preferências de hospedagem e de restaurantes, necessidades de serviços auxiliares e experiências de turismo poderão ser identificadas para cada viajante e, do mesmo modo, os serviços analíticos on-line, como a previsão de preços e os rankings de desequilíbrio, aumentando a probabilidade de compra. Decorrencia desses esforços já presente nas abordagens mais avançadas de segmentação de consumidores envolve a apresentação on-line de várias formas de publicidade turística, de modo que agências de viagens on-line, agregadores e sites de revisão praticam a segmentação

de anúncios com base no comportamento do consumidor (cf. Pries & Dunnigan, 2015; Davenport, 2013). Cabe registro adicional de que as questões que envolvem a utilização eficiente do potencial do Big Data não são apenas técnicas, como desenvolver ferramentas para extração e análise de dados. Há, também, questões relacionadas, de modo especial, com a mudança dos processos de gestão nas empresas, tanto no nível operacional quanto no plano dos processos de tomada de decisão. É necessário, por exemplo, verificar se os funcionários têm as habilidades necessárias para lidar com o novo contexto e, especialmente, se há espaço para mudança na cultura corporativa estabelecida (cf. Ey, 2014; Akerkar, 2012).

ESTUDO 2: Análise de avaliações *on-line* de destinos turísticos por turistas-consumidores: interpretações a partir do *Behavioral Perspective Model* (BPM).

Avaliar objetos e/ou seus aspectos (atributos, componentes, dimensões, características, etc.) é comportamento. Mais especificamente, em uma perspectiva analítico-comportamental (Skinner, 1938, 1953, 1969; cf. Catania, 1999; Baum, 2006), é comportamento verbal: subcategoria do repertório operante humano, distintivamente diferenciada e mantida por contingências de reforço mediado (sociais; cf. Skinner, 1957). Respostas verbais avaliativas diferenciam-se e são mantidas por contingências de reforço mediado (sociais) e terminam sob controle de estímulos ambientais (verbais ou não-verbais) sinalizadores daquelas contingências. Nesse sentido, o responder avaliativo característico de um indivíduo pode ser interpretado como a expressão verbal resultante da história passada de reforços e/ou punições mediados por outros indivíduos em contextos ambientais sociais onde as avaliações foram emitidas. As contingências de reforço mediado naturais presentes nos variados contextos sociais onde avaliações ocorrem podem ser sinalizadas por estímulos ou arranjos de estímulos com variados graus de funcionalidade e complexa interrelação (de tipo condicional, sequencial, temporalmente estendida, etc.), inclusive, de modo especial, a estimulação verbal produto do responder verbal dos indivíduos interagindo sob aquelas contingências (estímulos-produtos-de-resposta; cf. Skinner, 1957). Assim, em um determinado contexto (por exemplo, um *website*) a estimulação produto da emissão da resposta avaliativa emitida por um indivíduo (ou vários) pode funcionar como estímulo (ou padrão de estímulos) discriminativo para a emissão de respostas de avaliar por outro indivíduo (ou vários outros). Noutra possibilidade, respostas de busca por informação podem ser reforçadas ou punidas pela estimulação produto das emissões avaliativas na medida em que o indivíduo se depara com os registros verbais da emissão gravados na *webpage*. Em se tratando de avaliações de produtos ou serviços em um mercado de relações de consumo definidas, o avaliar — seja como a resposta de um consumidor, seja como fonte de estimulação para consumidores — pode ser considerado como instância da categoria mais geral “comportamento do consumidor”, frequentemente, mas não exclusivamente, observada no pós-compra. Nesse estágio, com base em ter se satisfeito ou não, consideradas as valências e magnitudes das expectativas prévias, o consumidor propagará, respectivamente,

avaliações positivas ou negativas sobre o produto ou serviço (Blythe, 2008; cf. Bunn, 1993; Foxall, 2005)

Na análise do comportamento do consumidor a partir da abordagem do *Behavioral Perspective Model* (BPM; Foxall, 2010; Oliveira-Castro & Foxall, 2005; Nalini, Cardoso, & Cunha, 2013), comportamentos de consumo são estabelecidos por reforços utilitários (consequências no plano do valor de uso e funcionalidade do produto), reforços informativos (consequências no plano das relações sociais decorrentes da aquisição do produto) e consequências aversivas (também no plano do valor de uso e funcionalidade, e da repercussão social). Trabalhos empíricos sobre o comportamento do consumidor com base no BPM apresentam análises do consumo de produtos visando verificar, por exemplo, a relação entre compra e preço do produto (elasticidade da demanda pelo produto), tempo de procura em relação ao nível informativo do produto, dentre outras situações (Sigurdsson, Foxall, & Saevarsson, 2010; Foxall, & Yani-De-Soriano, 2011). Contudo, existem poucos trabalhos fundamentados no BPM onde o comportamento do consumidor é analisado a partir de seus relatos em *websites*, apesar do uso ubíquo desses ambientes na exposição e disseminação de opiniões sobre os mais diversos produtos ou serviços na atualidade. Adicionalmente, inexitem estudos assentados no BPM relativos ao comportamento de avaliar produtos ou serviços via *websites*. A temática parece relevante, dado que o “opinar *on-line*” reflete o dinamismo que caracteriza a produção de conteúdos veiculados na internet sobre produtos ou serviços, tendo o consumidor encontrado nesse canal uma forma direta de expor seu agrado ou sua irritação decorrente de suas relações de consumo (cf. Hong, & Lee, 2008; Harrison-Walker, 2001).

No consumo de produtos complexos, tal como o produto turístico (Candela & Figini, 2012), reforços e punições advêm do contato do consumidor com os vários serviços e produtos que configuram o produto. Elevada frequência de opiniões e relatos avaliativos sobre destinos turísticos, a respeito dos seus atrativos (de lugar, de eventos, etc.), facilidades (de alojamento, de alimentação, de lazer, de exploração da região do destino etc.) e condições de acesso (tipos de transporte), pode ser encontrada em *websites* especializados. Nesse contexto, *sites* gerenciadores de viagens turísticas têm grande quantidade de relatos verbais nos mais diversos idiomas, apresentados em forma de comentários abertos ao público com opiniões relativas a restaurantes, hotéis e atrações de destinos turísticos. A quantidade de

comentários apresentada em postagens avaliativas da experiência de contato com os componentes de produtos turísticos sugere a pertinência da aplicação de técnicas computacionais para extração e análise de tais dados, vista a evidente impossibilidade econômica de realização manual da tarefa e as grandes vantagens em realizá-la.

Análises prospectivas de opiniões têm sido cada vez mais utilizadas em processos de tomada de decisão mercadológica (Vinerean, Cetina, Dumitrescu, & Tichindelean, 2013; Kaiquan, Shaoyi, Jiexun, & Yuxia, 2011). Isso torna a busca do conhecimento sobre o comportamento de consumidores, tal como ocorre em áreas como marketing e gestão, um fator estratégico para as organizações, certamente potencializada quando respaldada metodologicamente por sistemas computacionais confiáveis e, conceitualmente, por teorias da psicologia, como o BPM. Portanto, justificam-se os esforços de análise dos dados gerados por sistemas computacionais para a prospecção de propriedades de interações consumidor-produto em áreas como o turismo a partir de relatos verbais em *websites*. Os textos que definem tais relatos podem ser recuperados e tratados, e as opiniões sobre as propriedades das interações podem ser deles extraídas.

Nesse sentido, os dados das 197 variáveis primárias extraídas pelo sistema *SentimentALL* (Estudo 1) podem ser tratados e dar origem a variáveis secundárias cujas relações podem ser interpretadas com o uso do referencial conceitual do BPM. Admite-se aqui que uma análise assim empreendida é, necessariamente, não exaustiva, terá caráter embrionário, mas pode ilustrar um tipo promissor de integração de conhecimentos da computação e da psicologia, sinalizando potenciais benefícios mútuos para ambas as disciplinas. O posicionamento potencial de conhecimentos analítico-comportamentais na exploração e análise de dados da natureza dos que foram coletados no Estudo 1 pode contribuir de modo inovador como subsídios a processos de tomada de decisão inerentes às relações de consumo envolvendo produtos complexos (como o turístico) em mercados estruturados de grandes dimensões (como a indústria turística). Consideradas as informações já apresentadas anteriormente sobre turismo e o conceito de produto turístico, nas próximas seções, de modo a fundamentar o presente estudo (Estudo 2), serão apresentados alguns estudos com dados oriundos de avaliações *on-line* de produtos turísticos e os conceitos básicos do referencial teórico-conceitual do BPM aplicado à análise do comportamento do consumidor.

Avaliação de produtos turísticos em ambientes *on-line*

O impacto do uso de *websites* sobre a atividade turística e a consequente explosão de conteúdos gerados pelos usuários desses *websites* tem se constituído em temas de estudo de vários pesquisadores mundo afora. Miguéns, Baggio e Costa (2008) conduziram pesquisa com recurso ao *TripAdvisor*®, considerando apenas hotéis ($n = 191$) da cidade de Lisboa, Portugal. No estudo, além dos comentários sobre os hotéis, foram analisadas declarações de turistas-consumidores nos fóruns de discussão sobre atrações turísticas da localidade. O estudo indicou que as escolhas dos turistas-consumidores não se baseiam apenas nas categorizações ordenadas disponíveis na plataforma, mas, também, nas informações obtidas em salas de discussão onde questões específicas sobre a localidade são formuladas. Nos fóruns, os turistas-consumidores revelam-se interessados em vários aspectos do destino, com destaque para os transportes, os restaurantes e os preços.

A forma como as condições de hospedagem, alimentação e atrações são avaliadas em *sites* como o *TripAdvisor*® representa novo e complexo desafio para os profissionais da área do turismo, já que comentários e discussões podem influenciar, de forma contundente, a imagem de um destino turístico para viajantes. Isso pode repercutir na definição de novos modelos de negócio, particularmente aqueles que consideram a influência dos ambientes *on-line* no comportamento do consumidor (Constantinides, 2004; Paraskevas, 2005). Tal possibilidade, conjuntamente ao modo como, particularmente, a imagem dos hotéis no destino é impactada (Miguéns, Baggio, & Costa, 2008), atribuem peso diferenciado aos canais criados pelas tecnologias da informação para o entendimento de como produtos complexos (tal como o turístico), se “comportam”.

Limberger, Anjos, Meiria e Anjos (2014) realizaram pesquisa para analisar o nível de satisfação com hotéis a partir de avaliações feitas por turistas-consumidores usuários do *TripAdvisor*®. Para tanto, analisaram a correlação entre o nível de satisfação geral (valor de 1 a 5 atribuído pelo usuário ao hotel) e pontos atribuídos a alguns aspectos específicos por turistas-consumidores (por exemplo, custo-benefício, localização, qualidade de sono, quartos, limpeza e serviço). A análise considerou 660 avaliações de hotéis (classificados de 3 a 5 estrelas), tendo sido considerados no estudo apenas os comentários que possuíam a avaliação dos critérios específicos. Com o uso de análise multivariada de correspondência múltipla, os dados mostraram

que as variáveis quarto, serviço e custo-benefício mantiveram as maiores correlações com o nível de satisfação geral, resultado que corrobora com aqueles obtidos por Stringam, Gerdes e Vanleeuwen (2010) e Jeong e Jeon (2008).

Simms (2012) realizou análise sobre a importância do conteúdo gerado por usuários em sistemas *on-line* no planejamento das férias. A análise teve como foco o papel que algumas características da viagem (por exemplo, familiaridade com o destino, tempo de viagem, localização do destino) desempenham no nível de importância que turistas-consumidores dão aos conteúdos produzidos em *sites* sobre destinos turísticos. Para a realização do trabalho, foi utilizado o relatório anual “*The Use of Social Media by American Travelers*”, que continha entrevistas realizadas em janeiro de 2011 pelo Laboratório Nacional de Turismo e *e-Commerce* da Universidade de Temple (Pensilvânia, EUA). Amostra de viajantes americanos que usaram a internet para planejar a viagem responderam a seguinte questão: “*Para a sua mais recente viagem, quanto do planejamento foi realizado usando a internet?*”. Resposta a essa consistia em indicar uma das seguintes opções: 100%, 75% a 99%, 50% a 74%, 25% a 49% e 1% a 24%. Outra variável dependente do estudo foi obtida a partir da seguinte seguinte pergunta: “*No planejamento da viagem mais recente, você olhou (ou leu) comentários ou materiais (por exemplo, fotos, vídeo) em publicações on-line feitas por outros viajantes?*”. Estatísticas descritivas (frequências e tabulações cruzadas) foram usadas para analisar as diferenças entre os viajantes que haviam buscado informações *on-line* de outros viajantes para o planejamento da viagem e os que não haviam buscado, assim como as características das viagens em cada uma das situações. Simms constatou que mais de 57,4% dos viajantes consideraram os comentários/materiais postados *on-line*, e os comentários e as fotos das viagens foram os conteúdos indicados como os mais relevantes. Outro ponto observado foi que existem diferenças sociodemográficas (idade, níveis de renda e etnia) entre os viajantes que utilizam informações geradas por usuários *on-line* para o planejamento de suas férias e aqueles que não utilizam. Viajantes mais novos são mais propensos (66,4%) a utilizar essas informações, assim como os que tem melhor renda. Em relação a etnia, os entrevistados asiáticos mostraram-se os mais propensos a recorrer a informações *on-line* de outros viajantes (76,0%), seguido pelos afro-americanos (71,8%) e, por último, os caucasianos (54,1%). Para Simms (2012), variáveis comportamentais de consumo (por exemplo, experiência de viagem, uso do computador), também desempenharam papel importante no envolvimento dos

viajantes com o conteúdo gerado *on-line* no planejamento de viagens. E, por fim, envolvimento com conteúdo *on-line* sobre viagens parece ter relação com as características da viagem (ida a um destino turístico pela primeira vez, visita a um destino internacional, se viajou sozinho, etc.).

Miah, Vu, Gammack e McGrath (2016) fizeram pesquisa, no contexto do turismo, com análise de fotos publicadas no *site Flickr*²⁶. O trabalho foi realizado com base na abordagem *Design Science Research* (DSR), que envolve os processos de construir, investigar, validar e avaliar artefatos. Miah *et al.* (2016), a partir de definição do espaço computacional do problema, especificaram, implementaram e avaliaram artefato de solução em TI baseado em DSR. Em seguida, desenvolveram método para coletar, classificar, localizar e identificar informações turísticas significativas para apoiar a tomada de decisão estratégica de organizações de gerência de destinos turísticos. As fotos utilizadas no estudo foram tiradas por turistas usuários do *Flickr* ao longo do percurso de viagem usando dispositivos de captura de fotos habilitados com GPS (que grava automaticamente informações geográficas). Assim, as fotos *geotaggeadas* foram disponibilizadas no *site*. Além dos dados geográficos, os pesquisadores utilizaram informações temporais, como data e hora, e informações textuais. No processamento dos dados, foram usadas quatro técnicas: **1) processamento de metadados textuais**, para analisar as palavras-chave usadas na publicação das fotos de forma a possibilitar o entendimento de certos objetos de interesse dos turistas; **2) agrupamento de dados geográficos**, para a verificação das áreas de alta atividade fotográfica a partir de informações sobre o número de turistas enviando fotos para o *site* e o número de fotos carregadas, o que tornou possível encontrar *clusters* de locais turísticos populares; **3) identificação das fotos representativas**, para verificar que assunto apareciam mais frequentemente na postagem das fotos, e **4) modelagem de séries temporais**, com o objetivo prever a demanda turística futura no locais fotografados e revelar padrões de viagens sazonais visando planejamento e tomada de decisão estratégica.

Miah *et al.* (2016) combinaram as quatro técnicas no artefato para processar e analisar diferentes tipos de dados (por exemplo, *tags* de texto, *tags* geográficas, conteúdo de fotos e períodos de tempo) com o propósito de fornecer informações mais

²⁶ <https://www.flickr.com/>; plataforma de imagens que abriga amplo volume e variedade de dados enquadráveis no que está sendo denominado *Big Data* na literatura relacionada a área (CHEN; MAO; LIU, 2014).

consistentes e abrangentes sobre o comportamento e as percepções dos turistas. Na verificação da eficácia do artefato de solução proposto, Melbourne, Austrália, foi escolhida como destino turístico, visto atender mercados nacionais e internacionais, e ter recebido 2,4 milhões de visitantes internacionais entre o início de 2015 e março de 2016, e 8,4 milhões de visitantes domésticos em 2015. A origem geográfica de cada usuário foi recuperada do *Flickr*® a partir do identificador único de usuário na rede. Com esses dados, os turistas locais (residentes em Melbourne), os turistas domésticos (residentes em outros locais da Austrália) e os turistas internacionais foram definidos, tendo sido agrupados com base no seu continente natal. Visto que informar a localização não é obrigatório para criar conta no *Flickr*®, os dados de muitos usuários não foram incluídos no estudo, sendo analisados, ao final, os dados de 2.550 turistas com localização de residência identificada. Segundo Miah *et al*, os resultados mostraram que, a partir da solução desenvolvida, foi possível detectar os principais padrões e tendências de comportamento no destino turístico, com o incremento de detalhes relevantes para contribuir no processo de tomada de decisão estratégica por organizações de gerenciamento de destinos turísticos, tal como possibilitar aos gestores o desenvolvimento de ações de *marketing* direcionado — como, por exemplo, apresentar detalhes da arquitetura da cidade ou do jardim botânico, na medida em que turistas visitam determinadas áreas da localidade.

Em Miguéns, Baggio e Costa (2008), Limberger, Anjos, Meiria e Anjos (2014) e Miah, Vu, Gammack e McGrath (2016), os dados obtidos foram analisados a partir de categorias já previamente estruturadas em *websites* (o *TripAdvisor*® nos dois primeiros e o *Flickr*® no terceiro) e em Simms (2012) os dados foram obtidos a partir de entrevistas. No presente estudo, a categorização do *TripAdvisor*® também foi a referência para o agrupamento das avaliações feitas sobre os destinos turísticos considerados, não tendo sido desenvolvido qualquer instrumento não computacional de coleta de dados, nem adotada qualquer categorização prévia de aspectos do produto turístico eventualmente disponível em outros *websites*.

Análise do comportamento do consumidor: o *Behavioral Perspective Model* (BPM)

O comportamento de indivíduos nas interações caracterizadas como de consumo é influenciado por fatores ambientais e intraindividuais. A área

multidisciplinar do conhecimento denominada “Comportamento do Consumidor” ocupa-se, segundo Solomon (2008, p. 27), do estudo dos “... *processos envolvidos quando indivíduos ou grupos selecionam, compram, usam ou descartam produtos, serviços, ideias ou experiências para satisfazer necessidades e desejos*” (p. 27). Considerado, por exemplo, o contexto do *Marketing*, o conceito de consumidor vai além dos atos de compra em si, pois agrega atividades relacionadas a comportamentos pré-compra e pós-compra. Os comportamentos pré-compra estão relacionados às necessidades e desejos, às quais se segue a procura por produtos, serviços e/ou marcas. No consumo turístico, por exemplo, consultas a avaliações *on-line* feitas por turistas-consumidores sobre aspectos relacionados a restaurantes, hotéis e atrações de localidades turísticas podem ser fator relevante para que um determinado viajante escolha um determinado destino turístico para visitar. Os comportamentos pós-compra incluem o uso e a avaliação do item comprado, o que implica na geração de informações que podem ser (e, de fato, quase sempre são) repassadas verbalmente a outros consumidores. Isso fica evidenciado na definição de “comportamento do consumidor” apresentada por Blackwell, Miniard e Engel (2000, p. 15):

“O conjunto das atividades diretamente envolvidas em obter, fazer uso e dispor de produtos (marcas) e serviços, incluindo os processos de tomada de decisão que antecedem e sucedem estas ações” (Blackwell, Miniard, & Engel, 2000, p. 15).

Nos estudos em ciências da computação voltados ao desenvolvimento de tecnologias aplicáveis à descrição e análise de relações de consumo — tais como os mencionados no Estudo 1 referentes ao uso de plataformas que gerenciam *Big Data* e exploram algoritmos de análise de sentimento e mineração de opinião —, o conceito de comportamento do consumidor tende a ser adotado e explorado sem qualquer compromisso evidente com qualquer abordagem teórica do consumo (econômica, psicológica ou de qualquer outro tipo). O esforço é tecnológico, se dá em atendimento a demandas do mundo dos negócios, e pode ser melhor caracterizado como pragmático, no sentido de promover descrições mais abrangentes e precisas dos interesses e padrões de ação de consumidores com o objetivo de subsidiar e tornar cada vez mais eficiente a gestão mercadológica de produtos e serviços e, com isso, aumentar a satisfação de consumidores, a demanda e a lucratividade das organizações (cf. Davenport, 2013). Nos estudos computacionais, os fatos

comportamentais configurados como comportamentos de consumo são abordados quase que fenomenologicamente, descritos e conceituados com o uso de termos e noções no nível da linguagem comum, sem a inserção sistemática dos mesmos em quadros teóricos compreensivos onde a natureza, a origem, os determinantes, os níveis de análise, as correlações, as relações funcionais, etc., sejam discutidos e considerados em articulações com outros eventos, conceitos ou constructos (cf. Jaccard & Jacoby, 2010). Considerado esse cenário e admitindo ser possível a aplicação do referencial conceitual da análise do comportamento (cf. Skinner, 1938, 1953, 1957, 1969; Catania, 1999) como arcabouço para interpretar as evidências empíricas sobre o comportamento de consumo turístico extraídas da *World Wide Web* e tratadas por método computacional baseado no algoritmo de análise de sentimento (AS), o modelo analítico-comportamental para análise do comportamento do consumidor denominado *Behavioral Perspective Model* (BPM) é aqui explorado.

De acordo com Nalini, Cardoso e Cunha (2013), modelos comportamentais de análise do comportamento do consumidor dão relevo às interações (históricas e atuais) do indivíduo consumidor com o ambiente (passado e presente) de consumo como fundamento para a descrição, análise e explicação do consumir, esse admitido como processo comportamental aprendido de acordo com princípios de aprendizagem (cf. Catania, 1999). Modelo teórico analítico-comportamental, que concebe o comportamento do consumidor como comportamento operante (cf. Skinner, 1938, 1953; Catania, 1999) e cresce em importância na compreensão de processos de consumo humanos, é o *Behavioral Perspective Model* (BPM), proposta essa que, segundo Nalini, Cardoso e Cunha (2013), vem sendo formalmente desenvolvida desde o início dos anos 90 por Foxall (vide Foxall, 1990, 1993, 1996, 2001, 2002, 2005, 2010, 2016a, 2016b para aprofundamento sobre o desenvolvimento conceitual e pesquisa atual fundamentada no BPM; cf. Oliveira-Castro & Foxall, 2005; Nalini, Cardoso, & Cunha, 2013).

O *Behavioral Perspective Model* (BPM; somente a abreviação usada doravante) assenta-se teoricamente em noções do marketing e da análise do comportamento e, na explicação do consumir, dá ênfase às variáveis do contexto ambiental onde ocorrem as relações de consumo, a saber, genericamente, as contingências ambientais sob as quais o consumidor age em uma determinada situação de consumo e a história de aprendizagem do consumidor. Segundo Foxall (1997), o comportamento do consumidor resulta da interação entre duas grandes

tendências opostas: 1) a emissão de respostas de aproximação aos produtos ou serviços ofertados no mercado como função do contato do consumidor com consequências benéficas (tecnicamente, “consequências reforçadoras”; por exemplo, ganhos em eficiência, produtividade, conforto e imagem social) resultantes da aquisição e uso dos bens adquiridos, e 2) a emissão de respostas de afastamento ou evitação dos produtos ou serviços em função da ocorrência de consequências potencialmente malélicas (tecnicamente, “consequências punitivas”; por exemplo, dificuldades no uso, perda de recursos financeiros e aumento de restrições orçamentárias; cf. Alhadeff, 1982).

Em uma representação esquemática geral do BPM, adaptada de Foxall (2010, p. 9) e extraída de Nalini, Cardoso e Cunha (2013), a Figura 1 mostra que o comportamento do consumidor é função de variáveis que antecedem à sua ocorrência (variáveis antecedentes), assim como de variáveis consequentes à sua ocorrência (variáveis consequentes).



Figura 1: Representação esquemática do Behavioral Perspective Model (BPM; adaptada de Foxall, 2010, p. 9; extraída de Nalini, Cardoso, & Cunha, 2013, p. 494).

As variáveis antecedentes são o *cenário do comportamento do consumidor* e a *história passada de aprendizagem do consumidor*, e as variáveis consequentes são os *reforços utilitário e informativo*, e as *punições utilitária e informativa*. Pode ser observado que a estrutura do modelo se assenta no conceito analítico-comportamental de contingência tríplice (reforçadora ou punitiva; Skinner, 1953, 1969; cf. Todorov, 1985; Catania, 1999) e, nesse sentido, pode ser interpretada como uma aplicação do conceito a análise do consumir. Cabe registro de que, em análise do comportamento, os elementos das contingências são descritos em linguagem

extensional, o que, nesse contexto, significa dizer que tais elementos realizam seus papéis independentemente de vontades, crenças, intenções, atitudes, etc., de uma agência causal intraindividual — portanto, em sentido distinto daquele em que esses termos são conceituados em abordagens mentalistas (cf. Foxall, 2004).

Várias descrições do BPM, mais ou menos sintéticas, podem ser encontradas na literatura, tanto original (Foxall, 1990, 1993, 1996, 2001, 2002, 2005, 2010; vide Foxall, 2016a e 2016b para atualizações recentes) quanto derivada (Oliveira-Castro & Foxall, 2005; Nalini, Cardoso, & Cunha, 2013; e dezenas de artigos, livros, capítulos de livros, dissertações de mestrado e teses de doutorado elaborados em várias partes do mundo, incluindo o Brasil). Parece razoável afirmar que o BPM é um modelo em construção, porém com alicerces já firmemente erigidos. Nesse sentido, considerando as várias descrições existentes, configurações conceituais básicas das variáveis que estruturam o modelo podem ser descritas e exemplificadas do seguinte modo:

1) Cenário do comportamento do consumidor consiste no conjunto de variáveis ambientais contextuais moduladoras da função dos estímulos sinalizadores (discriminativos em graus variados de “força”) das consequências reforçadoras ou punitivas contingentes à emissão das respostas de consumo. Os estímulos que constituem o cenário podem ser: a) **físicos**; por exemplo, informações em um *website*; número de hotéis e/ou restaurantes em uma localidade turística; propagandas em um ponto de venda; o arranjo dos produtos em uma prateleira, a logomarca da loja, música ambiente, etc.; b) **sociais**; por exemplo, o fluxo (maior o menor) de pessoas nos atrativos de uma localidade turística; a presença de celebridades em uma promoção; a quantidade de atendentes e/ou vendedores em uma loja; a quantidade e o tipo de acompanhantes do consumidor, etc.; c) **temporais**; por exemplo, o período disponível para uma viagem; a quantidade disponível e as durações (maiores ou menores) de planos de voo para uma localidade turística; o período de tempo em que uma loja fica aberta, a duração de uma oferta; os períodos definidos por datas comemorativas, etc., ou d) **regulatórios**; por exemplo, regras de conduta comercial na loja ou *website* de *e-commerce* formuladas pela gerência do negócio ou auto formuladas pelo consumidor, especificadoras de contingências. A conjugação integrada dos estímulos das quatro categorias define o cenário do consumidor, que varia em graus de abertura (em um *continuum* aberto–fechado dependendo do número [maior ou menor,

respectivamente] de alternativas de consumo e do nível [maior ou menor, respectivamente] de restrições regulatórias ou topográficas sobre as respostas de consumo do consumidor). Em uma ocasião de consumo específica, a saliência discriminativa maior ou menor de um determinado estímulo do cenário ocorre como legado da história passada de aprendizagem do consumir, definindo o conceito *situação do consumidor* (vide Foxall, 2010, para distinção entre *cenário* e *situação do consumidor*, distinção essa não representada na Figura 1).

2) História de aprendizagem do consumidor é a noção que visa considerar as variáveis resultantes de processos históricos de seleção dos padrões de resposta de consumir idiossincráticos (repertório de respostas de consumo aprendido), decorrentes de experiências de exposição direta a contingências operantes de reforço ou punição, ou a formulações verbais dessas ao longo do desenvolvimento ontogenético do indivíduo. No BPM, portanto, idiossincrasias do repertório de consumo do indivíduo (por exemplo, preferências por certas localidades turísticas, hotéis ou restaurantes; tendência a buscar se inscrever e fazer uso de programas de milhagem; capacidade maior ou menor de planejamento dos gastos com a viagem; frequência e modo de uso dos produtos ou serviços adquiridos, etc.) resultam de processos de aprendizagem que transcorreram ao longo da história de vida do indivíduo em circunstâncias de consumo.

3) Reforço utilitário se refere a eventos consequentes contingentes à emissão de respostas de consumo que revelam afetar a ocorrência futura dessas respostas aumentando-as em frequência, e são classificáveis como ganhos de utilidade, ou seja, tem valor de uso para o consumidor frente a situações de resolução de problemas práticos. Reforço utilitário advêm das propriedades funcionais de um produto ou serviço adquirido e são mediados ou viabilizados por essas propriedades na medida em que o consumidor faz uso do produto ou serviço após adquiri-lo (ou obter permissão para usá-lo). Por exemplo, teoricamente, um turista-consumidor tem seu comportamento de consumo turístico obtendo consequências reforçadoras utilitárias quando, tendo adquirido um pacote de viagem e se deslocado até o destino turístico, recebe atendimento que avalia como de alta qualidade desde o início do voo até o último dia de estadia no hotel em que se hospeda; conhece e se alimenta com

iguarias que avalia como especiais e muito saborosas nos restaurantes que vem a conhecer; vive experiências avaliadas como interessantes, enriquecedoras, divertidas, relaxantes e percebidas como esteticamente belas ao visitar os atrativos culturais e ecológicos do destino, etc.

4) *Punição utilitária* se refere a eventos consequentes contingentes ao consumir também classificáveis no âmbito da utilidade, mas que auferem prejuízos e/ou perdas ao consumidor e, assim, acarretam na diminuição da probabilidade futura de ocorrência das respostas de consumo que os produziram. Considerado o exemplo sobre consumo turístico acima descrito, o turista-consumidor teria, teoricamente, seu comportamento obtendo consequências punitivas utilitárias caso, fazendo e passando por todas as situações descritas, recebesse atendimento, alimentos, e vivesse experiências que terminassem avaliados como insatisfatórios, ruins, decepcionantes, desgastantes, ineficientes, amedrontadores, etc.

5) *Reforço informativo* se refere a eventos consequentes contingentes a emissão de respostas de consumo — que, portanto, implicam no aumento da probabilidade de ocorrência futura das respostas que os produziram — e advêm das interações sociais que o consumidor trava com pessoas os seus grupos de referência em torno do produto ou serviço adquirido (ou aspectos da experiência de consumo), quando, então, ocorre elevação da imagem social do consumidor vista a aquisição do bem, devido a pelo menos um dos seus atributos, avaliado socialmente como importante, desejável ou valioso e/ou símbolo de *status*. Por exemplo, ao adquirir um pacote de viagem e gozar férias em determinado destino turístico (digamos, Dubai, nos Emirados Árabes Unidos), reforço informativo do comportamento turístico do consumidor ocorre quando, em publicizações de indícios da viagem (fotos em redes sociais, por exemplo) e/ou relatos ou comentários sobre a viagem para seus colegas de trabalho, o turista-consumidor obtêm destacada atenção positiva, na forma de elogios, enaltecimentos, parabenizações, apreciações prestigiosas (sugestivas de força, poder, especialidade, diferenciação), etc. A ocorrência de reforço informativo depende e é mediada por comportamentos de outras pessoas (natureza social), o que configura as interações sob as quais o conceito pode ser aplicado como episódios verbais, assim como é classificável como verbal a estimulação reforçadora que alcança o consumidor (cf. Skinner, 1957).

6) *Punição informativa*, por sua vez, se refere a eventos consequentes contingentes ao consumir também classificáveis no âmbito social, visto serem consequências mediadas por outras pessoas (verbais), mas que implicam na diminuição da probabilidade de ocorrência futura das respostas que os produziram. Considerado o exemplo de consumo turístico acima descrito, punição informativa ocorre quando, ao invés de receber atenção positiva (reforço informativo) de outros indivíduos pelas férias em Dubai, o turista-consumidor recebe atenção negativa, na forma de depreciações, atitudes invejosas, isolamento, desqualificações, questionamentos negativos da escolha feita visto sua condição financeira e a maior importância do atendimento a outras necessidades, imagem de esnobe, etc. Do mesmo modo que o reforço informativo, a punição informativa pode ser considerada um processo verbal de natureza social, assim como é verbal a estimulação punitiva que alcança o consumidor (cf. Skinner, 1957).

Em uma descrição genérica da dinâmica entre as variáveis do BPM, encontram-se as interrelações entre variáveis ambientais e comportamentais preconizadas pela noção de contingência tríplice, com o acréscimo de certas especificidades. O modelo admite que o consumo (procura, inspeção, escolha, compra, uso, avaliação, descarte) de um produto ou serviço é a resultante da confluência (representada pelo símbolo XX Figura 1) de variáveis do cenário (ambientais) e da história de aprendizagem do consumidor (repertório de consumo), confluência essa que potencializa relativamente funções discriminativas de estímulos do cenário e define a situação de consumo. Na situação de consumo, a emissão de respostas de consumo tem como consequências ambientais eventos reforçadores e punitivos, utilitários e informativos, conjuntamente, em graus variados. Decorrem, como efeitos combinados das consequências, a seleção, o fortalecimento e/ou a manutenção (reforço), ou enfraquecimento ou supressão (punição) das respostas (ações e verbalizações) de consumo, assim como se estabelecem ou se fortalecem funções discriminativas (em graus variados de força) dos estímulos (não-verbais ou verbais) componentes do cenário de consumo (cf. Foxall, 1996; 2010). Nesse sentido, por exemplo, a quantidade relativa de consequências reforçadoras utilitárias de um determinado produto ou serviço pode repercutir nas avaliações sociais de utilidade do mesmo para determinados segmentos consumidores, assim como, em relação às consequências informativas, poderá haver repercussão diferencial nas imagens

sociais dos consumidores. Assim, de acordo com o BPM, condições de estímulo que configuram socialmente produtos, serviços e/ou marcas, tais como, por exemplo, o nome de um destino turístico; o padrão das propagandas que anunciam esse destino, onde ocorrem associados o brasão do Estado onde o destino se localiza e um *jingle* característico; *souvenirs* com itens culturais típicos do destino representados graficamente ou miniaturizados, etc., podem se tornar estímulos discriminativos sinalizadores dos níveis de reforço utilitário e de reforço informativo que o destino pode, potencialmente, oferecer. Cardoso (2012) demonstrou ser possível, com base em avaliações dos componentes do produto turístico, determinar empiricamente os níveis reforçadores utilitário e informativo de 22 destinos turísticos litorâneos brasileiros. Os índices quantitativos obtidos para ambas as categorias de reforço para cada destino puderam ser ordenados de modo consistente e agrupados em três conjuntos de destinos com índices estatisticamente diferentes, agrupamentos esses interpretáveis como indicativos de nível reforçador potencial alto, intermediário ou baixo.

Adicionalmente, três estudos empíricos fundamentados no BPM tocam, ainda que remotamente, aspectos relevantes da presente proposta. Sigurdsson, Menon, Sigurdarson, Kristjansson e Foxall (2013) realizaram experimento baseado nos princípios do BPM no contexto do marketing via *e-mail* com o objetivo de investigar a escolha do consumidor nesse canal. E-mails foram enviados para dois grupos de pessoas de um mesmo banco de dados de consumidores interessados em livros infantis e que tinham dado prévia autorização para receber e-mails sobre o assunto. Um primeiro e-mail foi enviado a amostra de 7.265 pessoas do Grupo 1 e 7.227 pessoas do Grupo 2, e um segundo a 6.532 pessoas do Grupo 1 e 6.508 pessoas do Grupo 2. O produto alvo no experimento eram livros com histórias em quadrinhos do Pato Donald (Disney Inc.®). Na verificação das ações dos consumidores participantes do experimento, foi usado *software* que armazenava as seguintes informações de cada consumidor: **1)** se os *e-mails* haviam sido recebidos e abertos, **2)** se o consumidor havia clicado no *link* que fora encaminhado no *e-mail* e acessado a loja virtual, e **3)** a quantidade de livros-alvo vendida. Além disso, foram monitorados **4)** o número de cliques em outros *links* (além daqueles dos livros-alvo) na loja virtual e **5)** se os consumidores haviam cancelado sua participação na pesquisa depois de receber um *e-mail*. Os delineamentos experimentais podem ser representados por A–B–A–C–A e A–C–A–B–A, de modo a alterar a ordem de exposição às condições B e

C. A condição B foi definida pelo envio de *e-mails* com teor utilitário (econômico/funcional) e a condição C pelo envio de *e-mails* com teor informativo (social). As análises dos dados obtidos apresentados mostraram que os estímulos informativos (mensagem no assunto do *e-mail*: “*Você compra, nós damos*”) foram mais eficientes para induzir os consumidores a abrir os *e-mails* comparativamente aos estímulos utilitários (mensagem no assunto do *e-mail*: “*2 por 1*”), que foram mais eficazes no aumento das respostas de compra. Para Sigurdsson *et al.* (2013):

De acordo com o BPM, o consumo de produtos e serviços que estão na categoria entretenimento, tais como livros infantis, tendem a pertencer ao contexto relativamente aberto, com nível utilitário alto, mas com baixo nível de reforço informativo. Logo a interpretação foi confirmada experimentalmente com os dados obtidos nesse experimento. (Sigurdsson, *et al.*, 2013).

No experimento realizado por Alshurideh (2014), foi analisado o comportamento de retenção de clientes de operadoras de telefonia móvel do Reino Unido. Três grupos de usuários, cada grupo composto de 5 a 7 participantes, com variada experiência nesse tipo de negócio, tiveram as suas discussões, através da condução da pesquisadora, gravadas, transcritas, codificadas e analisadas. As perguntas levantadas nas discussões foram embasadas no BPM e a codificação das afirmações dos participantes foi realizada conforme exemplo a seguir:

A afirmação “*É bom ter um lugar, um meio, uma pessoa que você pode falar sobre diferentes tipos de ofertas*”, foi codificado como (*physical-place+*, *physical-employee+*) de acordo com a influência positiva que advém do fato da operadora ter um local físico para atendimento e da interação dos funcionários com os clientes. A declaração “*É inútil uma operadora de celulares ter uma loja física*”, foi codificado como (*physical-place-*), devido à influência negativa da disponibilidade de uma loja física. Os textos foram transcritos seguindo esse método para capturar com precisão todos os efeitos positivos e negativos sobre todos os elementos de BPM. (Alshurideh, 2014, p. 113).

Com base no BPM, Alshurideh (2014) postulou algumas proposições, como “*Quanto maior a quantidade de reforços utilitários fornecidos pela operadora móvel contratada, maior o efeito positivo sobre o cliente*” (p. 114), e com relação à natureza das propriedades informativas punitivas recebidas por um assinante durante período específico, “*Quanto menor a quantidade de propriedades informativas punitivas recebida, maior o efeito positivo sobre o comportamento de retenção do cliente (incluindo a renovação e atualização do contrato)*” (p. 114). Considerando essas especificações, a autora concluiu que os principais resultados mostraram que os

reforços utilitários e informativos, o escopo do cenário e a história de aprendizagem do indivíduo afetam a repetição do comportamento do consumidor positivamente, enquanto as propriedades punitivas utilitária e informativa afetam a repetição do comportamento negativamente

Cardoso (2012), naquele que talvez seja o único estudo sobre consumo turístico fundamentado no BPM, realizou pesquisa empírica visando a determinação dos níveis de reforço utilitário e informativo de 22 destinos turísticos litorâneos brasileiros. Duzentos (200) indivíduos de ambos os sexos, com idade entre 18 e 52 anos avaliaram componentes dos produtos turísticos (conforme Acerenza, 2002) por meio de vídeos de um minuto sobre cada destino turístico, estruturado a partir de aspectos definidos em Acerenza (2002), que tem relação com transporte, hospedagem, alimentação, lazer e atrativos. Variáveis dependentes foram medidas com escalas *Likert*: nível de importância e relevância de cada componente do produto turístico, o grau de conhecimento do destino e, com recurso aos vídeos, questões relacionadas às propriedades reforçadoras utilitárias e informativas localizáveis no destino ou prováveis a partir de uma visita ao destino. Cálculos envolvendo medidas da experiência turística geral do participante e das propriedades utilitárias e informativas de cada destino permitiram a definição empírica quantitativa do que, no estudo, foi denominado “nível (magnitude) reforçador utilitário” e “nível (magnitude) reforçador informativo”. Os índices quantitativos obtidos para ambas as categorias de reforço para cada destino puderam ser ordenados de modo consistente e agrupados (via análise de agrupamentos, método *K-means*) em três conjuntos de destinos com índices estatisticamente diferentes (técnica ANOVA para medidas repetidas), agrupamentos esses interpretáveis como indicativos de magnitudes reforçadoras potenciais altas, intermediárias ou baixas.

No esforço de montagem de um quadro descritivo-explicativo abrangente dos processos comportamentais de consumo (tal como verificados em pessoas de carne e osso, de fato), o BPM tem fundamentado a produção de conhecimento em três vertentes: **1)** a teoria da escolha do consumidor, que considera as relações entre o responder de consumo e valores diferenciais dos parâmetros do reforço e da punição; **2)** a análise, em termos operantes, da tomada de decisão do consumidor e do gerenciamento mercadológico, e **3)** interpretação do *marketing* como interações recíprocas entre consumidores (demandantes) e arranjos empresariais (ofertantes), modeladas e mantidas pelas consequências recíprocas que produzem (cf. Foxall,

1996, 1997, 2010; Evans, Jamal & Foxall, 2006). O modelo, contudo, é uma elaboração em desenvolvimento que, segundo Foxall (2010, p. 2) vem transcorrendo em cinco fases contínuas distintas, porém sobrepostas, sobre as quais assim se referiu o cientista britânico:

The BPM research program has five conceptually distinct but overlapping and continuing phases: (i) *conceptual*: a period of critical analysis of the prevailing cognitive paradigm from the standpoint of an alternative, behaviorist theory (1980–1990); (ii) *theoretical*: the development of the BPM as a means of representing the radical behaviorist methodology in a manner appropriate for the analysis of economic behavior and its use as an interpretive device (1989–2000); (iii) *empirical*: the use of the model to predict consumers' affective responses to purchase and consumption environments (1997–2009); (iv) *behavioral economics*: the development of matching theory and behavioral economics approaches based on the BPM variables (2000–2009); and (v) *philosophical*: the development of post-behaviorist models of behavioral explanation, intentional behaviorism, and superpersonal cognitive psychology (2003–2009). None of these is completed; indeed each is needed in order to stimulate the progress of the others; each has antecedents that predate and projections that postdate the temporal ranges suggested here. (Foxall, 2010, p. 2)

OBJETIVOS

Considerando: 1) as variáveis primárias obtidas no Estudo 1, oriundas da análise computacional de centenas de milhares de comentários de turistas-consumidores usuários do *TripAdvisor*® sobre variados destinos turísticos com recurso ao algoritmo de Análise de Sentimento (AS; Liu, 2010), e 2) que os conteúdos polarizados dos comentários tem natureza avaliativa, podendo ser admitidos como respostas verbais operantes resultantes e indicativas da exposição a contingências de reforço ou punição diretos ou mediados (Skinner, 1957) ao longo da história de consumo turístico dos viajantes comentadores, o presente estudo objetivou: **a)** descrever o responder avaliativo polarizado atribuído aos cem (100) destinos turísticos mais avaliados no período compreendido entre o início de fevereiro e final de março de 2017, e **b)** analisar interpretativamente tal responder com recurso aos conceitos do referencial conceitual do *Behavioral Perspective Model (BPM)*; Foxall, 1990; 2005; 2007; 2010).

MÉTODO

Base de dados

A base de dados utilizada foi composta por 6.438.497 comentários extraídos dos 100 destinos turísticos brasileiros mais avaliados no *TripAdvisor*® no período correspondente a início de fevereiro e final de março de 2017. Para cada comentário, foram extraídas e armazenadas, também, a quantidade de *likes* recebida e, sobre o seu autor, a quantidade de pontos acumulados e a opção escolhida na avaliação do destino turístico a partir da escala *Likert* específica do *TripAdvisor*®. Os dados configuram uma amostra não-probabilística anônima, com características sociodemográficas não alcançáveis.

Materiais e equipamentos

O tratamento e análise dos dados do estudo foram realizados com o *Microsoft Excel*® e o *Statistical Package for Social Sciences – SPSS 22*®. Esses programas estavam instalados em um *notebook* Lenovo G50-80, Intel® Core™ i3-50050U 4GHz, 16 GB RAM, 1 TB HD, com sistema operacional Windows 10®.

Procedimentos

Os dados das cento e noventa e sete (197) variáveis geradas pelos *SentimentALL v2* (Tabela 9, Estudo 1), organizados em um banco de dados (.xls, primeiro; .dat depois) com 19.700 entradas, foram analisados. Previamente às análises principais, condução de análise de varredura (*screening*) dos dados primários revelou perda de apenas 799 valores (4,05%) do total máximo possível. Visto o pequeno percentual de perda, dados perdidos da maioria das variáveis foram substituídos pelo valor da mediana, acrescido ou decrescido de um centésimo do desvio-padrão de cada distribuição. Para os dados perdidos das avaliações positivas e negativas, calculou-se também o número médio geral de avaliações por comentário e derivou-se as quantidades a substituir as avaliações perdidas considerando-se as porcentagens médias das distribuições (positivas e negativas) ponderadas em relação ao total de comentários, por aspecto (ACO, RES e ATR). Desta forma, as análises

principais foram conduzidas com, de fato, 100 valores para cada variável, perfazendo total de 19.700 valores.

RESULTADOS

A partir do módulo de AS da *SentimentALL* apresentado no Estudo 1, foi possível identificar nos milhares de comentários de turistas-consumidores do *TripAdvisor*® uma série de aspectos (características acerca do objeto avaliado), e suas relações com palavras opinativas, de destinos turísticos brasileiros. Com isso, foi definida a base de variáveis primárias da tese (Tabela 9) que, por sua vez, tem como foco a polaridade dos aspectos, o que permitiu inferir se as palavras que são usadas para opinar sobre um destino turístico, suas características ou atributos, eram positivas ou negativas. Os aspectos polarizados, extraídos dos textos, assim como as informações sobre o nível de colaboração no site dos autores dos comentários (pontos acumulados), a utilidade da avaliação (*likes* recebidos) e as escolhas feitas por meio da escala *Likert* do *TripAdvisor*®, têm natureza avaliativa e foram analisados, neste estudo, como respostas verbais operantes, interpretadas com base nos conceitos do *Behavioral Perspective Model* (BPM; FOXALL, 1990; 2005; 2007; 2010). Essa interpretação foi conduzida a partir do esquema apresentado na Figura 2, a seguir.

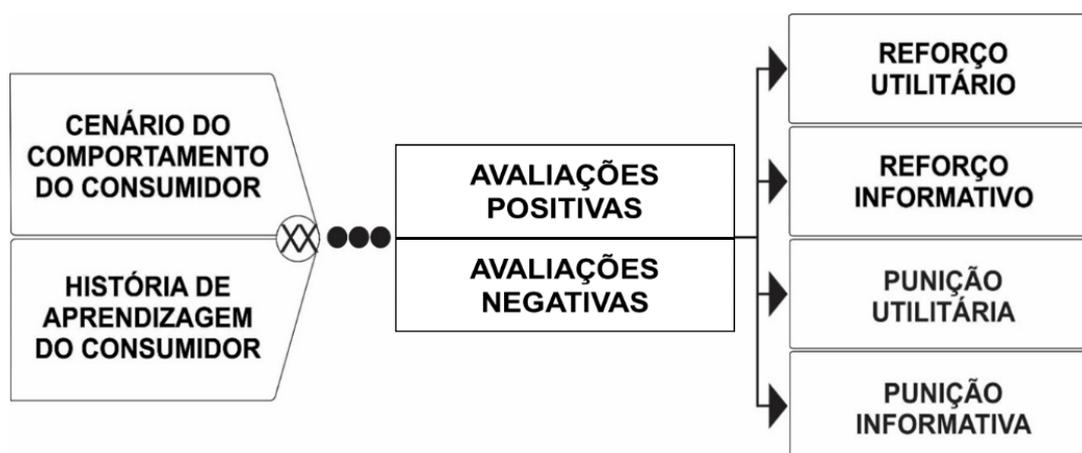


Figura 2: Adaptação, para o contexto de avaliações, da representação esquemática do *Behavioral Perspective Model* (BPM; adaptada de Foxall, 2010, p. 9; extraída e adaptada de Nalini; Cardoso; Cunha, 2013, p. 494).

No contexto do Estudo 2, o comportamento de consumo centralmente analisado foi o comportamento de avaliação de produtos (Figura 2), mais

especificamente de produtos turísticos. Assim, as avaliações verbais feitas pelos usuários do *TripAdvisor*®, apresentadas por meio de totalizações gerais e por tipo de polaridade e componente, a partir de análises ponderadas e correlações, compõem o cerne da discussão. Ao considerar esses dados, buscou-se interpretar as variáveis antecedentes (classificáveis como cenário do comportamento ou história de aprendizagem do turista-consumidor) e consequentes (classificáveis como reforços e punições, utilitárias e informativas) às respostas (positivas e negativas) de avaliar os destinos.

As representações em forma de gráficos das análises, apresentadas a seguir, também foram disponibilizadas em modelo de tabelas, com os nomes dos destinos e seus respectivos valores para facilitar a interpretação. Isso foi necessário porque a análise considerou dados dos 100 destinos turísticos mais avaliados no Brasil, logo se fossem adicionados os nomes desses locais nos gráficos, haveria uma sobrecarga de informação, o que dificultaria o processo de visualização. A seguir, na Tabela 1 é apresentado um *ranking* geral sobre o total de comentários geral e por componentes (ACO, RES e ATR) e, na Tabela 2, uma síntese da proporção entre avaliações e comentários. As tabelas, essas e a maioria das demais, são mais extensas do que o comumente encontrado, mas não foi possível apresentar os dados de outra forma.

Tabela 1: Mostra o número total geral de comentários (nTC) e o número total e percentuais de comentários por componente avaliado (Acomodações [Aco], Restaurantes [Res] e Atrações [Atr]) para os 100 destinos turísticos mais avaliados no *TripAdvisor*® nos meses de fevereiro e março de 2017. Os números totais agregados (Σ) e percentuais para o conjunto completo dos 100 destinos, assim como os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo, são mostrados nas linhas finais da tabela.

	Destino tur	Estado	nTC	nTC_Aco	%	nTC_Res	%	nTC_Atr	%
001	Alto Paraíso	GO	21.313	4.810	22,57	6.341	29,75	10.162	47,68
002	Angra dos Reis	RJ	16.294	5.349	32,83	5.035	30,90	5.910	36,27
003	Aquiraz	CE	10.310	5.689	55,18	2.108	20,45	2.513	24,37
004	Aracaju	SE	71.588	12.156	16,98	31.897	44,56	27.535	38,46
005	Armação dos Búzios	RJ	82.828	14.485	17,49	32.040	38,68	36.303	43,83
006	Arraial d'Ajuda	BA	38.668	5.004	12,94	15.466	40,00	18.198	47,06
007	Arraial do Cabo	RJ	21.573	1.444	6,69	6.826	31,64	13.303	61,67
008	Atibaia	SP	15.298	7.221	47,20	6.814	44,54	1.263	8,26
009	Balneário Camboriú	SC	64.066	12.702	19,83	28.536	44,54	22.828	35,63
010	Barreirinhas	MA	12.542	1.728	13,78	3.042	24,25	7.772	61,97
011	Barueri	SP	13.914	3.129	22,49	10.211	73,39	574	4,13
012	Bauru	SP	12.473	1.788	14,33	8.762	70,25	1.923	15,42
013	Belém	PA	70.177	8.857	12,62	32.628	46,49	28.692	40,89
014	Belo Horizonte	MG	210.469	31.537	14,98	114.175	54,25	64.757	30,77
015	Bento Gonçalves	RS	47.147	7.277	15,43	19.914	42,24	19.956	42,33
016	Blumenau	SC	29.052	4.162	14,33	14.067	48,42	10.823	37,25
017	Bombinhas	SC	20.826	2.065	9,92	8.194	39,35	10.567	50,74
018	Bonito	MS	48.327	4.254	8,80	17.962	37,17	26.111	54,03
019	Brasília	DF	243.462	33.940	13,94	130.562	53,63	78.960	32,43
020	Cabo Frio	RJ	25.588	3.251	12,71	12.599	49,24	9.738	38,06
021	Caldas Novas	GO	33.586	11.168	33,25	9.314	27,73	13.104	39,02
022	Campinas	SP	81.386	10.764	13,23	59.651	73,29	10.971	13,48
023	Campo Grande	MS	32.654	5.773	17,68	20.763	63,58	6.118	18,74
024	Campos do Jordão	SP	120.247	11.481	9,55	61.737	51,34	47.029	39,11
025	Canela	RS	74.247	7.404	9,97	22.432	30,21	44.411	59,82
026	Canoa Quebrada	CE	16.417	3.616	22,03	4.965	30,24	7.836	47,73
027	Caxias do Sul	RS	19.500	3.101	15,90	12.784	65,56	3.615	18,54
028	Cuiabá	MT	37.547	5.251	13,99	20.750	55,26	11.546	30,75
029	Curitiba	PR	182.330	37.875	20,77	45.640	25,03	98.815	54,20
030	Fernando de Noronha	PE	48.265	1.236	2,56	20.975	43,46	26.054	53,98
031	Florianópolis	SC	163.193	22.184	13,59	85.576	52,44	55.433	33,97
032	Fortaleza	CE	227.108	38.090	16,77	111.911	49,28	77.107	33,95
033	Foz do Iguaçu	PR	168.888	43.172	25,56	42.006	24,87	83.710	49,57
034	Goiânia	GO	68.432	11.584	16,93	41.029	59,96	15.819	23,12
035	Gramado	RS	169.910	36.772	21,64	1.025	0,60	132.113	77,75
036	Guarapari	ES	15.371	1.365	8,88	6.235	40,56	7.771	50,56
037	Guarujá	SP	28.218	5.688	20,16	12.070	42,77	10.460	37,07
038	Guarulhos	SP	27.503	7.751	28,18	17.158	62,39	2.594	9,43

Tabela 1: (Continuação) Mostra o número total geral de comentários (nTC) e o número total e percentuais de comentários por componente avaliado (Acomodações [Aco], Restaurantes [Res] e Atrações [Atr]) para os 100 destinos turísticos mais avaliados no *TripAdvisor*® nos meses de fevereiro e março de 2017. Os números totais agregados (Σ) e percentuais para o conjunto completo dos 100 destinos, assim como os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo, são mostrados nas linhas finais da tabela.

	Destino tur	Estado	nTC	nTC_Aco	%	nTC_Res	%	nTC_Atr	%
039	Ilha Grande	RJ	15.742	1.071	6,80	3.324	21,12	11.347	72,08
040	Ilhabela	SP	41.972	6.112	14,56	19.218	45,79	16.642	39,65
041	Ilhéus	BA	16.815	5.827	34,65	5.180	30,81	5.808	34,54
042	Itacaré	BA	24.848	2.745	11,05	10.004	40,26	12.099	48,69
043	Jericoacoara	CE	55.956	6.557	11,72	27.715	49,53	21.684	38,75
044	João Pessoa	PB	93.269	14.268	15,30	50.272	53,90	28.729	30,80
045	Joinville	SC	36.643	6.238	17,02	19.223	52,46	11.182	30,52
046	Juiz de Fora	MG	26.431	3.841	14,53	18.906	71,53	3.684	13,94
047	Jundiaí	SP	21.104	1.685	7,98	17.348	82,20	2.071	9,81
048	Lençóis	BA	12.816	2.122	16,56	5.568	43,45	5.126	40,00
049	Londrina	PR	32.855	5.192	15,80	22.055	67,13	5.608	17,07
050	Maceió	AL	154.488	30.227	19,57	63.150	40,88	61.111	39,56
051	Manaus	AM	79.406	11.463	14,44	36.098	45,46	31.845	40,10
052	Maragogi	AL	35.733	11.989	33,55	8.733	24,44	15.011	42,01
053	Maringá	PR	20.111	3.802	18,91	12.761	63,45	3.548	17,64
054	Monte Verde	MG	33.313	4.804	14,42	22.374	67,16	6.135	18,42
055	Morro de São Paulo	BA	26.703	3.479	13,03	13.125	49,15	10.099	37,82
056	Natal	RN	191.495	44.736	23,36	81.706	42,67	65.053	33,97
057	Niterói	RJ	49.147	1.709	3,48	29.929	60,90	17.509	35,63
058	Nova Petrópolis	RS	13.855	1.233	8,90	6.042	43,61	6.580	47,49
059	Olímpia	SP	22.323	9.001	40,32	5.627	25,21	7.695	34,47
060	Olinda	PE	16.697	929	5,56	7.191	43,07	8.577	51,37
061	Ouro Preto	MG	37.500	2.555	6,81	15.264	40,70	19.681	52,48
062	Palmas	TO	12.438	1.824	14,66	6.107	49,10	4.507	36,24
063	Paraty	RJ	48.528	4.116	8,48	22.061	45,46	22.351	46,06
064	Penedo	RJ	19.484	1.020	5,24	14.281	73,30	4.183	21,47
065	Penha	SC	20.413	1.356	6,64	6.050	29,64	13.007	63,72
066	Petrópolis	RJ	47.451	3.544	7,47	21.680	45,69	22.227	46,84
067	Piracicaba	SP	17.144	1.849	10,79	12.522	73,04	2.773	16,17
068	Pirenópolis	GO	21.619	3.758	17,38	8.701	40,25	9.160	42,37
069	Poços de Caldas	MG	28.354	7.588	26,76	10.304	36,34	10.462	36,90
070	Porto Alegre	RS	83.739	22.828	27,26	15.793	18,86	45.118	53,88
071	Porto de Galinhas	PE	86.125	21.817	25,33	28.460	33,04	35.848	41,62
072	Porto Seguro	BA	68.157	18.691	27,42	17.858	26,20	31.608	46,38
073	Praia da Pipa	RN	34.675	3.761	10,85	13.778	39,73	17.136	49,42
074	Praia do Forte	BA	23.021	7.792	33,85	5.853	25,42	9.376	40,73
075	Recife	PE	84.751	20.904	24,67	14.589	17,21	49.258	58,12
076	Ribeirão Preto	SP	50.063	7.386	14,75	34.906	69,72	7.771	15,52

Tabela 1: (Continuação) Mostra o número total geral de comentários (nTC) e o número total e percentuais de comentários por componente avaliado (Acomodações [Aco], Restaurantes [Res] e Atrações [Atr]) para os 100 destinos turísticos mais avaliados no TripAdvisor® nos meses de fevereiro e março de 2017. Os números totais agregados (Σ) e percentuais para o conjunto completo dos 100 destinos, assim como os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo, são mostrados nas linhas finais da tabela.

	Destino tur	Estado	nTC	nTC_Aco	%	nTC_Res	%	nTC_Atr	%
077	Rio de Janeiro	RJ	607.914	71.885	11,82	305.038	50,18	230.991	38,00
078	Rio Quente	GO	14.745	5.203	35,29	2.690	18,24	6.852	46,47
079	Salvador	BA	179.272	30.037	16,75	81.532	45,48	67.703	37,77
080	Santo André	SP	23.503	1.340	5,70	19.703	83,83	2.460	10,47
081	Santos	SP	51.291	4.816	9,39	26.782	52,22	19.693	38,39
082	São Bernardo do Campo	SP	18.790	447	2,38	15.666	83,37	2.677	14,25
083	São José do Rio Preto	SP	20.062	3.184	15,87	15.118	75,36	1.760	8,77
084	São José dos Campos	SP	32.339	3.963	12,25	24.955	77,17	3.421	10,58
085	São Lourenço	MG	13.638	3.171	23,25	5.675	41,61	4.792	35,14
086	São Luís	MA	45.446	7.549	16,61	22.953	50,51	14.944	32,88
087	São Paulo	SP	814.001	108.004	13,27	502.575	61,74	203.422	24,99
088	São Roque	SP	13.639	1.442	10,57	7.526	55,18	4.671	34,25
089	São Sebastião	SP	20.731	2.737	13,20	8.301	40,04	9.693	46,76
090	Sorocaba	SP	20.634	2.336	11,32	15.333	74,31	2.965	14,37
091	Tamandaré	PE	15.933	564	3,54	5.949	37,34	9.420	59,12
092	Taubaté	SP	12.250	2.486	20,29	8.261	67,44	1.503	12,27
093	Teresina	PI	20.208	2.186	10,82	13.641	67,50	4.381	21,68
094	Teresópolis	RJ	21.919	3.783	17,26	13.989	63,82	4.147	18,92
095	Tiradentes	MG	33.123	2.690	8,12	19.429	58,66	11.004	33,22
096	Trancoso	BA	19.328	2.334	12,08	6.639	34,35	10.355	53,58
097	Ubatuba	SP	38.139	2.713	7,11	13.552	35,53	21.874	57,35
098	Uberlândia	MG	21.189	3.817	18,01	13.540	63,90	3.832	18,08
099	Vila Velha	ES	25.429	1.876	7,38	12.652	49,75	10.901	42,87
100	Vitória	ES	59.179	12.249	20,70	29.341	49,58	17.589	29,72
	$\Sigma =$		6.438.603	1.025.754	15,93	3.027.796	47,03	2.385.053	37,04
	ME =		32.497	4.813	14,55	15.299	45,47	10.936	37,79
	MD =		64.386,03	10.257,54	16,49	30.277,96	47,27	23.850,53	36,24
	DP =		106.835,28	15.400,75	9,26	60.096,30	16,86	36.123,20	15,48
	Mín =		10.310	447		1.025		574	
	Máx =		814.001	108.004		502.575		230.991	

Tabela 2: Mostra as proporções medianas, médias (com desvios-padrão), mínimas e máximas entre os números de avaliações e os números de comentários por total geral, totais por tipos de avaliação (positivas e negativas) e totais por componentes avaliados (acomodações, restaurantes e atrações)

	Total geral	Avaliações POSITIVAS								Avaliações NEGATIVAS							
		Total		Acomodações		Restaurantes		Atrações		Total		Acomodações		Restaurantes		Atrações	
		pTAvC	pAvPC	%	pAvPC	%	pAvPC	%	pAvPC	%	pAvNC	%	pAvNC	%	pAvNC	%	pAvNC
ME =	3,2384	2,9218	90,02	4,1571	90,71	2,9008	89,48	2,3872	90,95	0,3211	9,98	0,4259	9,29	0,3409	10,52	0,2374	9,05
MD =	3,2597	2,9315	89,94	4,1534	90,48	2,9084	89,37	2,3918	90,78	0,3282	10,06	0,4372	9,52	0,3458	10,63	0,2428	9,22
DP =	0,19	0,17	1,24	0,35	0,41	0,17	0,23	0,16	0,29	0,05	1,24	0,09	0,86	0,04	0,67	0,06	0,61
Mín =	2,8433	2,5675		3,2649		2,5180		1,9936		0,2375		0,2847		0,2782		0,1352	
Máx =	3,9484	3,4514		4,9539		3,6811		2,8180		0,5377		0,7092		0,4999		0,5083	

Na Tabela 1, é apresentada a distribuição por destino turístico dos mais de seis milhões de comentários analisados (exatamente 6.438.603), sendo que deste total, 15,93% são do componente Acomodação (ACO), 47,03% de Restaurantes (RES) e 37,04% de Atrações (ATR). Dos 100 destinos turísticos apresentados na Tabela 1, 22% dos comentários foram extraídos de São Paulo (nTC = 814.001) e Rio de Janeiro (nTC = 607.914), treze destinos têm um total de comentários maior ou igual a 100.000 (Belo Horizonte, Brasília, Campos de Jordão, Curitiba, Florianópolis, Fortaleza, Foz do Iguaçu, Gramado, Maceió, Natal, Rio de Janeiro, Salvador e São Paulo), enquanto o destino com menor número de comentários, Aquiraz, teve um total de 10.310. Em relação a comentários por região do Brasil, a região Sudeste está na primeira posição absoluta, com 2.977.488, seguida da região Nordeste (1.630.634), Sul (1.146.775), Centro-Oeste (521.685, sendo que desse total, 243.462 comentários são do Distrito Federal) e, por último, com a participação de apenas 3 estados, dos 6 que compõem a região, a região Norte (162.021).

Como um comentário é um texto de um turista-consumidor em um site que reflete sua opinião acerca de um determinado destino turístico, mais especificamente sobre um componente desse produto turístico, um único comentário pode conter uma, algumas ou várias avaliações. Por exemplo, no comentário sobre um restaurante, “*a comida é saborosa, o preço é bom, mas o atendimento é ruim*”, tem-se três avaliações, a comida (positiva), o preço (positivo) e o atendimento (negativo). Na Tabela 2 são apresentadas algumas proporções entre os números de avaliações e os números de comentários por total geral, totais por tipos de avaliação (positivas e negativas), além de totais por componentes avaliados (ACO, RES, ATR) e evidencia que a maior média e mediana dessas proporções (em ambos os tipos de avaliação) é do componente ACO.

Quando os destinos são mais comentados, há um maior número de estimulação verbal discriminativa para respostas avaliativas. Nessa circunstância, uma variável contextual pode ser o número de avaliações (que, nesse trabalho, foram extraídas dos comentários) de um determinado destino turístico, ou seja, as avaliações funcionam como estimulação discriminativa para o avaliar pelo viajante. Uma possibilidade de interpretação dos diferentes posicionamentos dos destinos nos *rankings* passa por concebê-los a partir do conceito de cenário de consumo mais ou menos abertos (cf. Foxal, 1990; 1993; 2001). Assim, destinos “cenários mais abertos”, comparativamente a destinos “cenários mais fechados” caracterizam-se por maior

número de alternativas para o consumo turístico (muitas e variadas acomodações, restaurantes e atrativos, por exemplo) e, conseqüentemente, maior número de contingências para respostas de consumo que terminarão reforçadas e/ou punidas, induzindo maior número de comentários e avaliações positivas ou negativas. Nesse contexto, aqueles destinos que são bem posicionados no *ranking* de avaliações podem funcionar como cenários mais abertos. Assim como cenários fechados podem ser compreendidos, nessa interpretação, por aqueles destinos que estejam nas últimas posições do *ranking* (em quantidade de avaliações). Os diferentes posicionamentos dos destinos nos *rankings* podem ser concebidos a partir do *Continuum* de fechamento-abertura do cenário do comportamento (baseado em Foxall, 1993), conforme pode ser observado na Figura 3.

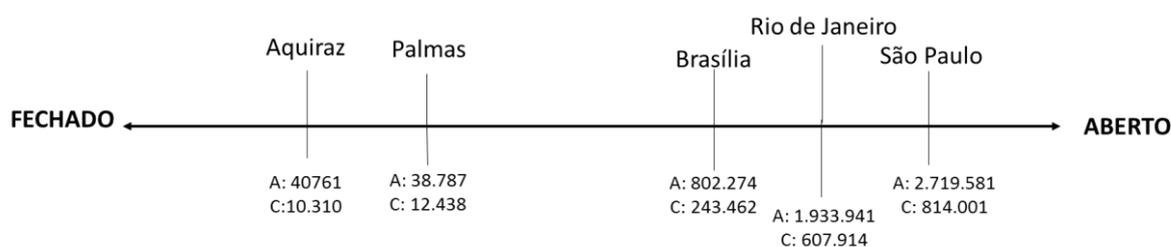


Figura 3: *Continuum* de fechamento-abertura do cenário do comportamento do consumidor-turista baseado na quantidade de avaliações/comentários. (Baseado em Foxall, 1993).

Ao considerar essa interpretação do *continuum* de fechamento-abertura do cenário do comportamento do consumidor-turista, pode ser mais provável que os destinos que estão nas primeiras posições da lista de proporção relativa ponderada de avaliações positivas e negativas (Tabela 3, apresentada a seguir), como São Paulo, Rio de Janeiro, Brasília, Fortaleza, Belo Horizonte, por exemplo, sejam cenários de consumos mais abertos. Ao considerar, como exemplificação, na cidade de São Paulos, foram extraídos 814.001 comentários e a partir disso foram analisados na *SentimentALL* um total de 2.719.581 aspectos polarizados. Por outro lado, cenários como Palmas, a capital do Tocantins, com um total de 12.438 comentários, com 38.787 aspectos polarizados, podem ser interpretados como cenários de consumo que tendem a ser mais fechados.

Nas Tabela 3 e Figura 4, a seguir, são apresentados os *rankings* dos destinos turístico dos valores das proporções de avaliações ponderadas em relação às proporções relativas de comentários.

Tabela 3: Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) ponderadas em relação às proporções relativas de comentários ($NTC_d/NC_{máx}$) atribuídos a cada um dos 100 destinos turísticos analisados.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
1º	São Paulo	0,8989	1º	São Paulo	0,1011
2º	Rio de Janeiro	0,6674	2º	Rio de Janeiro	0,0794
3º	Brasília	0,2689	3º	Brasília	0,0301
4º	Fortaleza	0,2511	4º	Fortaleza	0,0279
5º	Belo Horizonte	0,2349	5º	Belo Horizonte	0,0237
6º	Natal	0,2140	6º	Salvador	0,0230
7º	Curitiba	0,2057	7º	Natal	0,0212
8º	Salvador	0,1972	8º	Florianópolis	0,0203
9º	Gramado	0,1913	9º	Maceió	0,0189
10º	Foz do Iguaçu	0,1895	10º	Curitiba	0,0183
11º	Florianópolis	0,1802	11º	Foz do Iguaçu	0,0180
12º	Maceió	0,1709	12º	Gramado	0,0174
13º	Campos do Jordão	0,1321	13º	Campos do Jordão	0,0156
14º	João Pessoa	0,1035	14º	Armação dos Búzios	0,0114
15º	Recife	0,0950	15º	Porto de Galinhas	0,0113
16º	Porto de Galinhas	0,0945	16º	João Pessoa	0,0111
17º	Porto Alegre	0,0937	17º	Campinas	0,0110
18º	Armação dos Búzios	0,0904	18º	Porto Alegre	0,0091
19º	Campinas	0,0890	19º	Recife	0,0091
20º	Manaus	0,0888	20º	Aracaju	0,0089
21º	Canela	0,0834	21º	Porto Seguro	0,0089
22º	Aracaju	0,0790	22º	Manaus	0,0088
23º	Belém	0,0783	23º	Goiânia	0,0086
24º	Goiânia	0,0755	24º	Belém	0,0079
25º	Porto Seguro	0,0748	25º	Canela	0,0078
26º	Balneário Camboriú	0,0715	26º	Balneário Camboriú	0,0072
27º	Vitória	0,0660	27º	Jericoacoara	0,0069
28º	Jericoacoara	0,0619	28º	Vitória	0,0067
29º	Santos	0,0570	29º	Paraty	0,0067
30º	Ribeirão Preto	0,0553	30º	Niterói	0,0062
31º	Niterói	0,0542	31º	Ribeirão Preto	0,0062
32º	Bonito	0,0538	32º	Petrópolis	0,0060
33º	Fernando de Noronha	0,0534	33º	Santos	0,0060
34º	Bento Gonçalves	0,0533	34º	Fernando de Noronha	0,0059
35º	Paraty	0,0530	35º	São Luís	0,0055
36º	Petrópolis	0,0523	36º	Bonito	0,0055
37º	São Luís	0,0503	37º	Ilhabela	0,0053
38º	Ilhabela	0,0463	38º	Caldas Novas	0,0051
39º	Arraial d'Ajuda	0,0429	39º	Maragogi	0,0048
40º	Ubatuba	0,0421	40º	Ubatuba	0,0048

Tabela 3: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) ponderadas em relação às proporções relativas de comentários ($NTC_d/NC_{máx}$) atribuídos a cada um dos 100 destinos turísticos analisados.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
41º	Cuiabá	0,0419	41º	Monte Verde	0,0047
42º	Ouro Preto	0,0416	42º	Arraial d'Ajuda	0,0047
43º	Joinville	0,0410	43º	Bento Gonçalves	0,0047
44º	Maragogi	0,0391	44º	Ouro Preto	0,0044
45º	Praia da Pipa	0,0385	45º	Cuiabá	0,0042
46º	Londrina	0,0370	46º	Praia da Pipa	0,0041
47º	Campo Grande	0,0367	47º	Tiradentes	0,0041
48º	Tiradentes	0,0366	48º	Joinville	0,0040
49º	Monte Verde	0,0363	49º	São José dos Campos	0,0039
50º	Caldas Novas	0,0362	50º	Guarulhos	0,0038
51º	São José dos Campos	0,0358	51º	Cabo Frio	0,0037
52º	Blumenau	0,0326	52º	Guarujá	0,0036
53º	Poços de Caldas	0,0315	53º	Morro de São Paulo	0,0034
54º	Guarujá	0,0310	54º	Campo Grande	0,0034
55º	Guarulhos	0,0299	55º	Olímpia	0,0034
56º	Juiz de Fora	0,0295	56º	Londrina	0,0034
57º	Morro de São Paulo	0,0294	57º	Trancoso	0,0034
58º	Vila Velha	0,0284	58º	Poços de Caldas	0,0033
59º	Cabo Frio	0,0277	59º	Arraial do Cabo	0,0032
60º	Itacaré	0,0273	60º	Itacaré	0,0032
61º	Santo André	0,0261	61º	Blumenau	0,0031
62º	Praia do Forte	0,0256	62º	São José do Rio Preto	0,0031
63º	Teresópolis	0,0243	63º	Alto Paraíso de Goiás	0,0030
64º	Olímpia	0,0240	64º	Juiz de Fora	0,0030
65º	Pirenópolis	0,0237	65º	Penha	0,0029
66º	Uberlândia	0,0236	66º	Pirenópolis	0,0029
67º	Jundiaí	0,0234	67º	Vila Velha	0,0029
68º	Arraial do Cabo	0,0233	68º	Penedo	0,0029
69º	Alto Paraíso	0,0232	69º	Santo André	0,0028
70º	Sorocaba	0,0230	70º	Praia do Forte	0,0027
71º	Bombinhas	0,0230	71º	Teresópolis	0,0026
72º	São Sebastião	0,0230	72º	Bombinhas	0,0026
73º	Teresina	0,0225	73º	Rio Quente	0,0026
74º	Maringá	0,0225	74º	Jundiaí	0,0025
75º	Penha	0,0222	75º	São Sebastião	0,0025
76º	Caxias do Sul	0,0221	76º	Uberlândia	0,0025
77º	São José do Rio Preto	0,0216	77º	São Bernardo do Campo	0,0024
78º	Penedo	0,0211	78º	Angra dos Reis	0,0024
79º	São Bernardo do Campo	0,0206	79º	Sorocaba	0,0023

Tabela 3: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) ponderadas em relação às proporções relativas de comentários ($NTC_d/NC_{máx}$) atribuídos a cada um dos 100 destinos turísticos analisados.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
80º	Trancoso	0,0204	80º	Teresina	0,0023
81º	Piracicaba	0,0190	81º	Ilhéus	0,0022
82º	Olinda	0,0187	82º	Maringá	0,0022
83º	Ilhéus	0,0185	83º	Ilha Grande	0,0022
84º	Canoa Quebrada	0,0184	84º	Guarapari	0,0021
85º	Angra dos Reis	0,0176	85º	Piracicaba	0,0021
86º	Tamandaré	0,0176	86º	Tamandaré	0,0020
87º	Ilha Grande	0,0172	87º	Olinda	0,0019
88º	Atibaia	0,0170	88º	Caxias do Sul	0,0018
89º	Guarapari	0,0168	89º	Canoa Quebrada	0,0018
90º	Nova Petrópolis	0,0158	90º	Atibaia	0,0018
91º	Rio Quente	0,0156	91º	Barueri	0,0017
92º	Barueri	0,0154	92º	Lençóis	0,0017
93º	São Lourenço	0,0153	93º	Barreirinhas	0,0016
94º	São Roque	0,0153	94º	Aquiraz	0,0016
95º	Lençóis	0,0140	95º	Taubaté	0,0015
96º	Palmas	0,0139	96º	Bauru	0,0015
97º	Bauru	0,0138	97º	São Roque	0,0015
98º	Barreirinhas	0,0138	98º	São Lourenço	0,0015
99º	Taubaté	0,0135	99º	Palmas	0,0014
100º	Aquiraz	0,0111	100º	Nova Petrópolis	0,0012

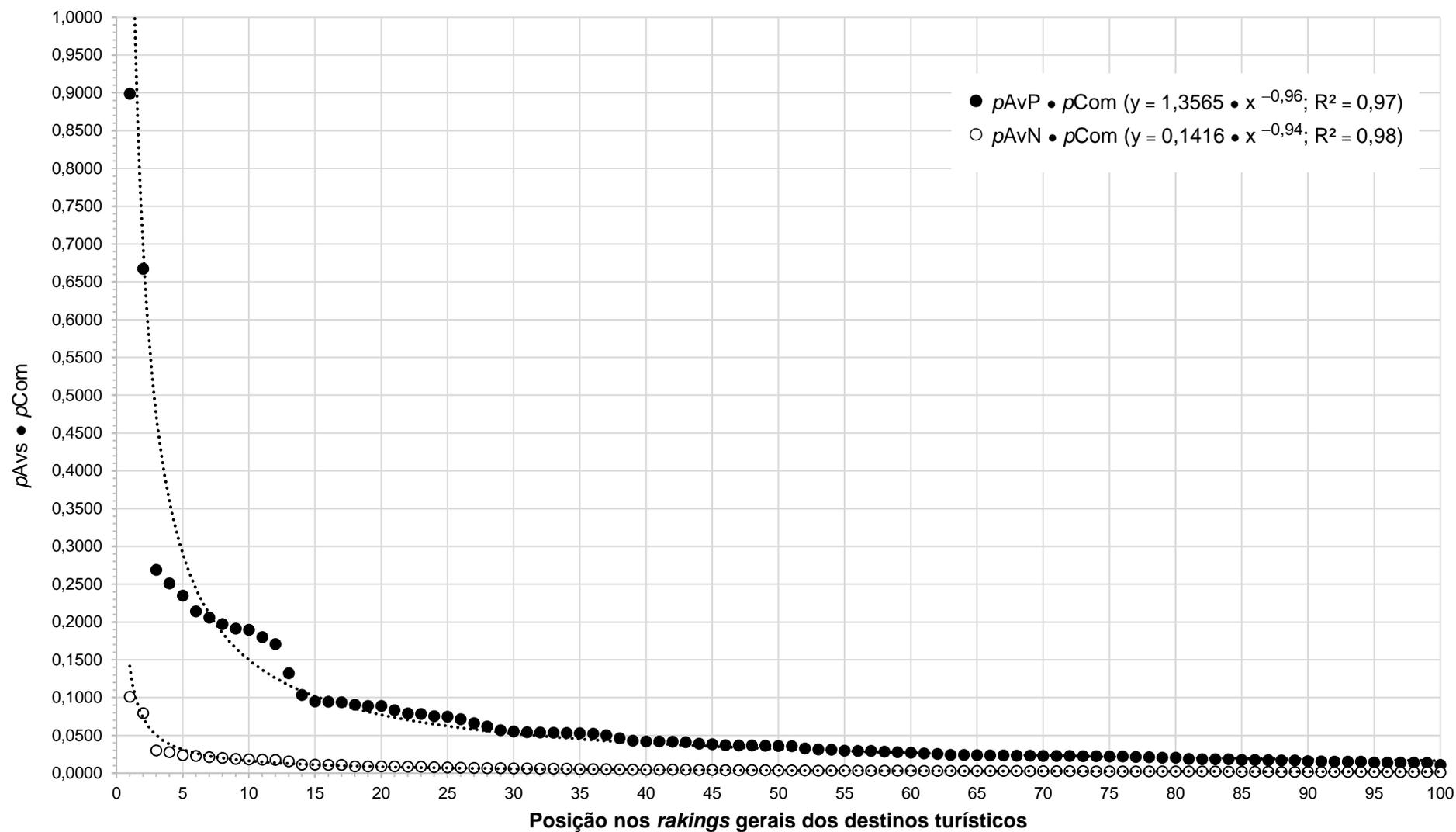


Figura 4: Mostra dos valores das proporções gerais de avaliações positivas (AvP) e negativas (AvN) ponderadas em relação às proporções relativas de comentários atribuídos a cada um dos 100 destinos turísticos analisados, conforme posição assumida por estes nos *rankings* obtidos com a disposição decrescente dos valores. Os expoentes das equações de ajuste (funções potência) indicam a taxa média de decréscimo dos valores da variável a cada posição sucessiva. Curvas correspondentes aos *rankings* na **Tabela 3**, onde os nomes dos destinos e os valores específicos para cada destino podem ser observados.

Conforme dados apresentados na Tabela 3 e nas curvas de *rankings* geradas a partir desses dados na Figura 4, podem ser observadas algumas tendências gerais relacionadas aos destinos turísticos considerando a proporção relativa ponderada das avaliações positivas ($pAvP \bullet pCom$) e negativas ($pAvN \bullet pCom$). Por exemplo, ao observar o gráfico da Figura 4, São Paulo ($pAvN \bullet pCom = 0,8989$) está isolada na ponta superior da curva de avaliações positivas, seguida da cidade do Rio de Janeiro ($pAvN \bullet pCom = 0,6674$). Depois, seguem-se as cidades de Brasília, Fortaleza, Belo Horizonte, Natal e Curitiba, que estão entre o intervalo de 0,27 e 0,20 ($pAvP \bullet pCom$), vale também ressaltar, que a partir do destino turístico Campos de Jordão, com $pAvN \bullet pCom = 0,1321$ (na 13ª Posição) ocorre uma ampliação no valor de decréscimo, que pode ser observado com Maceió, que vem em seguida, com $pAvN \bullet pCom = 0,1035$, o que dá início a parte mais horizontal e mais longa da curva. Em relação às respostas verbais negativas, as cinco primeiras posições são ocupadas pelos mesmos destinos turísticos com maiores valores em $pAvP \bullet pCom$. No entanto, a proporção relativa de avaliações negativas é significativamente menor que as das avaliações positivas, e a sua curva alcança um valor máximo de $pAvN \bullet pCom = 0,1011$. Dois destinos turísticos se sobressaem, os mesmos da curva relacionada às avaliações positivas, São Paulo ($pAvN \bullet pCom = 0,1011$) e Rio de Janeiro ($pAvN \bullet pCom = 0,0794$), e tem algum destaque, também, com $pAvN \bullet pCom$ superior a 0,025 e inferior a 0,050, Brasília e Fortaleza, mas os demais compõem a parte mais horizontal da curva com uma variação bem pequena dos valores.

Dez destinos turísticos mantiveram sua posição no ranking, considerando as duas polaridades dos aspectos (1º São Paulo, 2º Rio de Janeiro, 3º Brasília, 4º Fortaleza, 5º Belo Horizonte, 13º Campos do Jordão, 26º Balneário Camboriú, 48º Tiradentes, 60º Itacaré, 86º. Tamandaré), a grande maioria restante ficou em posições bem próximas em ambos os rankings, com exceção de Trancoso (80º posição nas avaliações positivas, 57º nas avaliações negativas) e Rio Quente (91º - avaliações positivas, 73º - avaliações negativas) que tiveram uma alteração de posição mais significativa (superior a 15). Assim, ao analisar a variável de cenário “quantidade de aspectos”, verifica-se que quanto maior o número de aspectos identificado nos comentários de um determinado destino turístico, maior a probabilidade desse destino estar presente nas primeiras posições de ambos os tipos de respostas verbais (positivas ou negativas). Nesse contexto, os aspectos podem representar consequências aversivas ou reforços utilitários, pois os comentários dos turistas

geralmente são sobre características dos destinos turísticos que visitaram, ou seja, descrevem benefícios ou punições referentes ao valor de uso dos serviços e produtos, por exemplo, sobre o atendimento no restaurante, a localização de uma praia, o quarto do hotel, a comida do restaurante, a piscina da pousada, etc.

Em relação às Capitais do Brasil, por exemplo, Palmas (capital do Tocantins) está entre os últimos destinos no ranking (96º e 99º, positivo e negativo, respectivamente), mas há outras capitais, como Rio Branco, Macapá e Porto Velho que nem aparecem no conjunto dos 100 destinos mais avaliados. Assim, de todas as capitais que aparecem nesse conjunto, mais de 90% estão entre os 50 destinos mais avaliados (em ambas as polaridades). Essas conclusões reforçam que as capitais, assim como as cidades de uma forma geral, que têm cenários de consumo com mais variáveis, por exemplo, mais pontos turísticos, atrações, locais diferenciados, espaço na mídia, opções de lazer, polos comerciais, etc., são aquelas que possuem a maior quantidade de respostas verbais no site.

Em uma síntese sobre o quantitativo de aspectos polarizados extraídos dos comentários, têm-se que as curvas das avaliações positivas e negativas apresentadas na Figura 4 têm um padrão semelhante, evidenciando a tendência à horizontalidade, com apenas alguns poucos destinos (na parte superior da curva) em destaque com o valor de proporção acima da média dos conjuntos. Mas, a diferença de altura inicial dessas duas curvas evidencia, principalmente, o fato de que os comentários dos turistas são em sua maioria (90%) positivos, o que permite a inferência de que a emissão de comentários nesse contexto deriva do reforço pelos benefícios oriundos diretamente do uso de produtos e/ou serviços dos destinos turísticos e não das consequências aversivas. A Tabela 4, Tabela 5, Tabela 6 e a Figura 5, apresentadas adiante, mostram a mesma forma de esquematização dos dados, mas no contexto de cada componente isoladamente.

Tabela 4: Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ACOMODAÇÕES (ACO) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_ACO_d/NC_{máx_ACO}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 5.

Rank	Destinos turísticos	$pAvP \cdot pCom$	Rank	Destinos turísticos	$pAvN \cdot pCom$
1º	São Paulo	,9054	1º	São Paulo	,0946
2º	Rio de Janeiro	,5995	2º	Rio de Janeiro	,0661
3º	Natal	,3723	3º	Natal	,0419
4º	Foz do Iguaçu	,3656	4º	Fortaleza	,0353
5º	Curitiba	,3201	5º	Foz do Iguaçu	,0342
6º	Fortaleza	,3174	6º	Salvador	,0313
7º	Gramado	,3170	7º	Curitiba	,0306
8º	Brasília	,2847	8º	Brasília	,0296
9º	Belo Horizonte	,2674	9º	Maceió	,0247
10º	Maceió	,2552	10º	Belo Horizonte	,0246
11º	Salvador	,2468	11º	Gramado	,0235
12º	Porto Alegre	,1902	12º	Florianópolis	,0219
13º	Florianópolis	,1835	13º	Porto Alegre	,0211
14º	Porto de Galinhas	,1809	14º	Porto de Galinhas	,0211
15º	Recife	,1744	15º	Porto Seguro	,0206
16º	Porto Seguro	,1525	16º	Recife	,0192
17º	Armação dos Búzios	,1200	17º	Armação dos Búzios	,0141
18º	João Pessoa	,1190	18º	Caldas Novas	,0138
19º	Balneário Camboriú	,1058	19º	João Pessoa	,0131
20º	Vitória	,1030	20º	Balneário Camboriú	,0118
21º	Aracaju	,1013	21º	Maragogi	,0115
22º	Maragogi	,0995	22º	Goiânia	,0113
23º	Campos do Jordão	,0975	23º	Aracaju	,0113
24º	Goiânia	,0959	24º	Olímpia	,0111
25º	Manaus	,0954	25º	Campinas	,0108
26º	Caldas Novas	,0896	26º	Manaus	,0108
27º	Campinas	,0889	27º	Vitória	,0104
28º	Belém	,0732	28º	Belém	,0088
29º	Olímpia	,0722	29º	Campos do Jordão	,0088
30º	Praia do Forte	,0656	30º	Guarulhos	,0082
31º	Guarulhos	,0636	31º	Angra dos Reis	,0074
32º	Canela	,0634	32º	São Luís	,0071
33º	Poços de Caldas	,0632	33º	Poços de Caldas	,0071
34º	São Luís	,0628	34º	Aquiraz	,0067
35º	Bento Gonçalves	,0622	35º	Ribeirão Preto	,0067
36º	Ribeirão Preto	,0617	36º	Praia do Forte	,0065
37º	Atibaia	,0604	37º	Rio Quente	,0064
38º	Jericoacoara	,0559	38º	Atibaia	,0064

Tabela 4: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ACOMODAÇÕES (ACO) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_ACO_d/NC_{máx_ACO}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 3.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
39º	Joinville	,0532	39º	Ilhéus	,0063
40º	Ilhabela	,0523	40º	Bento Gonçalves	,0052
41º	Campo Grande	,0492	41º	Canela	,0051
42º	Guarujá	,0478	42º	Guarujá	,0049
43º	Ilhéus	,0477	43º	Jericoacoara	,0048
44º	Aquiraz	,0459	44º	Arraial d'Ajuda	,0047
45º	Cuiabá	,0444	45º	Joinville	,0046
46º	Londrina	,0440	46º	Cabo Frio	,0046
47º	Angra dos Reis	,0422	47º	Ilhabela	,0043
48º	Rio Quente	,0417	48º	Campo Grande	,0043
49º	Arraial d'Ajuda	,0416	49º	Alto Paraíso de Goiás	,0042
50º	Monte Verde	,0407	50º	Santos	,0042
51º	Santos	,0404	51º	Cuiabá	,0042
52º	Alto Paraíso de Goiás	,0403	52º	Londrina	,0041
53º	Bonito	,0361	53º	Monte Verde	,0037
54º	Blumenau	,0353	54º	São José do Rio Preto	,0036
55º	Paraty	,0346	55º	Paraty	,0035
56º	São José dos Campos	,0334	56º	Petrópolis	,0034
57º	Uberlândia	,0323	57º	Juiz de Fora	,0034
58º	Juiz de Fora	,0322	58º	Praia da Pipa	,0034
59º	Maringá	,0321	59º	Teresópolis	,0033
60º	Pirenópolis	,0319	60º	Bonito	,0033
61º	Teresópolis	,0317	61º	Morro de São Paulo	,0033
62º	Praia da Pipa	,0314	62º	Blumenau	,0032
63º	Canoa Quebrada	,0309	63º	São José dos Campos	,0032
64º	Petrópolis	,0294	64º	Maringá	,0031
65º	Morro de São Paulo	,0289	65º	Uberlândia	,0030
66º	São Lourenço	,0268	66º	Pirenópolis	,0029
67º	Caxias do Sul	,0265	67º	Barueri	,0028
68º	Barueri	,0262	68º	Trancoso	,0026
69º	São José do Rio Preto	,0259	69º	Canoa Quebrada	,0026
70º	Cabo Frio	,0255	70º	São Lourenço	,0026
71º	São Sebastião	,0234	71º	Ubatuba	,0024
72º	Itacaré	,0231	72º	Itacaré	,0023
73º	Tiradentes	,0230	73º	Caxias do Sul	,0022
74º	Ubatuba	,0228	74º	Teresina	,0020
75º	Ouro Preto	,0217	75º	Taubaté	,0020
76º	Taubaté	,0210	76º	São Sebastião	,0019

Tabela 4: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ACOMODAÇÕES (ACO) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_ACO_d/NC_{máx_ACO}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 3.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
77º	Sorocaba	,0198	77º	Ouro Preto	,0019
78º	Trancoso	,0190	78º	Tiradentes	,0019
79º	Teresina	,0182	79º	Piracicaba	,0019
80º	Lençóis	,0181	80º	Sorocaba	,0019
81º	Bombinhas	,0175	81º	Bombinhas	,0017
82º	Vila Velha	,0159	82º	Barreirinhas	,0016
83º	Palmas	,0155	83º	Bauru	,0016
84º	Piracicaba	,0152	84º	Lençóis	,0016
85º	Bauru	,0149	85º	Niterói	,0015
86º	Barreirinhas	,0144	86º	Vila Velha	,0015
87º	Jundiaí	,0143	87º	Guarapari	,0014
88º	Niterói	,0143	88º	Palmas	,0014
89º	Arraial do Cabo	,0124	89º	Jundiaí	,0013
90º	São Roque	,0121	90º	São Roque	,0012
91º	Penha	,0115	91º	Penha	,0010
92º	Santo André	,0115	92º	Arraial do Cabo	,0009
93º	Guarapari	,0113	93º	Santo André	,0009
94º	Fernando de Noronha	,0108	94º	Ilha Grande	,0008
95º	Nova Petrópolis	,0108	95º	Olinda	,0007
96º	Ilha Grande	,0091	96º	São Bernardo do Campo	,0007
97º	Penedo	,0088	97º	Penedo	,0007
98º	Olinda	,0079	98º	Fernando de Noronha	,0006
99º	Tamandaré	,0047	99º	Nova Petrópolis	,0006
100º	São Bernardo do Campo	,0034	100º	Tamandaré	,0005

Tabela 5: Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente RESTAURANTES (RES) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_RES_d/NC_{máx_RES}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 5.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
1º	São Paulo	,8894	1º	São Paulo	,1106
2º	Rio de Janeiro	,5319	2º	Rio de Janeiro	,0750
3º	Brasília	,2297	3º	Brasília	,0301
4º	Belo Horizonte	,2035	4º	Belo Horizonte	,0237
5º	Fortaleza	,1991	5º	Fortaleza	,0236
6º	Florianópolis	,1527	6º	Florianópolis	,0176
7º	Natal	,1480	7º	Salvador	,0175
8º	Salvador	,1447	8º	Campos do Jordão	,0151
9º	Maceió	,1130	9º	Natal	,0145
10º	Campos do Jordão	,1077	10º	Campinas	,0134
11º	Campinas	,1053	11º	Maceió	,0127
12º	João Pessoa	,0900	12º	João Pessoa	,0101
13º	Curitiba	,0811	13º	Curitiba	,0097
14º	Foz do Iguaçu	,0747	14º	Foz do Iguaçu	,0089
15º	Goiânia	,0729	15º	Goiânia	,0087
16º	Manaus	,0646	16º	Armação dos Búzios	,0073
17º	Ribeirão Preto	,0622	17º	Ribeirão Preto	,0072
18º	Belém	,0585	18º	Manaus	,0072
19º	Aracaju	,0567	19º	Aracaju	,0068
20º	Armação dos Búzios	,0564	20º	Porto de Galinhas	,0067
21º	Niterói	,0529	21º	Niterói	,0067
22º	Vitória	,0524	22º	Belém	,0064
23º	Balneário Camboriú	,0513	23º	Vitória	,0059
24º	Porto de Galinhas	,0499	24º	Jericoacoara	,0058
25º	Jericoacoara	,0493	25º	Santos	,0057
26º	Santos	,0476	26º	Balneário Camboriú	,0055
27º	São José dos Campos	,0445	27º	Monte Verde	,0054
28º	São Luís	,0413	28º	Paraty	,0053
29º	Canela	,0405	29º	São José dos Campos	,0052
30º	Londrina	,0401	30º	Petrópolis	,0049
31º	Monte Verde	,0391	31º	Fernando de Noronha	,0046
32º	Paraty	,0386	32º	São Luís	,0044
33º	Petrópolis	,0383	33º	Ilhabela	,0042
34º	Campo Grande	,0375	34º	Tiradentes	,0042
35º	Cuiabá	,0375	35º	Canela	,0041
36º	Fernando de Noronha	,0371	36º	Santo André	,0040
37º	Bento Gonçalves	,0361	37º	Guarulhos	,0039
38º	Santo André	,0352	38º	Londrina	,0038

Tabela 5: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente RESTAURANTES (RES) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_RES_d/NC_{máx_RES}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 3.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
39º	Joinville	,0348	39º	Cuiabá	,0038
40º	Tiradentes	,0345	40º	Campo Grande	,0038
41º	Juiz de Fora	,0342	41º	São José do Rio Preto	,0037
42º	Ilhabela	,0340	42º	Porto Seguro	,0037
43º	Bonito	,0321	43º	Bonito	,0036
44º	Porto Seguro	,0319	44º	Penedo	,0036
45º	Jundiá	,0309	45º	Jundiá	,0036
46º	Guarulhos	,0302	46º	Bento Gonçalves	,0035
47º	Porto Alegre	,0281	47º	Joinville	,0035
48º	São Bernardo do Campo	,0279	48º	Juiz de Fora	,0035
49º	Sorocaba	,0277	49º	Porto Alegre	,0033
50º	Ouro Preto	,0276	50º	Arraial d'Ajuda	,0033
51º	Arraial d'Ajuda	,0275	51º	São Bernardo do Campo	,0032
52º	São José do Rio Preto	,0263	52º	Recife	,0031
53º	Recife	,0259	53º	Cabo Frio	,0031
54º	Blumenau	,0256	54º	Morro de São Paulo	,0030
55º	Teresópolis	,0250	55º	Guarujá	,0030
56º	Penedo	,0248	56º	Ubatuba	,0029
57º	Praia da Pipa	,0248	57º	Sorocaba	,0029
58º	Teresina	,0246	58º	Teresópolis	,0028
59º	Uberlândia	,0242	59º	Ouro Preto	,0028
60º	Ubatuba	,0241	60º	Uberlândia	,0028
61º	Caxias do Sul	,0234	61º	Praia da Pipa	,0027
62º	Morro de São Paulo	,0231	62º	Teresina	,0026
63º	Maringá	,0230	63º	Piracicaba	,0025
64º	Vila Velha	,0227	64º	Vila Velha	,0025
65º	Piracicaba	,0224	65º	Blumenau	,0024
66º	Cabo Frio	,0220	66º	Maringá	,0024
67º	Guarujá	,0211	67º	Maragogi	,0022
68º	Poços de Caldas	,0186	68º	Pirenópolis	,0021
69º	Barueri	,0182	69º	Barueri	,0021
70º	Itacaré	,0179	70º	Caxias do Sul	,0021
71º	Caldas Novas	,0166	71º	Itacaré	,0020
72º	Bauru	,0157	72º	Caldas Novas	,0020
73º	Pirenópolis	,0152	73º	São Sebastião	,0020
74º	Maragogi	,0152	74º	Trancoso	,0020
75º	Taubaté	,0147	75º	Poços de Caldas	,0019
76º	São Sebastião	,0145	76º	Arraial do Cabo	,0018

Tabela 5: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente RESTAURANTES (RES) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_RES_d/NC_{máx_RES}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 3.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
77º	Bombinhas	,0145	77º	Bombinhas	,0018
78º	São Roque	,0136	78º	Bauru	,0017
79º	Olinda	,0129	79º	Taubaté	,0017
80º	Atibaia	,0123	80º	Tamandaré	,0016
81º	Arraial do Cabo	,0118	81º	Penha	,0015
82º	Alto Paraíso de Goiás	,0113	82º	Olinda	,0014
83º	Trancoso	,0112	83º	Alto Paraíso de Goiás	,0014
84º	Guarapari	,0111	84º	São Roque	,0013
85º	Nova Petrópolis	,0110	85º	Praia do Forte	,0013
86º	Palmas	,0110	86º	Guarapari	,0013
87º	Penha	,0106	87º	Atibaia	,0013
88º	Praia do Forte	,0103	88º	Palmas	,0012
89º	Tamandaré	,0103	89º	Lençóis	,0012
90º	São Lourenço	,0102	90º	Olímpia	,0012
91º	Olímpia	,0100	91º	Canoa Quebrada	,0011
92º	Lençóis	,0099	92º	Angra dos Reis	,0011
93º	Ilhéus	,0093	93º	São Lourenço	,0011
94º	Angra dos Reis	,0089	94º	Ilhéus	,0010
95º	Canoa Quebrada	,0088	95º	Nova Petrópolis	,0010
96º	Ilha Grande	,0057	96º	Ilha Grande	,0009
97º	Barreirinhas	,0053	97º	Barreirinhas	,0007
98º	Rio Quente	,0047	98º	Rio Quente	,0006
99º	Aquiraz	,0037	99º	Aquiraz	,0005
100º	Gramado	,0018	100º	Gramado	,0002

Tabela 6: Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ATRATIVOS (ATR) ponderadas em relação às proporções relativas de comentários atribuídos ao componente ($NTC_ATR_d/NC_{máx_ATR}$), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 5.

Rank	Destinos turísticos	$pAvP \cdot pCom$	Rank	Destinos turísticos	$pAvN \cdot pCom$
1º	Rio de Janeiro	,9199	1º	Rio de Janeiro	,0801
2º	São Paulo	,8128	2º	São Paulo	,0679
3º	Gramado	,5193	3º	Gramado	,0526
4º	Curitiba	,4004	4º	Fortaleza	,0297
5º	Foz do Iguaçu	,3356	5º	Maceió	,0284
6º	Brasília	,3174	6º	Curitiba	,0274
7º	Fortaleza	,3041	7º	Salvador	,0273
8º	Salvador	,2658	8º	Foz do Iguaçu	,0268
9º	Belo Horizonte	,2608	9º	Brasília	,0244
10º	Natal	,2599	10º	Florianópolis	,0225
11º	Maceió	,2361	11º	Natal	,0218
12º	Florianópolis	,2174	12º	Belo Horizonte	,0195
13º	Recife	,1976	13º	Armação dos Búzios	,0178
14º	Campos do Jordão	,1864	14º	Campos do Jordão	,0172
15º	Porto Alegre	,1813	15º	Canela	,0161
16º	Canela	,1761	16º	Recife	,0157
17º	Porto de Galinhas	,1400	17º	Porto de Galinhas	,0152
18º	Armação dos Búzios	,1393	18º	Porto Alegre	,0141
19º	Manaus	,1286	19º	Porto Seguro	,0129
20º	Porto Seguro	,1240	20º	Aracaju	,0113
21º	Belém	,1149	21º	João Pessoa	,0109
22º	João Pessoa	,1135	22º	Paraty	,0104
23º	Aracaju	,1079	23º	Fernando de Noronha	,0102
24º	Bonito	,1029	24º	Bonito	,0101
25º	Fernando de Noronha	,1026	25º	Jericoacoara	,0098
26º	Balneário Camboriú	,0911	26º	Belém	,0094
27º	Petrópolis	,0877	27º	Ubatuba	,0093
28º	Paraty	,0864	28º	Manaus	,0093
29º	Ubatuba	,0854	29º	Ouro Preto	,0089
30º	Jericoacoara	,0841	30º	Petrópolis	,0085
31º	Bento Gonçalves	,0803	31º	Ilhabela	,0078
32º	Santos	,0790	32º	Balneário Camboriú	,0077
33º	Ouro Preto	,0763	33º	Arraial do Cabo	,0072
34º	Arraial d'Ajuda	,0720	34º	Praia da Pipa	,0071
35º	Vitória	,0706	35º	Caldas Novas	,0069
36º	Niterói	,0695	36º	Maragogi	,0068
37º	Praia da Pipa	,0671	37º	Arraial d'Ajuda	,0068
38º	Ilhabela	,0642	38º	São Luís	,0067

Tabela 6: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ATRATIVOS (ATR) ponderadas em relação às proporções relativas de comentários atribuídos ao componente (NTC_ATR/NC_{máx}_ATR), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 3.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
39º	Goiânia	,0629	39º	Penha	,0067
40º	Maragogi	,0582	40º	Trancoso	,0066
41º	São Luís	,0580	41º	Niterói	,0063
42º	Arraial do Cabo	,0504	42º	Santos	,0062
43º	Caldas Novas	,0498	43º	Bento Gonçalves	,0061
44º	Penha	,0496	44º	Itacaré	,0061
45º	Itacaré	,0463	45º	Alto Paraíso de Goiás	,0058
46º	Cuiabá	,0454	46º	Vitória	,0056
47º	Joinville	,0440	47º	Goiânia	,0056
48º	Ilha Grande	,0438	48º	Ilha Grande	,0053
49º	Vila Velha	,0431	49º	Rio Quente	,0047
50º	Tiradentes	,0431	50º	Cuiabá	,0046
51º	Campinas	,0431	51º	Tiradentes	,0045
52º	Blumenau	,0426	52º	Bombinhas	,0045
53º	Bombinhas	,0412	53º	Joinville	,0044
54º	Poços de Caldas	,0411	54º	Pirenópolis	,0044
55º	Guarujá	,0411	55º	Campinas	,0044
56º	Morro de São Paulo	,0397	56º	Blumenau	,0043
57º	São Sebastião	,0384	57º	Olímpia	,0042
58º	Cabo Frio	,0383	58º	Poços de Caldas	,0042
59º	Trancoso	,0383	59º	Guarujá	,0042
60º	Alto Paraíso de Goiás	,0382	60º	Vila Velha	,0041
61º	Tamandaré	,0374	61º	Guarapari	,0040
62º	Praia do Forte	,0369	62º	Morro de São Paulo	,0040
63º	Pirenópolis	,0352	63º	Cabo Frio	,0039
64º	Olinda	,0341	64º	Praia do Forte	,0037
65º	Canoa Quebrada	,0311	65º	São Sebastião	,0036
66º	Ribeirão Preto	,0309	66º	Tamandaré	,0034
67º	Barreirinhas	,0302	67º	Barreirinhas	,0034
68º	Guarapari	,0296	68º	Monte Verde	,0031
69º	Olímpia	,0291	69º	Olinda	,0030
70º	Nova Petrópolis	,0269	70º	Lençóis	,0030
71º	Rio Quente	,0250	71º	Ribeirão Preto	,0028
72º	Campo Grande	,0247	72º	Canoa Quebrada	,0028
73º	Monte Verde	,0235	73º	Angra dos Reis	,0024
74º	Angra dos Reis	,0232	74º	Ilhéus	,0024
75º	Ilhéus	,0228	75º	Penedo	,0020
76º	Londrina	,0226	76º	Palmas	,0018

Tabela 6: (Continuação) Mostra os *rankings* dos destinos turísticos obtidos com a disposição decrescente dos valores das proporções de avaliações positivas ($pAvP$) e negativas ($pAvN$) para o componente ATRATIVOS (ATR) ponderadas em relação às proporções relativas de comentários atribuídos ao componente (NTC_ATR/NC_{máx}_ATR), considerado o conjunto dos 100 destinos turísticos analisados. A representação gráfica dos *rankings* encontra-se na Figura 3.

Rank	Destinos turísticos	$pAvP \bullet pCom$	Rank	Destinos turísticos	$pAvN \bullet pCom$
77º	Lençóis	,0192	77º	Campo Grande	,0018
78º	São Lourenço	,0191	78º	São Roque	,0017
79º	São Roque	,0185	79º	Londrina	,0016
80º	Palmas	,0177	80º	Teresópolis	,0016
81º	Teresina	,0175	81º	Nova Petrópolis	,0016
82º	Teresópolis	,0163	82º	São Lourenço	,0016
83º	Penedo	,0161	83º	Teresina	,0014
84º	Uberlândia	,0154	84º	Aquiraz	,0014
85º	Caxias do Sul	,0148	85º	Juiz de Fora	,0013
86º	Juiz de Fora	,0146	86º	São Bernardo do Campo	,0012
87º	Maringá	,0144	87º	Uberlândia	,0012
88º	São José dos Campos	,0139	88º	Guarulhos	,0011
89º	Sorocaba	,0118	89º	São José do Rio Preto	,0011
90º	Piracicaba	,0112	90º	Sorocaba	,0010
91º	São Bernardo do Campo	,0104	91º	Maringá	,0010
92º	Guarulhos	,0101	92º	São José dos Campos	,0009
93º	Santo André	,0098	93º	Caxias do Sul	,0008
94º	Aquiraz	,0094	94º	Santo André	,0008
95º	Jundiaí	,0085	95º	Piracicaba	,0008
96º	Bauru	,0076	96º	Bauru	,0007
97º	São José do Rio Preto	,0066	97º	Taubaté	,0007
98º	Taubaté	,0059	98º	Atibaia	,0006
99º	Atibaia	,0049	99º	Jundiaí	,0005
100º	Barueri	,0023	100º	Barueri	,0001

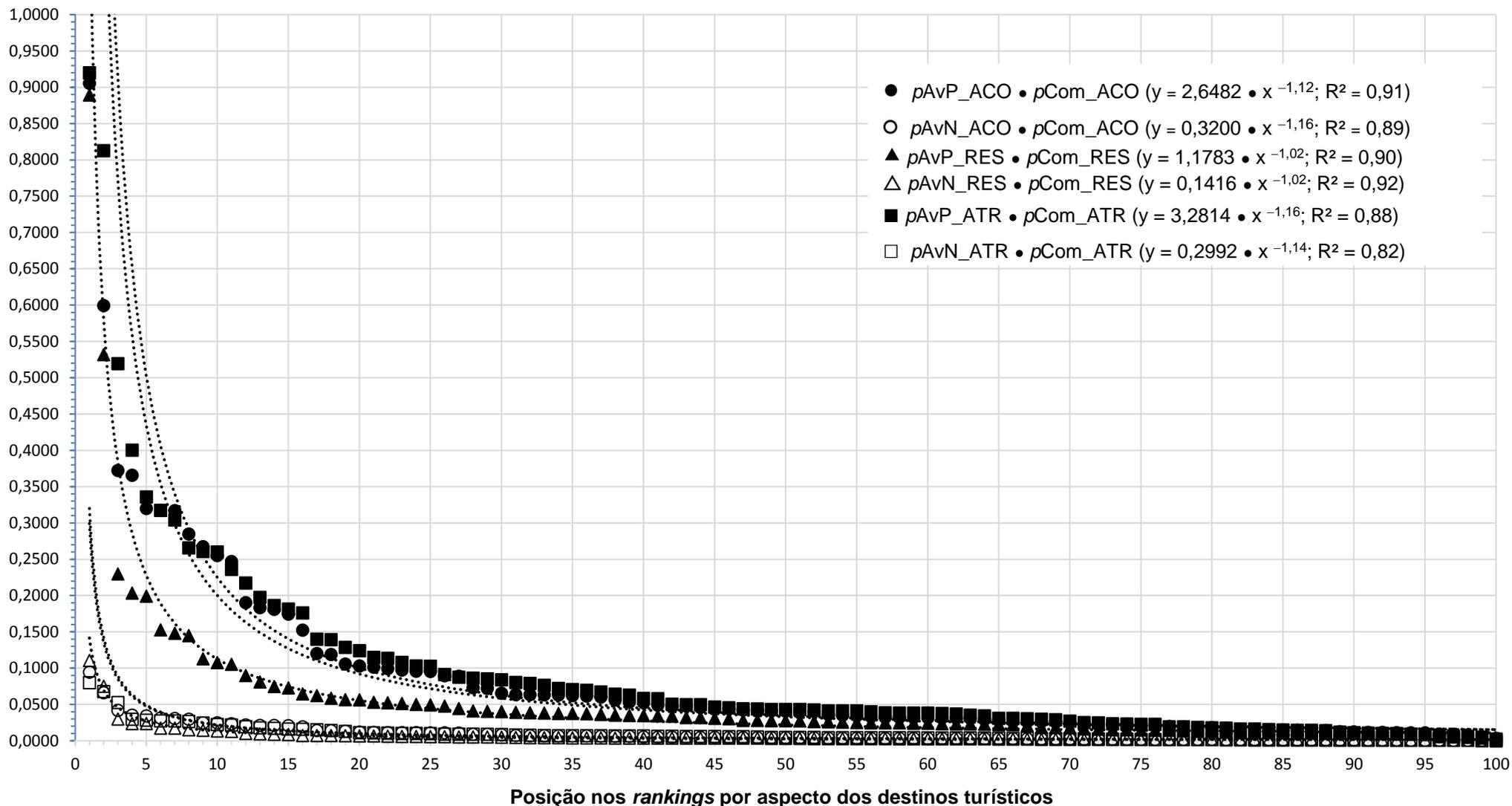


Figura 5: Mostra os valores das proporções de avaliações positivas (AvP; pontos pretos) e negativas (AvN; pontos brancos) para acomodações (ACO; círculos), restaurantes (RES; triângulos) e atrativos (ATR; quadrados) ponderados em relação às proporções relativas de comentários atribuídos a cada componente avaliado, considerado o conjunto dos 100 destinos turísticos analisados, conforme posição assumida por estes nos rankings por componente obtidos com a disposição decrescente dos valores. Os expoentes das equações de ajuste (funções potência) indicam a taxa média de decréscimo dos valores da variável a cada posição sucessiva. Curvas correspondentes aos rankings na Tabela 4, Tabela 5, Tabela 6, onde os nomes dos destinos e os valores específicos para cada destino podem ser observados.

Tabela 7: Mostra os coeficientes de correlação *rho* de Spearman entre os valores das proporções de avaliações positivas (AvP) para acomodações (ACO), restaurantes (RES) e atrativos (ATR) ponderados em relação às proporções relativas de comentários atribuídos a cada componente avaliado, considerado o conjunto dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%).

	$\rho_{AvP_ACO} \bullet \rho_{Com_ACO}$	$\rho_{AvP_RES} \bullet \rho_{Com_RES}$	$\rho_{AvP_ATR} \bullet \rho_{Com_ATR}$
$\rho_{AvP_ACO} \bullet \rho_{Com_ACO}$	---	,584**	,646**
$\rho_{AvP_RES} \bullet \rho_{Com_RES}$		---	,582**
$\rho_{AvP_ATR} \bullet \rho_{Com_ATR}$			---

Tabela 8: Mostra os coeficientes de correlação *rho* de Spearman entre os valores das proporções de avaliações negativas (AvN) para acomodações (ACO), restaurantes (RES) e atrativos (ATR) ponderados em relação às proporções relativas de comentários atribuídos a cada componente avaliado, considerado o conjunto dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%).

	$\rho_{AvN_ACO} \bullet \rho_{Com_ACO}$	$\rho_{AvN_RES} \bullet \rho_{Com_RES}$	$\rho_{AvN_ATR} \bullet \rho_{Com_ATR}$
$\rho_{AvN_ACO} \bullet \rho_{Com_ACO}$	---	,548**	,624**
$\rho_{AvN_RES} \bullet \rho_{Com_RES}$		---	,550**
$\rho_{AvN_ATR} \bullet \rho_{Com_ATR}$			---

A Tabela 4, Tabela 5 e Tabela 6, apresentadas anteriormente, mostram os *rankings* dos destinos turísticos (na ordem decrescente de valores), considerando a proporção relativa ponderada de avaliações positivas e negativas para os três componentes, Acomodações (ACO), Restaurantes (RES) e Atrações (ATR). Em relação ao componente ACO (Tabela 4), observou-se que os 15 destinos com os maiores valores no *ranking* diferem pouco dos destinos que estão nessas mesmas posições no *ranking* geral (Tabela 3), apenas Porto Alegre (12º) e Porto de Galinhas (14º), que aparecem em ACO, não estão no TOP15 do *ranking* geral, mas essa diferença não é significativa, pois esses destinos aparecem nesse *ranking* (Tabela 3) nas posições 17º e 16º, respectivamente; e nas avaliações negativas, o que difere em relação aos destinos turísticos do TOP15 do *ranking* geral é novamente Porto Alegre ([ACO, 13º], [Geral, 18º]) e Porto Seguro ([ACO, 15º], [Geral, 21º]).

No componente RES (Tabela 5), há mais diferenças em relação ao *ranking* geral (Tabela 3), especialmente pelo fato de que neste componente específico, Gramado figura nas últimas posições (nas avaliações positivas e negativas), enquanto no geral, Gramado está no TOP15. Campinas (11º) e Goiânia (15º) estão presentes no TOP15 somente no *ranking* do componente RES. Em Atrações (Tabela 6),

Gramado volta ao TOP15 de um componente específico. Nessa lista, o que difere do *ranking* geral (Tabela 3), nas avaliações positivas, é a inserção da capital Porto Alegre (15º) e, nas negativas, do destino turístico Canela (15º). Nos demais casos, evidenciou-se que o padrão presente no *ranking* geral das quinze primeiras posições (Tabela 3) permanece (não necessariamente na mesma ordem) mesmo quando essas avaliações estão segmentadas nos componentes.

Ao usar como parâmetro destinos turísticos que pertençam ao TOP5 de avaliações (positivas e negativas) nos três componentes, as duas primeiras posições permanecem intactas, São Paulo e Rio de Janeiro, esse último alcançando a primeira posição em avaliações positivas/negativas apenas na categoria Atração. Mas há uma mudança significativa nas três posições seguintes. Nas avaliações positivas, há destinos que aparecem somente em um dos componentes, por exemplo, Natal (3º, ACO), Brasília (3º, RES), Belo Horizonte (4º, RES), Fortaleza (5º, RES), Gramado (3º, ATR), nas negativas, o que há de exceção é Maceió, que aparece somente na tabela do componente ATR na 5ª. posição.

Uma interpretação possível sobre a cidade de Gramado figurar nas últimas posições no componente Restaurantes (RES) e estar presente nas primeiras posições gerais e, especialmente, no TOP5 de Atrações (ATR), pode ter relação com o fato dos turistas, ao emitir respostas verbais sobre os ambientes de alimentação de Gramado, considerarem mais representativos as atrações do lugar do que um restaurante específico, talvez pelos reforços informativos que são potencializados pelas diversas mídias (programas de TV, redes sociais, reportagens), que aumentam a probabilidade de ocorrência de avaliações (positivas em sua maioria e algumas poucas negativas) voltadas para esse componente. Assim, os turistas tendem a emitir respostas verbais em número significativamente maior para os locais considerados, por exemplo, pontos turísticos (as atrações), do que especificamente para os restaurantes desses locais. É como se os restaurantes, sem o contexto da atração, não funcionassem discriminativamente para a emissão de comentários.

A Figura 5 apresenta a distribuição dos dados da Tabela 4, Tabela 5 e Tabela 6 na forma gráfica, com as curvas de dados que representam a proporção relativa ponderada de avaliações (positivas e negativas) para cada um dos componentes (ACO, RES, ATR). Ainda que as alturas iniciais das curvas sejam diferentes, pois os primeiros destinos de cada componente (positivo/negativo) têm valores de proporção relativa ponderada bem superiores aos demais, os expoentes, que são a medida

média dos decréscimos dos valores, são bem similares, logo a forma que cada curva se horizontaliza é semelhante. No ponto mais alto de todas as curvas está o componente ATR, com o Rio de Janeiro atingindo o maior valor da proporção relativa ponderada de avaliações positivas (0,9199), seguido por São Paulo (0,9054) no componente ACO e, também, no componente RES com o valor de 0,8894. A partir disso, há uma queda considerável nos valores, com Gramado (0,5193) no componente ATR e, depois, Natal (0,3723) do componente ACO. Pelo que pode ser observado ainda no contexto das avaliações positivas na Figura 5, na terceira curva é onde se encontram os pontos relacionados ao componente RES (avaliações positivas), que se afastam consideravelmente das duas outras curvas, onde os dados relacionados aos componentes ATR e ACO estão presentes. Para o cálculo da AvP do componente RES, foi feita a ponderação em relação às proporções relativas de comentários atribuídos ao componente. Por São Paulo ser um destino turístico com uma abundância de variáveis ambientais contextuais para o cenário do consumo do turista, tem um total de 814.001 comentários (a maior quantidade verificada), e destes, 502.575 são do componente RES. Assim, a ponderação dos valores com o número de comentário máximo nesse seguimento, fez com que parte dos valores finais por destino turístico de RES, considerando $pAVP[RES] \bullet pCom[RES]$, ficassem com valores inferiores aos demais componentes (ao menos na parte destacada antes da horizontalização da curva, quando os valores tendem a ficar mais homogêneos). Uma outra situação que merece destaque é o fato de que no componente Atração (ATR), o Rio de Janeiro mostrou que pode ser um cenário de consumo mais aberto que São Paulo, e isso pode estar relacionado a fatores que promovem certo *status* social e que servem como reforço informativo. Nesse sentido, o número de celebridades em suas praias, ter um de seus eventos apontados pela mídia como a maior festa do planeta, paisagens e monumentos considerados cartão postal do Brasil e o fato do Rio de Janeiro ser um dos locais mais usados em cenários de novelas, etc. parecem atenuar as características aversivas presentes nos informes jornalísticos, que evidenciam a violência e a falta de segurança da cidade.

Nas avaliações negativas representadas na Figura 5, tem-se no ponto mais alto da curva, o destino turístico São Paulo ($[pAvN \bullet pCom = 0,1106]$, $[pAvN \bullet pCom = 0,0946]$), nos componentes RES e ACO, respectivamente, e, depois pela cidade do Rio de Janeiro ($pAvN \bullet pCom = 0,0801$), no componente ATR. Rapidamente, as curvas horizontalizam, com os dados relacionados ao componente RES apresentando

os valores mais baixos, enquanto as curvas com os dados de ACO e ATR praticamente se sobrepõem, o que reforça uma certa similaridade no comportamento dos dados na curva.

As tabelas 7 e 8 mostram os coeficientes ρ de Spearman entre os valores das proporções positivas (AvP) e negativas (AvN) para os três componentes de produtos turísticos, Acomodações (ACO), Restaurantes (RES) e Atrações (ATR), ponderados em relação às proporções relativas de comentários atribuídos a cada aspecto avaliado nos 100 destinos turísticos analisados. Na análise dos valores resultantes, verifica-se que há uma associação moderada entre os valores das proporções de avaliações por componentes. O que permite a inferência de que, nesse caso, avaliar positivamente acomodações, por exemplo, não implica necessariamente em avaliações positivas de atrações, o que sugere certa independência dos processos de controles diferenciais sobre as respostas de avaliar.

Os *likes*, que são interpretados no *TripAdvisor*® como votos úteis, estão sistematizados nas Tabela 9 e Figura 6, a seguir.

Tabela 9: Mostra, para cada um dos 100 destinos turísticos analisados, o número total de *likes* (*nTLikes*) atribuídos aos comentários, o número total de comentários (*nTC*), a proporção de *likes* por comentário ($nTLikes/nTC$), a proporção de comentários em relação ao número máximo de comentários observado ($nTC/nC_{máx}$) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	nTLikes	nTC	nTLikes/nTC	nTC/nC_{máx}	wpLikes
1º	São Paulo	222.505	814.001	0,2733	1,0000	0,2733
2º	Rio de Janeiro	187.773	607.914	0,3089	0,7468	0,2307
3º	Recife	170.946	84.751	2,0170	0,1041	0,2100
4º	Natal	95.490	191.495	0,4987	0,2353	0,1173
5º	Maceió	92.110	154.488	0,5962	0,1898	0,1132
6º	Fortaleza	85.626	227.108	0,3770	0,2790	0,1052
7º	Brasília	74.663	243.462	0,3067	0,2991	0,0917
8º	Belo Horizonte	72.932	210.469	0,3465	0,2586	0,0896
9º	Foz do Iguaçu	66.860	168.888	0,3959	0,2075	0,0821
10º	Curitiba	65.881	182.330	0,3613	0,2240	0,0809
11º	Campos do Jordão	61.884	120.247	0,5146	0,1477	0,0760
12º	Gramado	60.429	169.910	0,3557	0,2087	0,0742
13º	Salvador	60.032	179.272	0,3349	0,2202	0,0737
14º	Porto de Galinhas	58.309	86.125	0,6770	0,1058	0,0716
15º	Florianópolis	57.826	163.193	0,3543	0,2005	0,0710
16º	Porto Seguro	56.377	68.157	0,8272	0,0837	0,0693
17º	Santos	44.035	51.291	0,8585	0,0630	0,0541
18º	João Pessoa	43.809	93.269	0,4697	0,1146	0,0538
19º	Armação dos Búzios	37.258	82.828	0,4498	0,1018	0,0458
20º	Maragogi	36.161	35.733	1,0120	0,0439	0,0444
21º	Aracaju	33.056	71.588	0,4618	0,0879	0,0406
22º	Balneário Camboriú	29.678	64.066	0,4632	0,0787	0,0365
23º	Campinas	27.003	81.386	0,3318	0,1000	0,0332
24º	Caldas Novas	25.431	33.586	0,7572	0,0413	0,0312
25º	Monte Verde	25.072	33.313	0,7526	0,0409	0,0308
26º	Praia do Forte	24.477	23.021	1,0632	0,0283	0,0301
27º	Canela	22.907	74.247	0,3085	0,0912	0,0281
28º	Manaus	22.853	79.406	0,2878	0,0976	0,0281
29º	Poços de Caldas	22.357	28.354	0,7885	0,0348	0,0275
30º	Porto Alegre	22.221	83.739	0,2654	0,1029	0,0273
31º	Arraial d'Ajuda	21.686	38.668	0,5608	0,0475	0,0266
32º	Ilhabela	20.918	41.972	0,4984	0,0516	0,0257
33º	Jericoacoara	20.344	55.956	0,3636	0,0687	0,0250
34º	Paraty	20.102	48.528	0,4142	0,0596	0,0247
35º	Goiânia	19.272	68.432	0,2816	0,0841	0,0237
36º	Bento Gonçalves	19.138	47.147	0,4059	0,0579	0,0235
37º	Ubatuba	18.910	38.139	0,4958	0,0469	0,0232
38º	Cabo Frio	18.718	25.588	0,7315	0,0314	0,0230

Tabela 9: (Continuação) Mostra, para cada um dos 100 destinos turísticos analisados, o número total de *likes* (*nTLikes*) atribuídos aos comentários, o número total de comentários (*nTC*), a proporção de *likes* por comentário (*nTLikes/nTC*), a proporção de comentários em relação ao número máximo de comentários observado (*nTC/nC_{máx}*) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	nTLikes	nTC	nTLikes/nTC	nTC/nC _{máx}	wpLikes
39º	Belém	18.270	70.177	0,2603	0,0862	0,0224
40º	Bonito	17.877	48.327	0,3699	0,0594	0,0220
41º	Vitória	16.901	59.179	0,2856	0,0727	0,0208
42º	Guarujá	15.823	28.218	0,5607	0,0347	0,0194
43º	Ilhéus	15.780	16.815	0,9384	0,0207	0,0194
44º	Petrópolis	15.349	47.451	0,3235	0,0583	0,0189
45º	Pirenópolis	15.226	21.619	0,7043	0,0266	0,0187
46º	Tiradentes	15.071	33.123	0,4550	0,0407	0,0185
47º	Olímpia	14.625	22.323	0,6552	0,0274	0,0180
48º	Morro de São Paulo	14.362	26.703	0,5378	0,0328	0,0176
49º	São Luís	14.139	45.446	0,3111	0,0558	0,0174
50º	Angra dos Reis	13.824	16.294	0,8484	0,0200	0,0170
51º	Praia da Pipa	13.682	34.675	0,3946	0,0426	0,0168
52º	Olinda	12.989	16.697	0,7779	0,0205	0,0160
53º	Ribeirão Preto	12.777	50.063	0,2552	0,0615	0,0157
54º	Ouro Preto	12.588	37.500	0,3357	0,0461	0,0155
55º	Niterói	12.429	49.147	0,2529	0,0604	0,0153
56º	Atibaia	12.123	15.298	0,7925	0,0188	0,0149
57º	Rio Quente	11.924	14.745	0,8087	0,0181	0,0146
58º	São José dos Campos	11.902	32.339	0,3680	0,0397	0,0146
59º	Blumenau	11.473	29.052	0,3949	0,0357	0,0141
60º	Itacaré	11.435	24.848	0,4602	0,0305	0,0140
61º	Aquiraz	11.422	10.310	1,1079	0,0127	0,0140
62º	Joinville	11.343	36.643	0,3096	0,0450	0,0139
63º	Cuiabá	11.050	37.547	0,2943	0,0461	0,0136
64º	Teresópolis	10.578	21.919	0,4826	0,0269	0,0130
65º	Campo Grande	10.520	32.654	0,3222	0,0401	0,0129
66º	Trancoso	10.077	19.328	0,5214	0,0237	0,0124
67º	São Sebastião	9.931	20.731	0,4790	0,0255	0,0122
68º	Fernando de Noronha	9.715	48.265	0,2013	0,0593	0,0119
69º	Penha	9.633	20.413	0,4719	0,0251	0,0118
70º	Alto Paraíso de Goiás	9.454	21.313	0,4436	0,0262	0,0116
71º	Bombinhas	9.398	20.826	0,4513	0,0256	0,0115
72º	São Lourenço	9.188	13.638	0,6737	0,0168	0,0113
73º	Arraial do Cabo	9.028	21.573	0,4185	0,0265	0,0111
74º	Londrina	8.655	32.855	0,2634	0,0404	0,0106
75º	São Roque	8.578	13.639	0,6289	0,0168	0,0105

Tabela 9: (Continuação) Mostra, para cada um dos 100 destinos turísticos analisados, o número total de *likes* (*nTLikes*) atribuídos aos comentários, o número total de comentários (*nTC*), a proporção de *likes* por comentário (*nTLikes/nTC*), a proporção de comentários em relação ao número máximo de comentários observado (*nTC/nC_{máx}*) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	nTLikes	nTC	nTLikes/nTC	nTC/nC _{máx}	wpLikes
76º	Penedo	8.235	19.484	0,4227	0,0239	0,0101
77º	Guarulhos	8.178	27.503	0,2973	0,0338	0,0100
78º	Guarapari	7.986	15.371	0,5195	0,0189	0,0098
79º	Jundiá	7.674	21.104	0,3636	0,0259	0,0094
80º	Tamandaré	7.303	15.933	0,4584	0,0196	0,0090
81º	Juiz de Fora	7.084	26.431	0,2680	0,0325	0,0087
82º	Caxias do Sul	6.980	19.500	0,3579	0,0240	0,0086
83º	Maringá	6.917	20.111	0,3439	0,0247	0,0085
84º	Sorocaba	6.841	20.634	0,3315	0,0253	0,0084
85º	São Bernardo do Campo	6.800	18.790	0,3619	0,0231	0,0084
86º	Piracicaba	6.580	17.144	0,3838	0,0211	0,0081
87º	Canoa Quebrada	6.416	16.417	0,3908	0,0202	0,0079
88º	Vila Velha	5.976	25.429	0,2350	0,0312	0,0073
89º	São José do Rio Preto	5.678	20.062	0,2830	0,0246	0,0070
90º	Bauru	5.511	12.473	0,4418	0,0153	0,0068
91º	Santo André	5.478	23.503	0,2331	0,0289	0,0067
92º	Ilha Grande	5.463	15.742	0,3470	0,0193	0,0067
93º	Teresina	5.248	20.208	0,2597	0,0248	0,0064
94º	Lençóis	5.248	12.816	0,4095	0,0157	0,0064
95º	Barreirinhas	5.096	12.542	0,4063	0,0154	0,0063
96º	Taubaté	5.027	12.250	0,4104	0,0150	0,0062
97º	Uberlândia	4.490	21.189	0,2119	0,0260	0,0055
98º	Nova Petrópolis	4.009	13.855	0,2894	0,0170	0,0049
99º	Palmas	3.738	12.438	0,3005	0,0153	0,0046
100º	Barueri	3.214	13.914	0,2310	0,0171	0,0039
	Σ =	2.675.358	6.438.603			
	ME =	13.753	32.497			
	MD =	26.754	64.386,03			
	DP =	36.420,12	106.835,28			
	Mín =	3.214	10.310			
	Máx =	222.505	814.001			

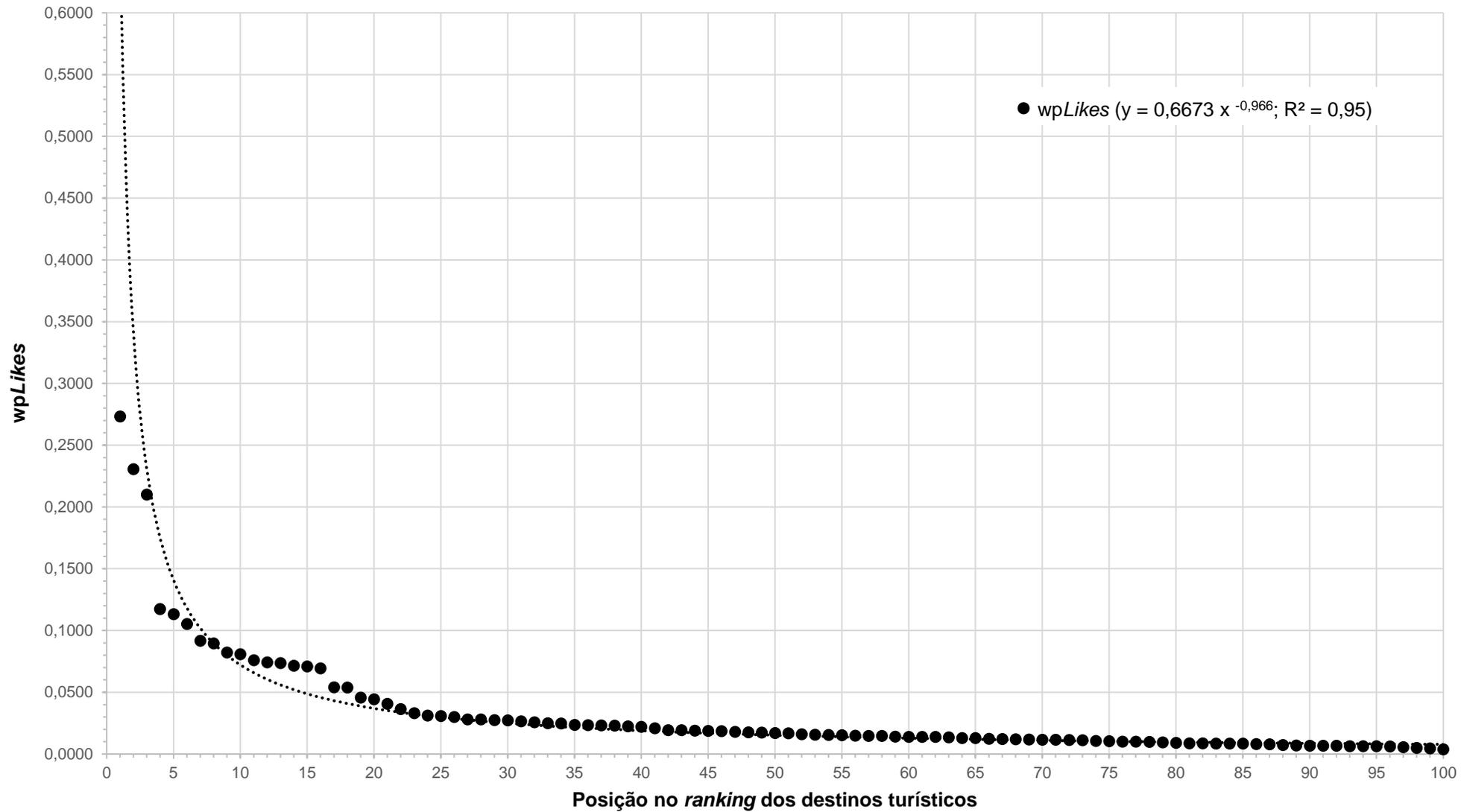


Figura 6: Mostra os valores da proporção relativa ponderada de *likes* para comentários sobre cada um dos 100 destinos turísticos analisados, conforme posição assumida por estes no *ranking* obtido com a disposição decrescente dos valores (Tabela 9).

Os dados apresentados da Tabela 9 e Figura 6 mostram, de modo geral, a mesma tendência das representações anteriores: há um conjunto de 10 a 15 destinos que induziram significativamente mais avaliações e, nos dados acima, mais *likes* sobre avaliações que os demais. Por exemplo, todos os 15 destinos com mais *likes* também estão presentes no *ranking* geral (Tabela 3, avaliações positivas). Há uma grande diferença no valor *wpLikes* entre os destinos nas primeiras posições (São Paulo, Rio de Janeiro, Recife) com o da quarta posição, Natal, ou seja, os 3 primeiros tem uma quantidade superior a 170.000 *likes* ($wpLikes > 0,2$), enquanto que Natal tem uma quantidade inferior a 100.000 ($wpLikes < 0,12$). Na 17ª posição, voltará a ocorrer uma diferença considerável no valor *wpLikes*, quando Porto Seguro (16º, $nTLikes = 56.377$, $wpLikes = 0,071$) antecede Santos (17º, $nTLikes = 44.035$, $wpLikes = 0,0541$). Uma situação fora do padrão, que é evidenciada nessas representações, é o fato de Aquiraz (que está na última posição no *ranking* geral) ocupar a 61ª posição no *ranking* de *likes*, além da quantidade de *likes* ser superior à quantidade de comentários nesse destino. Aquiraz, diferente de outros destinos turísticos do Nordeste, não é tão exposta na mídia, mas por ficar tão próxima à Fortaleza (cerca de 37 km), é especialmente relevante para os turistas que tem como destino a capital, pois além desse fato, uma de suas praias, a Porto das Dunas, abriga o maior parque aquático do Brasil. Assim, seus comentários, ainda que poucos, podem produzir respostas de consumo em quem os lê. A ação de hospedar-se em um hotel da cidade cujo comentário foi positivo em relação a alguns aspectos importantes para o turista, pode aumentar a probabilidade do comentário ganhar um voto útil, dado as suas especificidades e ao tipo de estímulo que levou o turista a acessar informações sobre este local no *TripAdvisor*®. Mas, ainda assim, o reforço informativo maior para a ação de emitir ou ler um comentário pode ser a visibilidade do destino turístico. Assim, é mais provável que destinos como São Paulo, Natal, Rio de Janeiro e Fortaleza, que apresentam um maior número de variáveis ambientais, sejam mais lidos, o que aumenta a probabilidade de vir a ganhar um voto útil.

As próximas tabelas e gráficos consideram a análise dos *likes* por componente do produto turístico isoladamente (ACO, RES, ATR).

Tabela 10: Mostra, considerado o aspecto ACOMODAÇÕES (ACO) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários (*nTLikes_ACO*), o número total de comentários (*nTC_ACO*), a proporção de *likes* por comentário (*nTLikes_ACO/nTC_ACO*), a proporção de comentários em relação ao número máximo de comentários observado (*nTC_ACO/nC_{máx}*) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes_ACO*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes_ACO*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes_ACO*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	ACOMODAÇÕES (ACO)				
		<i>nTLikes</i>	<i>nTC</i>	<i>nTLikes/nTC</i>	<i>nTC/nC_{máx}</i>	<i>wpLikes</i>
1º	Natal	60459	44736	1,3515	,4142	,5598
2º	Rio de Janeiro	57133	71885	,7948	,6656	,5290
3º	São Paulo	52753	108004	,4884	1,0000	,4884
4º	Maceió	49644	30227	1,6424	,2799	,4596
5º	Foz do Iguaçu	40666	43172	,9420	,3997	,3765
6º	Porto de Galinhas	39955	21817	1,8314	,2020	,3699
7º	Fortaleza	39411	38090	1,0347	,3527	,3649
8º	Porto Seguro	39218	18691	2,0982	,1731	,3631
9º	Gramado	34082	36772	,9268	,3405	,3156
10º	Maragogi	26987	11989	2,2510	,1110	,2499
11º	Salvador	22119	30037	,7364	,2781	,2048
12º	Curitiba	21944	37875	,5794	,3507	,2032
13º	João Pessoa	20665	14268	1,4483	,1321	,1913
14º	Armação dos Búzios	19946	14485	1,3770	,1341	,1847
15º	Florianópolis	18682	22184	,8421	,2054	,1730
16º	Caldas Novas	18178	11168	1,6277	,1034	,1683
17º	Recife	17509	20904	,8376	,1935	,1621
18º	Campos do Jordão	16264	11481	1,4166	,1063	,1506
19º	Belo Horizonte	15268	31537	,4841	,2920	,1414
20º	Balneário Camboriú	15235	12702	1,1994	,1176	,1411
21º	Praia do Forte	14904	7792	1,9127	,0721	,1380
22º	Aracaju	14509	12156	1,1936	,1126	,1343
23º	Monte Verde	14076	4804	2,9301	,0445	,1303
24º	Ilhéus	12518	5827	2,1483	,0540	,1159
25º	Brasília	12415	33940	,3658	,3142	,1149
26º	Poços de Caldas	11590	7588	1,5274	,0703	,1073
27º	Arraial d'Ajuda	11403	5004	2,2788	,0463	,1056
28º	Angra dos Reis	10252	5349	1,9166	,0495	,0949
29º	Jericoacoara	9509	6557	1,4502	,0607	,0880
30º	Aquiraz	9466	5689	1,6639	,0527	,0876
31º	Olímpia	9399	9001	1,0442	,0833	,0870
32º	Ilhabela	9062	6112	1,4827	,0566	,0839
33º	Porto Alegre	8926	22828	,3910	,2114	,0826
34º	Rio Quente	8374	5203	1,6095	,0482	,0775
35º	Atibaia	8323	7221	1,1526	,0669	,0771

Tabela 10: (Continuação) Mostra, considerado o aspecto ACOMODAÇÕES (ACO) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários ($nTLikes_ACO$), o número total de comentários (nTC_ACO), a proporção de *likes* por comentário ($nTLikes_ACO/nTC_ACO$), a proporção de comentários em relação ao número máximo de comentários observado ($nTC_ACO/nC_{máx}$) e a proporção de *likes* ponderada pelo número proporcional de comentários ($wpLikes_ACO$). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores $wpLikes_ACO$. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis ($nTLikes_ACO$) e nTC , são mostrados nas linhas finais da tabela.

Rank	Destino turístico	ACOMODAÇÕES (ACO)				
		$nTLikes$	nTC	$nTLikes/nTC$	$nTC/nC_{máx}$	$wpLikes$
36º	Guarujá	6915	5688	1,2157	,0527	,0640
37º	Paraty	6764	4116	1,6433	,0381	,0626
38º	Pirenópolis	6626	3758	1,7632	,0348	,0613
39º	Cabo Frio	6509	3251	2,0022	,0301	,0603
40º	Manaus	6384	11463	,5569	,1061	,0591
41º	Morro de São Paulo	6296	3479	1,8097	,0322	,0583
42º	Vitória	6276	12249	,5124	,1134	,0581
43º	Ubatuba	6046	2713	2,2285	,0251	,0560
44º	Belém	5771	8857	,6516	,0820	,0534
45º	São Lourenço	5762	3171	1,8171	,0294	,0533
46º	Bento Gonçalves	5737	7277	,7884	,0674	,0531
47º	Goiânia	5703	11584	,4923	,1073	,0528
48º	São Luís	5639	7549	,7470	,0699	,0522
49º	Itacaré	5114	2745	1,8630	,0254	,0474
50º	Bonito	4861	4254	1,1427	,0394	,0450
51º	Campinas	4776	10764	,4437	,0997	,0442
52º	Canela	4705	7404	,6355	,0686	,0436
53º	Praia da Pipa	4156	3761	1,1050	,0348	,0385
54º	Trancoso	4031	2334	1,7271	,0216	,0373
55º	Tiradentes	3782	2690	1,4059	,0249	,0350
56º	Teresópolis	3763	3783	,9947	,0350	,0348
57º	Bombinhas	3616	2065	1,7511	,0191	,0335
58º	Petrópolis	3575	3544	1,0087	,0328	,0331
59º	Santos	3395	4816	,7049	,0446	,0314
60º	Guarulhos	3346	7751	,4317	,0718	,0310
61º	São Sebastião	3300	2737	1,2057	,0253	,0306
62º	Canoa Quebrada	3266	3616	,9032	,0335	,0302
63º	Arraial do Cabo	3207	1444	2,2209	,0134	,0297
64º	Alto Paraíso de Goiás	3151	4810	,6551	,0445	,0292
65º	Bauru	2985	1788	1,6695	,0166	,0276
66º	Ouro Preto	2867	2555	1,1221	,0237	,0265
67º	Joinville	2574	6238	,4126	,0578	,0238
68º	Ribeirão Preto	2567	7386	,3475	,0684	,0238
69º	Blumenau	2481	4162	,5961	,0385	,0230
70º	Taubaté	2369	2486	,9529	,0230	,0219

Tabela 10: (Continuação) Mostra, considerado o aspecto ACOMODAÇÕES (ACO) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários (*nTLikes_ACO*), o no. total de comentários (*nTC_ACO*), a proporção de *likes* por comentário (*nTLikes_ACO/nTC_ACO*), a proporção de comentários em relação ao no. máximo de comentários observado (*nTC_ACO/nC_{máx}*) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes_ACO*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes_ACO*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes_ACO*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	ACOMODAÇÕES (ACO)				
		<i>nTLikes</i>	<i>nTC</i>	<i>nTLikes/nTC</i>	<i>nTC/nC_{máx}</i>	<i>wpLikes</i>
71º	Campo Grande	2301	5773	,3986	,0535	,0213
72º	Guarapari	2258	1365	1,6542	,0126	,0209
73º	Juiz de Fora	2172	3841	,5655	,0356	,0201
74º	Penha	2162	1356	1,5944	,0126	,0200
75º	Barreirinhas	2102	1728	1,2164	,0160	,0195
76º	Ilha Grande	2084	1071	1,9458	,0099	,0193
77º	São Roque	2031	1442	1,4085	,0134	,0188
78º	São José dos Campos	2026	3963	,5112	,0367	,0188
79º	Fernando de Noronha	1911	1236	1,5461	,0114	,0177
80º	São José do Rio Preto	1862	3184	,5848	,0295	,0172
81º	Cuiabá	1854	5251	,3531	,0486	,0172
82º	Londrina	1803	5192	,3473	,0481	,0167
83º	Lençóis	1794	2122	,8454	,0196	,0166
84º	Maringá	1724	3802	,4534	,0352	,0160
85º	Caxias do Sul	1604	3101	,5173	,0287	,0149
86º	Penedo	1371	1020	1,3441	,0094	,0127
87º	Uberlândia	1345	3817	,3524	,0353	,0125
88º	Palmas	1325	1824	,7264	,0169	,0123
89º	Niterói	1299	1709	,7601	,0158	,0120
90º	Sorocaba	1169	2336	,5004	,0216	,0108
91º	Olinda	1126	929	1,2121	,0086	,0104
92º	Tamandaré	1089	564	1,9309	,0052	,0101
93º	Vila Velha	1055	1876	,5624	,0174	,0098
94º	Teresina	874	2186	,3998	,0202	,0081
95º	Piracicaba	842	1849	,4554	,0171	,0078
96º	Barueri	767	3129	,2451	,0290	,0071
97º	Jundiaí	733	1685	,4350	,0156	,0068
98º	Nova Petrópolis	705	1233	,5718	,0114	,0065
99º	Santo André	376	1340	,2806	,0124	,0035
100º	São Bernardo do Campo	355	447	,7942	,0041	,0033
$\Sigma =$		993.278	1.025.754			
ME =		4.818,50	4.813,00			
MD =		9.932,78	10.257,54			
DP =		13.022,15	15.478,34			
Mín =		355	447			
Máx =		60.459	108.004			

Tabela 11: Mostra, considerado o aspecto RESTAURANTES (RES) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários ($nTLikes_RES$), o número total de comentários (nTC_RES), a proporção de *likes* por comentário ($nTLikes_RES / nTC_RES$), a proporção de comentários em relação ao número máximo de comentários observado ($nTC_RES / nC_{máx}$) e a proporção de *likes* ponderada pelo número proporcional de comentários ($wpLikes_RES$). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores $wpLikes_RES$. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis ($nTLikes_RES$) e nTC , são mostrados nas linhas finais da tabela.

Rank	Destino turístico	RESTAURANTES (RES)				
		$nTLikes$	nTC	$nTLikes/nTC$	$nTC/nC_{máx}$	$wpLikes$
1º	São Paulo	144628	502575	,2878	1,0000	,2878
2º	Rio de Janeiro	97170	305038	,3186	,6070	,1933
3º	Recife	56982	14589	3,9058	,0290	,1134
4º	Brasília	48045	130562	,3680	,2598	,0956
5º	Belo Horizonte	44416	114175	,3890	,2272	,0884
6º	Santos	37026	26782	1,3825	,0533	,0737
7º	Campos do Jordão	36599	61737	,5928	,1228	,0728
8º	Florianópolis	29771	85576	,3479	,1703	,0592
9º	Fortaleza	29371	111911	,2624	,2227	,0584
10º	Salvador	21888	81532	,2685	,1622	,0436
11º	Maceió	21244	63150	,3364	,1257	,0423
12º	Natal	20167	81706	,2468	,1626	,0401
13º	Campinas	20158	59651	,3379	,1187	,0401
14º	João Pessoa	14723	50272	,2929	,1000	,0293
15º	Foz do Iguaçu	14089	42006	,3354	,0836	,0280
16º	Curitiba	13978	45640	,3063	,0908	,0278
17º	Armação dos Búzios	12321	32040	,3846	,0638	,0245
18º	Porto de Galinhas	10488	28460	,3685	,0566	,0209
19º	Aracaju	10241	31897	,3211	,0635	,0204
20º	Goiânia	9941	41029	,2423	,0816	,0198
21º	Balneário Camboriú	9893	28536	,3467	,0568	,0197
22º	Manaus	9487	36098	,2628	,0718	,0189
23º	Monte Verde	9338	22374	,4174	,0445	,0186
24º	São José dos Campos	9329	24955	,3738	,0497	,0186
25º	Ribeirão Preto	8835	34906	,2531	,0695	,0176
26º	Canela	8667	22432	,3864	,0446	,0172
27º	Tiradentes	8461	19429	,4355	,0387	,0168
28º	Niterói	7922	29929	,2647	,0596	,0158
29º	Jericoacoara	7901	27715	,2851	,0551	,0157
30º	Porto Seguro	7806	17858	,4371	,0355	,0155
31º	Petrópolis	7784	21680	,3590	,0431	,0155
32º	Belém	7507	32628	,2301	,0649	,0149
33º	Porto Alegre	7407	15793	,4690	,0314	,0147
34º	Campo Grande	7129	20763	,3434	,0413	,0142
35º	Praia da Pipa	6992	13778	,5075	,0274	,0139

Tabela 11:(Continuação) Mostra, considerado o aspecto RESTAURANTES (RES) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários (*nTLikes_RES*), o número total de comentários (*nTC_RES*), a proporção de *likes* por comentário ($nTLikes_RES / nTC_RES$), a proporção de comentários em relação ao número máximo de comentários observado ($nTC_RES / nC_{máx}$) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes_RES*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes_RES*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes_RES*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	RESTAURANTES (RES)				
		<i>nTLikes</i>	<i>nTC</i>	$nTLikes/nTC$	$nTC/nC_{máx}$	<i>wpLikes</i>
36º	Ilhabela	6849	19218	,3564	,0382	,0136
37º	Vitória	6842	29341	,2332	,0584	,0136
38º	Paraty	6824	22061	,3093	,0439	,0136
39º	Jundiaí	6653	17348	,3835	,0345	,0132
40º	Bento Gonçalves	6418	19914	,3223	,0396	,0128
41º	Arraial d'Ajuda	6172	15466	,3991	,0308	,0123
42º	Cabo Frio	5970	12599	,4738	,0251	,0119
43º	Londrina	5890	22055	,2671	,0439	,0117
44º	Penedo	5767	14281	,4038	,0284	,0115
45º	São Bernardo do Campo	5527	15666	,3528	,0312	,0110
46º	Cuiabá	5513	20750	,2657	,0413	,0110
47º	Piracicaba	5200	12522	,4153	,0249	,0103
48º	Blumenau	5168	14067	,3674	,0280	,0103
49º	Teresópolis	5115	13989	,3656	,0278	,0102
50º	São Luís	5109	22953	,2226	,0457	,0102
51º	Sorocaba	5099	15333	,3326	,0305	,0101
52º	Joinville	4988	19223	,2595	,0382	,0099
53º	Santo André	4779	19703	,2426	,0392	,0095
54º	Ouro Preto	4775	15264	,3128	,0304	,0095
55º	Maringá	4563	12761	,3576	,0254	,0091
56º	Guarulhos	4472	17158	,2606	,0341	,0089
57º	Caxias do Sul	4395	12784	,3438	,0254	,0087
58º	Ubatuba	4361	13552	,3218	,0270	,0087
59º	Bonito	4327	17962	,2409	,0357	,0086
60º	Pirenópolis	4206	8701	,4834	,0173	,0084
61º	Juiz de Fora	4172	18906	,2207	,0376	,0083
62º	Fernando de Noronha	4018	20975	,1916	,0417	,0080
63º	São Roque	3954	7526	,5254	,0150	,0079
64º	Guarujá	3634	12070	,3011	,0240	,0072
65º	São José do Rio Preto	3588	15118	,2373	,0301	,0071
66º	Teresina	3554	13641	,2605	,0271	,0071
67º	Itacaré	3485	10004	,3484	,0199	,0069
68º	Atibaia	3329	6814	,4886	,0136	,0066
69º	Poços de Caldas	3315	10304	,3217	,0205	,0066
70º	Morro de São Paulo	3279	13125	,2498	,0261	,0065

Tabela 11:(Continuação) Mostra, considerado o aspecto RESTAURANTES (RES) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários (*nTLikes_RES*), o número total de comentários (*nTC_RES*), a proporção de *likes* por comentário (*nTLikes_RES/nTC_RES*), a proporção de comentários em relação ao número máximo de comentários observado (*nTC_RES/nC_{máx}*) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes_RES*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes_RES*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes_RES*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	RESTAURANTES (RES)				
		<i>nTLikes</i>	<i>nTC</i>	<i>nTLikes/nTC</i>	<i>nTC/nC_{máx}</i>	<i>wpLikes</i>
71º	Maragogi	3264	8733	,3738	,0174	,0065
72º	Bombinhas	3165	8194	,3863	,0163	,0063
73º	Guarapari	2804	6235	,4497	,0124	,0056
74º	São Sebastião	2800	8301	,3373	,0165	,0056
75º	Trancoso	2788	6639	,4199	,0132	,0055
76º	Caldas Novas	2656	9314	,2852	,0185	,0053
77º	Arraial do Cabo	2564	6826	,3756	,0136	,0051
78º	Uberlândia	2488	13540	,1838	,0269	,0050
79º	Vila Velha	2483	12652	,1963	,0252	,0049
80º	Barueri	2394	10211	,2345	,0203	,0048
81º	Taubaté	2312	8261	,2799	,0164	,0046
82º	Tamandaré	2151	5949	,3616	,0118	,0043
83º	Alto Paraíso de Goiás	2143	6341	,3380	,0126	,0043
84º	Lençóis	2142	5568	,3847	,0111	,0043
85º	Bauru	2113	8762	,2412	,0174	,0042
86º	Nova Petrópolis	2098	6042	,3472	,0120	,0042
87º	Olinda	1884	7191	,2620	,0143	,0037
88º	Olímpia	1824	5627	,3242	,0112	,0036
89º	Penha	1762	6050	,2912	,0120	,0035
90º	Canoa Quebrada	1666	4965	,3355	,0099	,0033
91º	Angra dos Reis	1647	5035	,3271	,0100	,0033
92º	São Lourenço	1606	5675	,2830	,0113	,0032
93º	Ilhéus	1451	5180	,2801	,0103	,0029
94º	Praia do Forte	1414	5853	,2416	,0116	,0028
95º	Ilha Grande	1404	3324	,4224	,0066	,0028
96º	Palmas	1194	6107	,1955	,0122	,0024
97º	Aquiraz	1090	2108	,5171	,0042	,0022
98º	Barreirinhas	603	3042	,1982	,0061	,0012
99º	Rio Quente	556	2690	,2067	,0054	,0011
100º	Gramado	401	1025	,3912	,0020	,0008
$\Sigma =$		1.047.881	3.027.796			
ME =		5.104,00	15.298,50			
MD =		10.478,81	30.277,96			
DP =		19.132,98	60.399,05			
Mín =		401	1.025			
Máx =		144.628	502.575			

Tabela 12: Mostra, considerado o aspecto ATRAÇÕES (ATR) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários (*nTLikes_ATR*), o número total de comentários (*nTC_ATR*), a proporção de *likes* por comentário ($nTLikes_ATR / nTC_ATR$), a proporção de comentários em relação ao número máximo de comentários observado ($nTC_ATR / nC_{máx}$) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes_ATR*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes_ATR*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes_ATR*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	ATRAÇÕES (ATR)				
		<i>nTLikes</i>	<i>nTC</i>	$nTLikes/nTC$	$nTC/nC_{máx}$	<i>wpLikes</i>
1º	Recife	96455	49258	1,9582	,2132	,4176
2º	Rio de Janeiro	33470	230991	,1449	1,0000	,1449
3º	Curitiba	29959	98815	,3032	,4278	,1297
4º	Gramado	25946	132113	,1964	,5719	,1123
5º	São Paulo	25124	203422	,1235	,8806	,1088
6º	Maceió	21222	61111	,3473	,2646	,0919
7º	Fortaleza	16844	77107	,2184	,3338	,0729
8º	Salvador	16025	67703	,2367	,2931	,0694
9º	Natal	14864	65053	,2285	,2816	,0643
10º	Brasília	14203	78960	,1799	,3418	,0615
11º	Belo Horizonte	13248	64757	,2046	,2803	,0574
12º	Foz do Iguaçu	12105	83710	,1446	,3624	,0524
13º	Olinda	9979	8577	1,1635	,0371	,0432
14º	Canela	9535	44411	,2147	,1923	,0413
15º	Florianópolis	9373	55433	,1691	,2400	,0406
16º	Porto Seguro	9353	31608	,2959	,1368	,0405
17º	Campos do Jordão	9021	47029	,1918	,2036	,0391
18º	Bonito	8689	26111	,3328	,1130	,0376
19º	Ubatuba	8503	21874	,3887	,0947	,0368
20º	João Pessoa	8421	28729	,2931	,1244	,0365
21º	Aracaju	8306	27535	,3017	,1192	,0360
22º	Praia do Forte	8159	9376	,8702	,0406	,0353
23º	Porto de Galinhas	7866	35848	,2194	,1552	,0341
24º	Poços de Caldas	7452	10462	,7123	,0453	,0323
25º	Bento Gonçalves	6983	19956	,3499	,0864	,0302
26º	Manaus	6982	31845	,2192	,1379	,0302
27º	Paraty	6514	22351	,2914	,0968	,0282
28º	Cabo Frio	6239	9738	,6407	,0422	,0270
29º	Maragogi	5910	15011	,3937	,0650	,0256
30º	Porto Alegre	5888	45118	,1305	,1953	,0255
31º	Penha	5709	13007	,4389	,0563	,0247
32º	Guarujá	5274	10460	,5042	,0453	,0228
33º	Ilhabela	5007	16642	,3009	,0720	,0217
34º	Belém	4992	28692	,1740	,1242	,0216
35º	Armação dos Búzios	4991	36303	,1375	,1572	,0216
36º	Ouro Preto	4946	19681	,2513	,0852	,0214

Tabela 12: (Continuação) Mostra, considerado o aspecto ATRAÇÕES (ATR) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários (*nTLikes_ATR*), o número total de comentários (*nTC_ATR*), a proporção de *likes* por comentário ($nTLikes_ATR / nTC_ATR$), a proporção de comentários em relação ao número máximo de comentários observado ($nTC_ATR / nC_{máx}$) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes_ATR*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes_ATR*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nTLikes_ATR*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	ATRAÇÕES (ATR)				
		<i>nTLikes</i>	<i>nTC</i>	$nTLikes/nTC$	$nTC/nC_{máx}$	<i>wpLikes</i>
37º	Morro de São Paulo	4787	10099	,4740	,0437	,0207
38º	Caldas Novas	4597	13104	,3508	,0567	,0199
39º	Balneário Camboriú	4550	22828	,1993	,0988	,0197
40º	Pirenópolis	4394	9160	,4797	,0397	,0190
41º	Alto Paraíso de Goiás	4160	10162	,4094	,0440	,0180
42º	Arraial d'Ajuda	4111	18198	,2259	,0788	,0178
43º	Tamandaré	4063	9420	,4313	,0408	,0176
44º	Petrópolis	3990	22227	,1795	,0962	,0173
45º	São Sebastião	3831	9693	,3952	,0420	,0166
46º	Blumenau	3824	10823	,3533	,0469	,0166
47º	Fernando de Noronha	3786	26054	,1453	,1128	,0164
48º	Vitória	3783	17589	,2151	,0761	,0164
49º	Joinville	3781	11182	,3381	,0484	,0164
50º	Cuiabá	3683	11546	,3190	,0500	,0159
51º	Goiânia	3628	15819	,2293	,0685	,0157
52º	Santos	3614	19693	,1835	,0853	,0156
53º	Olímpia	3402	7695	,4421	,0333	,0147
54º	São Luís	3391	14944	,2269	,0647	,0147
55º	Trancoso	3258	10355	,3146	,0448	,0141
56º	Arraial do Cabo	3257	13303	,2448	,0576	,0141
57º	Niterói	3208	17509	,1832	,0758	,0139
58º	Rio Quente	2994	6852	,4370	,0297	,0130
59º	Jericoacoara	2934	21684	,1353	,0939	,0127
60º	Guarapari	2924	7771	,3763	,0336	,0127
61º	Itacaré	2836	12099	,2344	,0524	,0123
62º	Tiradentes	2828	11004	,2570	,0476	,0122
63º	Bombinhas	2617	10567	,2477	,0457	,0113
64º	São Roque	2593	4671	,5551	,0202	,0112
65º	Praia da Pipa	2534	17136	,1479	,0742	,0110
66º	Vila Velha	2438	10901	,2236	,0472	,0106
67º	Barreirinhas	2391	7772	,3076	,0336	,0104
68º	Campinas	2069	10971	,1886	,0475	,0090
69º	Ilha Grande	1975	11347	,1741	,0491	,0086
70º	Angra dos Reis	1925	5910	,3257	,0256	,0083
71º	São Lourenço	1820	4792	,3798	,0207	,0079
72º	Ilhéus	1811	5808	,3118	,0251	,0078

Tabela 12: (Continuação) Mostra, considerado o aspecto ATRAÇÕES (ATR) para cada um dos 100 destinos turísticos analisados, o número total de *likes* atribuídos aos comentários (*nLikes_ATR*), o número total de comentários (*nTC_ATR*), a proporção de *likes* por comentário (*nLikes_ATR / nTC_ATR*), a proporção de comentários em relação ao número máximo de comentários observado (*nTC_ATR / nC_{máx}*) e a proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes_ATR*). O *ranking* dos destinos turísticos foi obtido com a disposição decrescente dos valores *wpLikes_ATR*. Os valores medianos (ME), médios (MD; com desvios-padrão, DP) e valores mínimo e máximo para as variáveis (*nLikes_ATR*) e *nTC*, são mostrados nas linhas finais da tabela.

Rank	Destino turístico	ATRAÇÕES (ATR)				
		<i>nLikes</i>	<i>nTC</i>	<i>nLikes/nTC</i>	<i>nTC/nC_{máx}</i>	<i>wpLikes</i>
73º	Teresópolis	1700	4147	,4099	,0180	,0074
74º	Monte Verde	1658	6135	,2703	,0266	,0072
75º	Canoa Quebrada	1484	7836	,1894	,0339	,0064
76º	Ribeirão Preto	1375	7771	,1769	,0336	,0060
77º	Lençóis	1312	5126	,2560	,0222	,0057
78º	Palmas	1219	4507	,2705	,0195	,0053
79º	Nova Petrópolis	1206	6580	,1833	,0285	,0052
80º	Penedo	1097	4183	,2623	,0181	,0047
81º	Campo Grande	1090	6118	,1782	,0265	,0047
82º	Caxias do Sul	981	3615	,2714	,0156	,0042
83º	Londrina	962	5608	,1715	,0243	,0042
84º	São Bernardo do Campo	918	2677	,3429	,0116	,0040
85º	Aquiraz	866	2513	,3446	,0109	,0037
86º	Teresina	820	4381	,1872	,0190	,0035
87º	Juiz de Fora	740	3684	,2009	,0159	,0032
88º	Uberlândia	657	3832	,1715	,0166	,0028
89º	Maringá	630	3548	,1776	,0154	,0027
90º	Sorocaba	573	2965	,1933	,0128	,0025
91º	São José dos Campos	547	3421	,1599	,0148	,0024
92º	Piracicaba	538	2773	,1940	,0120	,0023
93º	Atibaia	471	1263	,3729	,0055	,0020
94º	Bauru	413	1923	,2148	,0083	,0018
95º	Guarulhos	360	2594	,1388	,0112	,0016
96º	Taubaté	346	1503	,2302	,0065	,0015
97º	Santo André	323	2460	,1313	,0106	,0014
98º	Jundiaí	288	2071	,1391	,0090	,0012
99º	São José do Rio Preto	228	1760	,1295	,0076	,0010
100º	Barueri	53	574	,0923	,0025	,0002
Σ =		634.199	2.385.053			
ME =		3.655,50	10.936,00			
MD =		6.341,99	23.850,53			
DP =		11.055,47	36.305,18			
Mín =		53	574			
Máx =		96.455	230.991			

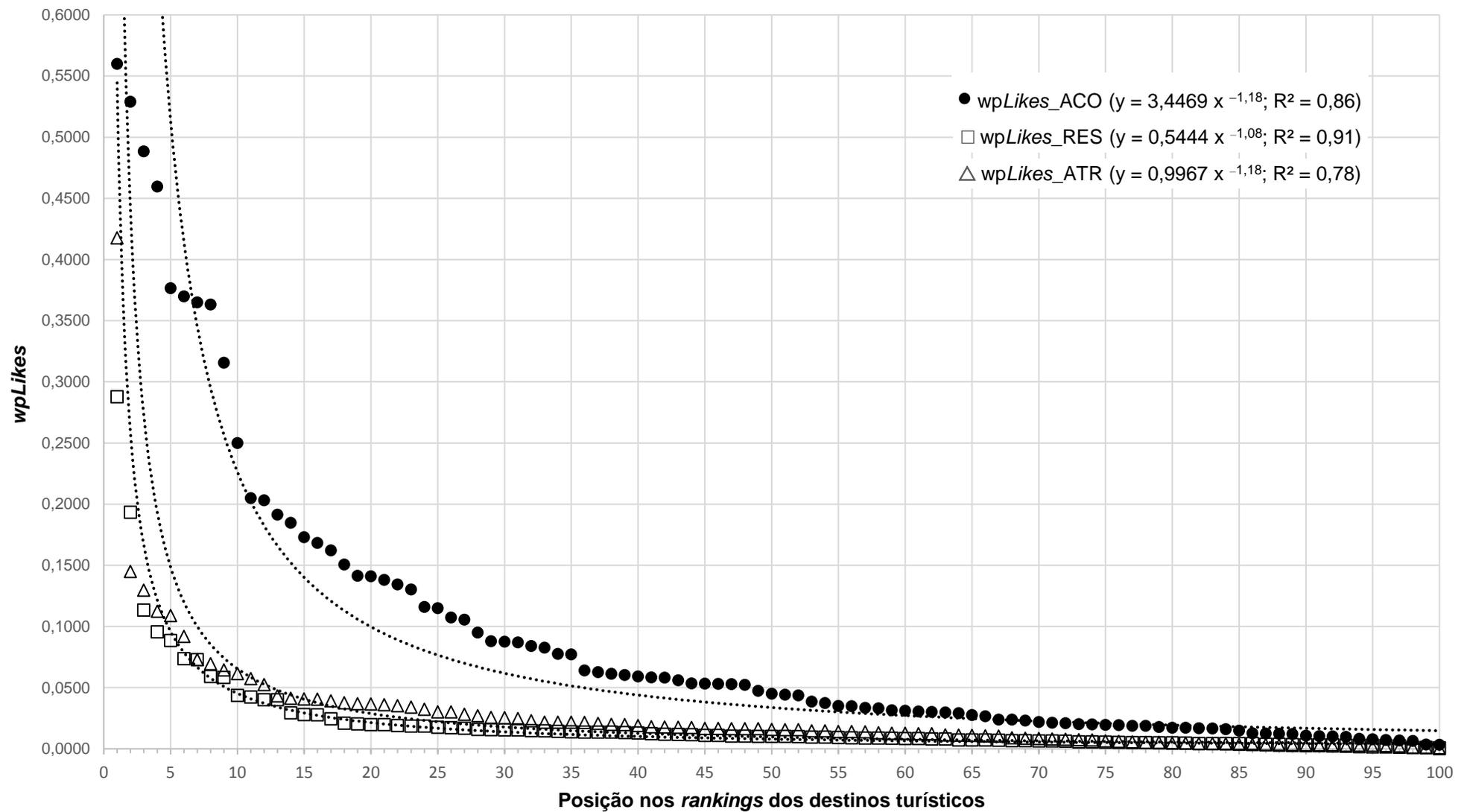


Figura 7: Mostra os valores da proporção relativa ponderada de *likes* (*wpLikes*) para comentários sobre os componentes (ACO, RES e ATR) de cada um dos 100 destinos turísticos analisados, conforme posição assumida pelos destinos nos *rankings* obtidos com a disposição decrescente dos valores (Tabela 10, Tabela 11 e Tabela 12). Os expoentes das equações de ajuste (funções potência) indicam a taxa média de decréscimo dos valores da variável a cada posição sucessiva.

Tabela 13: Mostra os coeficientes de correlação *rho* de Spearman entre os valores das proporções ponderadas de *likes* (*wpLikes*) para comentários sobre acomodações (ACO), restaurantes (RES) e atrativos (ATR) emitidos para cada um dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%; bicaudado).

	<i>wpLikes</i> _ACO	<i>wpLikes</i> _RES	<i>wpLikes</i> _ATR
<i>wpLikes</i> _ACO	---	0,429**	0,725**
<i>wpLikes</i> _RES		---	0,448**
<i>wpLikes</i> _ATR			---

Tabela 14: Mostra os coeficientes de correlação *rho* de Spearman entre os valores das proporções relativas ponderadas de *likes* (*wpLikes*) e as proporções relativas de comentários sobre acomodações (ACO), restaurantes (RES) e atrativos (ATR) emitidos para cada um dos 100 destinos turísticos analisados. (n = 100; ** Nível de significância = 0,01%; bicaudado).

	<i>wpLikes</i> _ACO	<i>wpLikes</i> _RES	<i>wpLikes</i> _ATR
pC_ACO	0,843**	---	---
pC_RES		0,934**	---
pC_ATR			0,897**

Na análise dos *likes* por componentes (Tabela 10, Tabela 11 e Tabela 12), o componente ACO teve o maior valor de proporção de *likes* ponderada pelo número proporcional de comentários (*wpLikes*), com a cidade de Natal na primeira posição do *ranking* (*wpLikes* = 0,5598). Ao verificar a disposição dos destinos na Figura 7, vale ressaltar que até a 4ª posição do *ranking*, ACO é superior à primeira posição dos outros componentes, tendo apenas uma incidência de ATR (*wpLikes* = 0,4176) e RES (*wpLikes* = 0,2878) com *wpLikes* > 0,2, enquanto que ACO tem 12 destinos com proporção de *likes* superior a esse valor. O componente ATR, tem Recife na primeira posição; e São Paulo, que esteve na primeira posição na maioria das análises, alcança esse *status*, nesse item, somente no componente RES. Há uma similaridade entre os 20 destinos nas primeiras posições entre os três componentes, excetuando-se os seguintes: Maragogi ([10º, ACO], [71º, RES], [29º, ATR]), Santos ([6º, RES], [59º, ACO], [52º, ATR]), Campinas ([13º, RES], [51º, ACO], [68º, ATR]), Olinda ([13º, ATR], [91º, ACO], [87º, RES]) e Canela ([14º, ATR], [52º, ACO], [26º, RES]).

Na Tabela 13 são apresentados os coeficientes de correlação *rho* de Spearman entre os valores das proporções ponderadas de *likes* (*wpLikes*) para comentários sobre acomodações (ACO), restaurantes (RES) e atrativos (ATR). As associações

que se estabelecem são moderadas (para baixo), com mais ênfase (um aumento de associação) entre ACO e ATR. No âmbito da análise baseada no BPM, essas correlações podem ser interpretadas como indicativas de maior ou menor grau de independência das emissões de *likes* em relação aos componentes do produto turístico. A independência (não associação) é particularmente verdadeira para as relações entre ACO e RES, e ACO e ATR. Os dados desses dois pares de componentes podem ser interpretados como indicativos de controle diferencial sobre a emissão de *likes* pelos estímulos discriminativos verbais que, nas interações dos avaliadores emissores com as dimensões do site e entre eles, configuram as condições de hospedagem e alimentação nas localidades por um lado, e as condições de acomodação e os atrativos da localidade por outro.

Nos coeficientes de correlação, apresentados na Tabela 14, entre os valores das proporções relativas ponderadas de *likes* (*wpLikes*) e as proporções relativas de comentários sobre os componentes, verificou-se o quanto essas duas variáveis estão associadas. Como pode ser visto, as correlações são positivas, bem elevadas e muito significativas estatisticamente. Esses números significam que quanto mais comentários (sobre ACO, RES e ATR) um destino recebeu, mais *likes* a esses comentários ocorreram.

Na Tabela 15, a seguir, é realizada uma comparação entre os números totais de *likes* emitidos por *TripColaboradores* em duas classes de níveis.

Tabela 15: Mostra as estatísticas *U* de Mann-Whitney para comparações entre os números totais médios de *likes* emitidos por *TripColaboradores* classificados nos níveis mínimo (N1 e N2) e máximo (N5 e N6) de experiência como avaliadores, para os componentes acomodações (ACO), restaurantes (RES) e atrações (ATR), e os valores médios comparados (em número de *likes* e posts), o valor *p.* observado e os valores de tamanho do efeito (*r*) para cada comparação (*n* = 100; ** Nível de confiança = 0,01%; bicaudado).

Comparações	MD Likes	MD Postos	<i>U</i>	<i>p.</i>	<i>r</i>
N _T L _i k _e s_ACO_TripCol_N1+N2 vs. N _T L _i k _e s_ACO_TripCol_N5+N6	1.719,80	83,63	3.313,00	0,00	- 0,29
	3.069,29	117,37			
N _T L _i k _e s_RES_TripCol_N1+N2 vs. N _T L _i k _e s_RES_TripCol_N5+N6	1.215,59	68,76	1.825,50	0,00	- 0,55
	4.765,28	132,25			
N _T L _i k _e s_ATR_TripCol_N1+N2 vs. N _T L _i k _e s_ATR_TripCol_N5+N6	582,77	66,73	1.623,00	0,00	- 0,58
	3.468,88	134,27			

Nas estatísticas *U* de Mann-Whitney, apresentadas na Tabela 15, para comparações entre os números totais médios de *likes* emitidos por *TripColaboradores* classificados nos níveis mínimo (N1 e N2), que correspondem aos usuários que tem entre 300 e 500 pontos no sistema gamificado do *TripAdvisor*®, e máximo (N5 e N6) de experiência como avaliadores (usuários com 5000 pontos ou mais), para os componentes acomodações (ACO), restaurantes (RES) e atrações (ATR). Com as comparações realizadas, verificou-se que os *TripColaboradores* nos dois níveis emitem *likes* em quantidades bem diferenciadas, especialmente os usuários dos componentes Restaurantes e Atrações, cujas diferenças foram mais acentuadas ($MDLikes_{N1+N2} = 1.215,59$ e $MDLikes_{N5+N6} = 4.765,28$, $r = -0,55$; $MDLikes_{N1+N2} = 582,77$ e $MDLikes_{N5+N6} = 3.468,88$, $r = -0,55$, respectivamente). Isso evidenciou que consumidores com diferentes *backgrounds* (experiência) como avaliadores recebem *likes* em quantidades significativamente diferentes, sendo as diferenças ainda mais acentuadas para os componentes RES e ATR, quando a experiência tende a um grande efeito (valores *r* superiores a 0,50). Se for considerada a história passada de aprendizagem do consumir pelo turista, especialmente no que diz respeito a avaliação de produtos e o efeito disso na comunidade em que estão inseridos (*TripAdvisor*®), então segue-se que os *TripColaboradores* com mais experiência recebem mais *likes*, pois tem mais visibilidade no sistema (geralmente os comentários que aparecem nas primeiras posições das páginas são os mais curtidos),

seus perfis têm informações que reforçam o comportamento dos demais em considerar as suas opiniões sobre os destinos turísticos, dando mais autoridade aos seus comentários no ambiente. Por exemplo, o *e-bay*²⁷, um serviço de comércio eletrônico para compra e venda de produtos novos e usados, tem um estruturado sistema de *feedback* do vendedor e do comprador²⁸, assim, quanto mais *feedback* positivo recebe um perfil, mais autoridade no processo de compra e venda ele passa a ter no ambiente. Ainda que sistema do *TripAdvisor*® não haja transações comerciais, logo a investigação sobre o perfil talvez não seja tão minuciosa, há relevância sobre sua história naquele ambiente e nas variáveis que compõem seu comportamento de consumo.

Nas tabelas a seguir (Tabela 16 e Tabela 17) serão apresentados os *rankings* dos destinos turísticos a partir de duas análises, o número total de avaliações positivas e negativas realizadas nos comentários por meio da *SentimentALL* e as avaliações realizadas pelos usuários a partir da escala *likert* do *TripAdvisor*®.

²⁷ <https://www.ebay.com/>

²⁸ <https://pages.ebay.com/br/pt-br/help/feedback/howitworks.html>

Tabela 16: Mostra *rankings* dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações positivas (nT_AvP; à esquerda) e do somatório dos números totais de avaliações “Muito bom” (MB) e “Excelente” (EX), nT_(MB+EX) à direita, conforme escala avaliativa do *TripAdvisor*®.

Rank	Destino turístico	nT_AvP	Rank	Destino turístico	nT_(MB+EX)
1º	São Paulo	2442716	1º	São Paulo	674129
2º	Rio de Janeiro	1725477	2º	Rio de Janeiro	507778
3º	Brasília	720238	3º	Brasília	202595
4º	Belo Horizonte	665670	4º	Fortaleza	187065
5º	Fortaleza	643987	5º	Belo Horizonte	176457
6º	Natal	595488	6º	Natal	162750
7º	Salvador	527356	7º	Curitiba	153954
8º	Curitiba	521103	8º	Salvador	144817
9º	Florianópolis	496752	9º	Foz do Iguaçu	144292
10º	Foz do Iguaçu	486568	10º	Gramado	140709
11º	Maceió	468788	11º	Florianópolis	137406
12º	Gramado	442752	12º	Maceió	127685
13º	Campos do Jordão	320058	13º	Campos do Jordão	93658
14º	João Pessoa	279424	14º	João Pessoa	76054
15º	Recife	259715	15º	Porto de Galinhas	73380
16º	Porto de Galinhas	253004	16º	Armação dos Búzios	70107
17º	Porto Alegre	247535	17º	Manaus	66833
18º	Campinas	238657	18º	Campinas	64866
19º	Armação dos Búzios	235448	19º	Canela	63904
20º	Manaus	223339	20º	Belém	59679
21º	Aracaju	208124	21º	Aracaju	58096
22º	Belém	206701	22º	Recife	57715
23º	Canela	204667	23º	Porto Alegre	55735
24º	Goiânia	198684	24º	Goiânia	55377
25º	Balneário Camboriú	194741	25º	Porto Seguro	55098
26º	Porto Seguro	193349	26º	Balneário Camboriú	53562
27º	Vitória	171828	27º	Jericoacoara	49652
28º	Jericoacoara	150503	28º	Vitória	48510
29º	Bento Gonçalves	148700	29º	Santos	42976
30º	Ribeirão Preto	147635	30º	Fernando de Noronha	42644
31º	Bonito	143941	31º	Bonito	42281
32º	Santos	142668	32º	Paraty	41288
33º	Paraty	139160	33º	Niterói	41080
34º	Niterói	135917	34º	Bento Gonçalves	40981
35º	São Luís	135585	35º	Ribeirão Preto	40941
36º	Petrópolis	135063	36º	Petrópolis	40813
37º	Fernando de Noronha	123920	37º	Ilhabela	36003
38º	Ilhabela	120086	38º	São Luís	35905

Tabela 16: (Continuação) Mostra *rankings* dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações positivas (nT_AvP; à esquerda) e do somatório dos números totais de avaliações “Muito bom” (MB) e “Excelente” (EX), nT_(MB+EX) à direita, conforme escala avaliativa do *TripAdvisor*®.

Rank	Destino turístico	nT_AvP	Rank	Destino turístico	nT_(MB+EX)
39º	Arraial d'Ajuda	111057	39º	Arraial d'Ajuda	33533
40º	Maragogi	110706	40º	Ouro Preto	33253
41º	Cuiabá	109817	41º	Ubatuba	32523
42º	Ouro Preto	109166	42º	Maragogi	30692
43º	Joinville	106733	43º	Praia da Pipa	30597
44º	Ubatuba	103133	44º	Tiradentes	27948
45º	Tiradentes	101673	45º	Monte Verde	27946
46º	Londrina	101288	46º	Campo Grande	27600
47º	Campo Grande	100528	47º	Londrina	27508
48º	Caldas Novas	98402	48º	São José dos Campos	26236
49º	Praia da Pipa	97663	49º	Caldas Novas	25125
50º	São José dos Campos	93822	50º	Juiz de Fora	22432
51º	Monte Verde	93355	51º	Vila Velha	22028
52º	Poços de Caldas	87113	52º	Itacaré	21821
53º	Blumenau	84799	53º	Cuiabá	21793
54º	Juiz de Fora	80698	54º	Joinville	21024
55º	Guarujá	79866	55º	Guarulhos	20310
56º	Morro de São Paulo	79352	56º	Arraial do Cabo	19089
57º	Guarulhos	76955	57º	Santo André	19088
58º	Praia do Forte	76713	58º	Pirenópolis	18687
59º	Cabo Frio	70906	59º	Bombinhas	18288
60º	Teresópolis	70102	60º	Teresópolis	18235
61º	Vila Velha	70032	61º	São Sebastião	17999
62º	Itacaré	69965	62º	Olímpia	17968
63º	Pirenópolis	68622	63º	Penha	17899
64º	Alto Paraíso de Goiás	66635	64º	Uberlândia	17583
65º	Santo André	65020	65º	Jundiaí	17252
66º	Caxias do Sul	62850	66º	Sorocaba	16794
67º	Olímpia	62785	67º	São José do Rio Preto	16739
68º	Uberlândia	62784	68º	Maringá	16670
69º	São Sebastião	60716	69º	Trancoso	16629
70º	Teresina	60609	70º	Caxias do Sul	16584
71º	São José do Rio Preto	60144	71º	Teresina	16389
72º	Sorocaba	59964	72º	Penedo	15529
73º	Bombinhas	59622	73º	Blumenau	15366
74º	Jundiaí	59480	74º	São Bernardo do Campo	15017
75º	Maringá	59362	75º	Poços de Caldas	14607
76º	Penha	56132	76º	Alto Paraíso de Goiás	14443

Tabela 16: (Continuação) Mostra *rankings* dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações positivas (nT_AvP; à esquerda) e do somatório dos números totais de avaliações “Muito bom” (MB) e “Excelente” (EX), nT_(MB+EX) à direita, conforme escala avaliativa do *TripAdvisor*®.

Rank	Destino turístico	nT_AvP	Rank	Destino turístico	nT_(MB+EX)
77º	Arraial do Cabo	55628	77º	Canoa Quebrada	14341
78º	Trancoso	54842	78º	Guarujá	14227
79º	Ilhéus	54139	79º	Olinda	14158
80º	Penedo	53178	80º	Piracicaba	14026
81º	Atibaia	50946	81º	Tamandaré	13806
82º	Canoa Quebrada	50776	82º	Ilha Grande	13733
83º	Piracicaba	49997	83º	Morro de São Paulo	13697
84º	São Bernardo do Campo	49867	84º	Ilhéus	13407
85º	Angra dos Reis	49504	85º	Angra dos Reis	12879
86º	Rio Quente	48258	86º	Guarapari	12730
87º	Olinda	47181	87º	Atibaia	12557
88º	Tamandaré	46612	88º	Cabo Frio	12347
89º	Nova Petrópolis	44383	89º	Nova Petrópolis	12085
90º	São Lourenço	43504	90º	Praia do Forte	11777
91º	Guarapari	42120	91º	São Lourenço	11771
92º	São Roque	41373	92º	Lençóis	11584
93º	Ilha Grande	40973	93º	Rio Quente	11345
94º	Lençóis	39371	94º	São Roque	11272
95º	Barueri	38705	95º	Barueri	11122
96º	Bauru	37277	96º	Barreirinhas	10962
97º	Taubaté	36696	97º	Bauru	10348
98º	Aquiraz	35584	98º	Palmas	10113
99º	Palmas	35152	99º	Taubaté	9949
100º	Barreirinhas	34256	100º	Aquiraz	8204

Tabela 17: Mostra *rankings* dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações negativas (nT_AvN; à esquerda) e do somatório dos números totais de avaliações “Horrível” (HO) e “Ruim” (RU), nT_(HO+RU) à direita, conforme escala avaliativa do *TripAdvisor*®.

Rank	Destino turístico	nT_AvN	Rank	Destino turístico	nT_(HO+RU)
1º	São Paulo	274594	1º	São Paulo	40546
2º	Rio de Janeiro	205350	2º	Rio de Janeiro	29257
3º	Brasília	80726	3º	Brasília	11728
4º	Fortaleza	71552	4º	Fortaleza	10923
5º	Belo Horizonte	67098	5º	Belo Horizonte	9448
6º	Salvador	61642	6º	Salvador	9314
7º	Natal	58986	7º	Campos do Jordão	8100
8º	Florianópolis	55942	8º	Natal	7913
9º	Maceió	51799	9º	Florianópolis	7538
10º	Curitiba	46394	10º	Maceió	7246
11º	Foz do Iguaçu	46104	11º	Gramado	6603
12º	Gramado	40356	12º	Curitiba	6553
13º	Campos do Jordão	37852	13º	Foz do Iguaçu	6475
14º	Porto de Galinhas	30168	14º	Campinas	5147
15º	João Pessoa	30044	15º	João Pessoa	4248
16º	Armação dos Búzios	29637	16º	Armação dos Búzios	3925
17º	Campinas	29499	17º	Porto Seguro	3907
18º	Recife	24942	18º	Porto de Galinhas	3760
19º	Porto Alegre	24132	19º	Aracaju	3486
20º	Aracaju	23565	20º	Goiânia	3445
21º	Porto Seguro	23026	21º	Manaus	3034
22º	Goiânia	22627	22º	Balneário Camboriú	3000
23º	Manaus	22012	23º	Porto Alegre	2850
24º	Belém	20873	24º	Canela	2810
25º	Balneário Camboriú	19735	25º	Recife	2680
26º	Canela	19150	26º	Ribeirão Preto	2605
27º	Paraty	17491	27º	Belém	2496
28º	Vitória	17462	28º	Vitória	2492
29º	Jericoacoara	16762	29º	Caldas Novas	2489
30º	Ribeirão Preto	16459	30º	Paraty	2346
31º	Niterói	15509	31º	São Luís	2327
32º	Petrópolis	15467	32º	Santos	2197
33º	São Luís	14955	33º	Niterói	2164
34º	Santos	14912	34º	Guarulhos	2155
35º	Bonito	14758	35º	Petrópolis	1877
36º	Caldas Novas	13890	36º	Monte Verde	1806
37º	Ilhabela	13711	37º	Jericoacoara	1790
38º	Fernando de Noronha	13578	38º	São José dos Campos	1760

Tabela 17: (Continuação) Mostra *rankings* dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações negativas (nT_AvN; à esquerda) e do somatório dos números totais de avaliações “Horível” (HO) e “Ruim” (RU), nT_(HO+RU) à direita, conforme escala avaliativa do *TripAdvisor*®.

Rank	Destino turístico	nT_AvN	Rank	Destino turístico	nT_(HO+RU)
39º	Maragogi	13554	39º	Ilhabela	1641
40º	Bento Gonçalves	12988	40º	Olímpia	1555
41º	Arraial d'Ajuda	12061	41º	Bento Gonçalves	1554
42º	Monte Verde	11993	42º	Arraial d'Ajuda	1553
43º	Ubatuba	11659	43º	Maragogi	1539
44º	Ouro Preto	11593	44º	Tiradentes	1521
45º	Tiradentes	11400	45º	Ubatuba	1432
46º	Cuiabá	11008	46º	Bonito	1405
47º	Praia da Pipa	10430	47º	Fernando de Noronha	1398
48º	Joinville	10392	48º	Londrina	1389
49º	São José dos Campos	10250	49º	Santo André	1355
50º	Guarulhos	9886	50º	Cabo Frio	1299
51º	Cabo Frio	9490	51º	Campo Grande	1238
52º	Campo Grande	9390	52º	Guarujá	1234
53º	Guarujá	9363	53º	Jundiaí	1191
54º	Morro de São Paulo	9323	54º	Joinville	1178
55º	Londrina	9290	55º	Penedo	1152
56º	Poços de Caldas	9256	56º	Teresópolis	1135
57º	Trancoso	9116	57º	Rio Quente	1100
58º	Olímpia	8878	58º	São Bernardo do Campo	1087
59º	São José do Rio Preto	8609	59º	Morro de São Paulo	1070
60º	Alto Paraíso de Goiás	8570	60º	Praia da Pipa	1041
61º	Pirenópolis	8375	61º	Angra dos Reis	1041
62º	Itacaré	8214	62º	Juiz de Fora	1022
63º	Blumenau	8172	63º	Ouro Preto	1014
64º	Juiz de Fora	8152	64º	Poços de Caldas	1007
65º	Praia do Forte	8145	65º	Cuiabá	1003
66º	Rio Quente	7928	66º	Sorocaba	998
67º	Arraial do Cabo	7680	67º	Maringá	976
68º	Teresópolis	7651	68º	Ilhéus	969
69º	Penha	7394	69º	Uberlândia	895
70º	Penedo	7233	70º	São José do Rio Preto	869
71º	Vila Velha	7133	71º	Teresina	851
72º	Santo André	7015	72º	Pirenópolis	840
73º	Angra dos Reis	6870	73º	Vila Velha	826
74º	Bombinhas	6668	74º	Blumenau	812
75º	Uberlândia	6551	75º	Piracicaba	812
76º	São Sebastião	6546	76º	Trancoso	808
77º	Ilhéus	6470	77º	Itacaré	796

Tabela 17: (Continuação) Mostra *rankings* dos destinos turísticos obtidos a partir da disposição decrescente do número total de avaliações negativas (nT_AvN; à esquerda) e do somatório dos números totais de avaliações “Horível” (HO) e “Ruim” (RU), nT_(HO+RU) à direita, conforme escala avaliativa do *TripAdvisor*®.

Rank	Destino turístico	nT_AvN	Rank	Destino turístico	nT_(HO+RU)
78º	Jundiaí	6432	78º	Barueri	781
79º	Teresina	6172	79º	Caxias do Sul	766
80º	Sorocaba	6006	80º	São Sebastião	755
81º	São Bernardo do Campo	5908	81º	Bombinhas	741
82º	Maringá	5807	82º	Atibaia	733
83º	Piracicaba	5454	83º	Arraial do Cabo	699
84º	Atibaia	5374	84º	Aquiraz	699
85º	Tamandaré	5346	85º	Tamandaré	668
86º	Guarapari	5274	86º	Penha	667
87º	Caxias do Sul	5195	87º	Guarapari	631
88º	Ilha Grande	5144	88º	Olinda	616
89º	Aquiraz	5124	89º	Taubaté	610
90º	Canoa Quebrada	4976	90º	São Roque	589
91º	Lençóis	4796	91º	Praia do Forte	558
92º	Olinda	4710	92º	Canoa Quebrada	532
93º	Barueri	4306	93º	Bauru	516
94º	São Lourenço	4154	94º	Ilha Grande	514
95º	Barreirinhas	4095	95º	São Lourenço	490
96º	Taubaté	4086	96º	Palmas	426
97º	Bauru	4017	97º	Alto Paraíso de Goiás	425
98º	São Roque	4008	98º	Nova Petrópolis	355
99º	Palmas	3618	99º	Barreirinhas	304
100º	Nova Petrópolis	3392	100º	Lençóis	289

Tabela 18: Mostra os coeficientes de correlação *rho* de Spearman entre as variáveis número total de avaliações positivas (nT_AvP), número total de avaliações negativas (nT_AvN), somatório dos números totais de avaliações “Horível” e “Ruim” (nT_[HO+RU]) e somatório dos números totais de avaliações “Muito bom” e “Excelente” (nT_[MB+EX]). (n = 100; ** Nível de significância = 0,01%).

	nT_AvP	nT_AvN	nT_(HO+RU)	nT_(MB+EX)
nT_AvP	---	,976**	,923**	,958**
nT_AvN		---	,942**	,943**
nT_(HO+RU)			---	,910**
nT_(MB+EX)				---

Além das análises dos comentários realizadas na *SentimentALL* para verificar a polaridade dos aspectos de um destino turístico, também foram extraídas das páginas de cada destino, para cada autor de comentário, a nota que este indicou no

modelo de avaliação do *TripAdvisor*®, considerando a seguinte escala (horrrível, ruim, razoável, muito bom, excelente). Para essa análise, não foi utilizada a opção “razoável”, pois para fazer uma associação entre os aspectos positivos e negativos, as demais opções eram mais adequadas. Observou-se, com isso, nos dados apresentados na Tabela 18, uma forte associação entre a quantidade de características polarizadas analisadas nos comentários (nT_{AvP} , nT_{AvN}) e da quantidade de respostas nos dois conjuntos de opções ($nT_{(MB+EX)}$, $nT_{(HO+RUM)}$).

Na próxima seção serão discutidos os resultados aqui apresentados. Essa discussão será norteada pelos conceitos do *BPM*, de forma a apresentar como os dados analisados podem contribuir para os estudos relacionados ao comportamento do consumidor, mais especificamente do setor de turismo.

DISCUSSÃO

Na seção de Resultados foram apresentados dados que descrevem o responder avaliativo polarizado relacionado aos cem (100) destinos turísticos mais avaliados no site *TripAdvisor*® no período compreendido entre o início de fevereiro e final de março de 2017. Enquanto nas pesquisas apresentadas em Miguéns, Baggio e Costa (2008) e Limberger, Anjos, Meiria e Anjos (2014), que também utilizaram dados do *TripAdvisor*®, foram considerados apenas hotéis ($n = 191$ e $n = 660$, respectivamente), os dados analisados nesta tese foram extraídos dos três componentes (ACO, RES, ATR) dos 100 destinos turísticos com maior quantidade de avaliação no *ranking* e isso resultou em uma grande quantidade de dados, mais de 6 milhões de comentários, com um total de mais de 41 milhões de avaliações (aspectos positivos/negativos). Então, o primeiro desafio na estruturação dos resultados foi a definição de um seguimento de análise, mesmo que vários outros caminhos também fossem vislumbrados com amplas possibilidades para interessantes conclusões.

A partir de uma análise mais geral, os resultados apresentados nesse estudo mostraram que os destinos turísticos variam nos níveis reforçador e punitivo, utilitários e informativos, visto que as avaliações usadas na análise são respostas verbais acerca de características desses destinos no âmbito geral ou como parte de um dos componentes do produto turístico (Acomodações, Restaurantes e Atrativos). Em uma das primeiras análises realizadas, verificou-se que apenas 15,93% do total de

comentários são do componente Acomodação, enquanto que nos componentes Restaurante e Atrações essa porcentagem é mais significativa, 47,03% e 37,04%, respectivamente. Mas, considerando que o turista-consumidor acomoda-se, geralmente, em um único hotel (salvo exceções) por viagem, e tende a ir a mais de um restaurante e conhecer e/ou se aventurar em algumas atrações, é possível interpretar que, de certa forma, enquanto a acomodação tende a ser única (ou com pouca variação) por destino turístico, há uma exposição maior em uma viagem a uma variedade de restaurantes e atrações, assim as contingências de reforço ou punição diretos ou mediados nesses contextos resultam em mais respostas verbais no site.

Os dados das avaliações verbais cadastradas no site do *TripAdvisor*® pelos turistas-consumidores em forma de comentários, e analisados pela *SentimentALL* como aspectos polarizados (positivos e negativos), podem ser interpretados como respostas verbais resultante da história passada de reforços e/ou punições mediados por outros indivíduos no próprio site. Assim, as variáveis que podem contribuir na interpretação dessas respostas avaliativas são aquelas que lançam luz sobre as circunstâncias que provavelmente envolveram os indivíduos avaliadores quando esses emitiram as respostas de avaliar. Nesse sentido, o contato com a estimulação verbal gerada pelos comentários já presentes no site pode ser relevante nesse contexto, pois há todo um espectro de contingências sociais presentes nas interações via site, que vão desde as interações estabelecidas entre os usuários, até aquelas formadas entre os usuários e o próprio site por meio do seu sistema gamificado (apresentado adiante).

As análises realizadas no Estudo 2 envolveram, sobretudo, a ordenação dos destinos a partir dos quantitativos de avaliações positivas e negativas, e nas implicações que isso tem para a classificação empírica dos mesmos como mais ou menos reforçadores para quem os consome. Primeiro, há uma incidência notavelmente superior de avaliações positivas, no gráfico da Figura 4 (seção Resultados) isso foi evidenciado, por exemplo, no fato da proporção relativa ponderada de avaliações positivas atingir uma altura inicial da curva com valor $pAvN \bullet pCom = 0,8989$, enquanto a altura inicial das avaliações negativas, o valor máximo atingido na curva é $pAvN \bullet pCom = 0,1011$. Estabelecendo uma relação, ainda que tênue, com o trabalho baseado em *BPM*, desenvolvido em Alshurideh (2014), pode ser sugerido que quanto maior a quantidade de reforços utilitários e informativos

fornecidos nas respostas verbais positivas dos turistas-consumidores acerca de um determinado destino turístico, maior a probabilidade das pessoas da comunidade do *TripAdvisor*® comentarem sobre as características positivas de suas viagens.

Segundo, sobre a classificação empírica dos destinos como mais ou menos reforçadores, um ponto observado na análise dos dados foi que os destinos turísticos, com poucas exceções, mantêm posições próximas no *ranking* por tipo de avaliação (positiva e negativa) e/ou por tipo de componente (ACO, RES, ATR). Há, inclusive, no TOP15 dos *rankings* apresentados na Tabela 4, Tabela 5 e Tabela 6, a partir de uma relação com o *ranking* geral apresentado na Tabela 3, uma constância dos mesmos destinos (ainda que em posições diferentes, exceto as duas primeiras posições que permanecem entre São Paulo e Rio de Janeiro). Para verificar a relação desse *ranking* com dados que vão além dos trabalhados nesta tese, foi utilizada, em caráter de exemplificação, uma tabela disponível na área de Dados Abertos do site²⁹ do Ministério do Turismo (MTur), relacionada a dados de turismo sistematizados por categoria, ano 2017, que mostra o desempenho da economia do setor nos municípios que constam no Mapa do Turismo Brasileiro. Ao considerar os destinos turísticos que estão categorizados por A na tabela do MTur (em uma sistematização de A a E, os destinos A são os que apresentaram os melhores desempenhos) e os destinos que estão no TOP15 do *ranking* de avaliações (Tabela 3, desta tese), verificou-se que há uma forte correspondência entre os dados das duas tabelas. Assim, o TOP15 do *ranking* de avaliações de comentários extraídos no *TripAdvisor*® estão categorizados por A no site do Ministério do Turismo. Além disso, se for considerada a variável “Demanda Doméstica” (presente na tabela de dados abertos), os mesmos TOP15 estão nessa lista com os maiores valores nas 20 primeiras posições (considerando que essa análise foi feita usando apenas os 100 destinos com mais avaliações no *TripAdvisor*®, que correspondem a 97 destinos da tabela do Ministério do Turismo).

Isso mostra um caminho para trabalhos futuros, a possibilidade de relacionar as variáveis primárias e secundárias desta tese com outros dados secundários sobre o fluxo turístico nas localidades estudadas (ou outras que poderão ser estudadas) e relativizar as quantidades observadas de comentários e avaliações (positivas e negativas) pelo tamanho dos fluxos. São Paulo, por exemplo, apareceu nos dados

²⁹ <http://dados.turismo.gov.br/categorizacao>

desta tese como o local com mais primeiras posições nos *rankings*, com mais de 800.000 comentários e milhões de avaliações, destas quase 90% são positivas. Nos limites desse estudo, apontar São Paulo como um produto turístico com magnitude reforçadora utilitária e informativa superior aos demais destinos é uma constatação válida. No entanto, se for considerada a demanda de turistas em São Paulo, até que ponto a proporção de comentários nesse destino seria representativa para estas afirmações? Por exemplo, São Paulo, na tabela do MTur, apresentou 13.478.628 de demanda doméstica (dados 2017), e nos dados apresentados em Resultados teve 814.001 comentários e 2.719.581 avaliações, o que corresponde a 20,17% de avaliações em relação à demanda. Palmas (TO), por exemplo, que foi apresentada como um cenário de consumo mais fechado, tem 439.469 de demanda doméstica (MTur), 12.438 comentários e 38.787 avaliações, o que corresponde a 8,82% considerando a mesma relação anterior, avaliações x demanda. Nesse caso, São Paulo continua sendo o destino com maior magnitude reforçadora utilitária e informativa, vista que é a maior expressão da base empírica analisada, mas se ocorresse um fluxo de demanda doméstica em Palmas na faixa de 30.000, por exemplo, o impacto das avaliações verbais dessa cidade considerando esse suposto valor de demanda seria maior que em um local com cenário de consumo bem mais aberto, com grande poder midiático e uma enorme estrutura de produtos turísticos. Mas isso é uma análise para estudos futuros que podem corroborar com essa tese, ou apontar novos caminhos de análise.

Os dados das representações da seção anterior mostraram, de modo geral, que há um conjunto de 10 a 15 destinos que induziram significativamente mais avaliações e essa tendência permaneceu na análise de quantidade dos *likes* nos comentários. Um destaque, considerando a análise por componente, foi o componente ACO, que teve o maior valor de proporção de *likes* ponderada pelo número proporcional de comentários, com a cidade de Natal na primeira posição do *ranking*, e essa foi a primeira análise em que a primeira posição não esteve entre São Paulo e Rio de Janeiro. São Paulo, nesse caso, teve uma quantidade total de *likes* bem superior a Natal (222.505 para São Paulo e 95.490 para Natal), mas em ACO foi considerada a proporção de *likes* por comentário no componente específico e, nesse contexto, Natal foi superior.

Para apresentar a interpretação do *like* utilizada neste estudo, foi preciso verificar a forma como esse comportamento é interpretado no *TripAdvisor*®³⁰. No site, parece haver uma diferença entre o *curtir*, que é similar ao *like* do *Facebook*, ou seja, qualquer pessoa pode *curtir* a página de um destino, acomodação, restaurante ou atração e o voto útil nos comentários (esse foi o dado contabilizado pela *SentimentALL* e denominado na tese por *like*). Pelas informações do site, esse voto reflete o quão a informação foi útil para outro turista, por exemplo, se o comentário discorreu sobre uma característica em um hotel que era crucial para a decisão do turista em relação a sua hospedagem. Mas a interpretação dos responsáveis pelo sistema do *TripAdvisor*® pode não ser a mesma que as pessoas têm ao acessar o botão útil em um determinado comentário. A representação do voto útil (*like*) é mostrado na Figura 8.

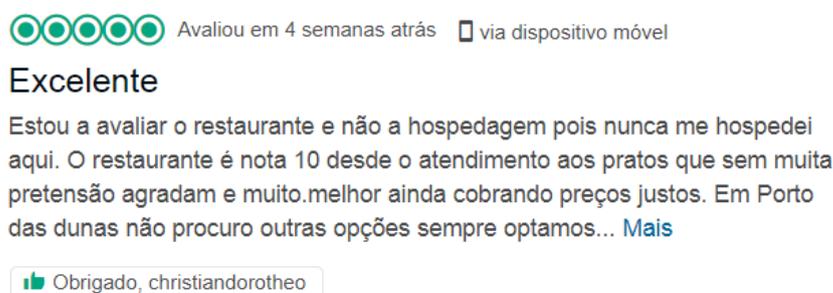


Figura 8: Representação do voto útil no *TripAdvisor*® (Fonte: Site *TripAdvisor*®)

Nesse estudo, o voto útil (denominado *like*) foi interpretado como uma espécie de *feedback*. Mais precisamente, um *feedback* interpretável como verbal a um comentário feito (seja sobre ACO, RES ou ATR, para um determinado destino), que também é uma emissão verbal. Isso configura uma interação verbal entre quem emitiu o comentário e quem emitiu o *like* (*feedback*) ao comentário. Mas, não há informação sobre o emissor do *like*, há apenas um número (total e proporcional) de *likes* na relação com um número (total ou proporcional) de comentários.

Não foi possível extrair o *like* por tipo de avaliação (positiva e negativa), pois o *like* é dado para um comentário completo (que pode ser composto por n avaliações) e o nível de análise realizado na *SentimentALL* considerou o aspecto, não o documento como um todo. Contudo, a grande maioria das avaliações para todos os destinos (90,02%, Tabela 3, Tabela 4, Tabela 5, Tabela 6), total e proporcionalmente por componente do produto turístico, foram positivas. Acrescenta-se a isso a alta

³⁰ <https://www.tripadvisor.com/hc/pt-br/articles/204291188-Diretrizes-para-votos-uteis>

correlação apresentada na Tabela 14, é possível inferir que a grande maioria dos *likes* foi emitida em relação a comentários formados por uma maioria de avaliações positivas dos destinos.

Nas análises realizadas no Estudo 2, as avaliações positivas são respostas verbais decorrentes de contatos mais frequentes (e/ou mais “significativos”) do turista-consumidor com contingências de reforço positivo do que com contingências punitivas. De certa forma, essas avaliações verbais, fontes de estímulos discriminativos verbais para aqueles indivíduos que emitiram os *likes*, sinalizaram para esses indivíduos as contingências sob as quais os emissores das avaliações positivas foram expostos, contingências essas que, ainda que apenas verbalmente representadas para os emissores dos *likes*, configuraram-se também como contingências de reforço. Mas, novamente, assim como esses *likes* não indicam necessariamente que seu uso foi interpretado por quem o emitiu com a finalidade definida no projeto de gamificação do sistema, muito provavelmente a interpretação que está sendo conduzida nesta tese também não será suficiente para abarcar todas as possibilidades.

O fato do *TripAdvisor*® ser um sistema mais fechado do que uma rede social genérica, por exemplo, como o *Facebook* ou o *Twitter*, pode estreitar os cenários. Geralmente, não é comum no *TripAdvisor*® discursos que fujam do contexto de viagens/turismo, então o *like* emitido tem mais probabilidade de ter sido útil para quem o leu considerando essa premissa. Assim prevalece a ideia de que há algumas contingências de reforços de natureza verbal para a emissão de um *like*, e são essas contingências que permitem interpretações do avaliar como comportamento de consumo. Por exemplo, foi apresentado anteriormente que o componente Acomodação teve o maior valor ponderado de *likes*, além disso, Natal foi o destino turístico que apareceu no topo do *ranking*. Duas condições foram essenciais para isso acontecer: os turistas que emitiram *likes* nos comentários do destino turístico Natal (cidade que tem a quantidade total de *likes* bem inferior a de São Paulo) fizeram isso, em sua maioria, em uma mesma categoria (ACO), esse fato e considerando que o número total de comentários de Natal está na quarta posição do *ranking* geral e em uma quantidade bem inferior às três cidade que a antecederam (todas tem mais de 170.000 comentários enquanto Natal tem menos de 100.000), fez com que o valor das proporções ponderadas de *likes* para comentários sobre ACO se tornasse o maior no âmbito geral (conforme pode ser visualizado na Figura 7, seção Resultados). Os

turistas que emitiram *likes* para São Paulo o fizeram em grande parte no componente Restaurantes, mas como o número total de comentários de SP é de maior quantidade, isso diminuiu o peso dessa emissão e, mesmo que nesse componente São Paulo tenha atingido o topo do *ranking*, ainda ficou em posição bem inferior a Natal no âmbito geral de *likes* (proporção ponderada). Contudo, estabelecer as condições de estímulos para explicar esse cenário por meio de mais variáveis e dados pode ser tópico de estudos futuros, com investigação de outras variáveis secundárias e/ou dados secundários que possam contribuir para o entendimento, por exemplo, da relevância de Natal no componente ACO, ou verificar se o fato do voto útil não contabilizar ponto nem para quem o emite, nem para quem o recebe (ponto no sistema gamificado do *TripAdvisor*®, informação disponível no site³¹) possa ser um fator preponderante para que, na época da extração dos dados para essa pesquisa, a quantidade de votos úteis (*likes*) contabilizada tenha sido menor que a metade da quantidade de comentários.

Um fator interpretativo geral condizente com os dados do Estudo 2, nos limites do que foi analisado nesse estudo até esse ponto, é o de que reforço ou punição por contato direto com acomodações, restaurantes e atrativos estão relacionados com contingências de reforço ou punição utilitários. Mas, no plano do responder verbal de avaliar esses componentes, incluindo os *feedbacks* (*likes*) a essas avaliações/comentários, contingências de reforço ou punição informativas (que são verbais em natureza) podem ter um papel de destaque. Uma avaliação em forma de comentário é um texto que pode gerar um efeito social (impacto, repercussões, etc.) a quem a emite, e um voto útil em um comentário (*like*) pode gerar visibilidade no perfil do colaborador que fez o comentário que recebeu esse voto, conforme pode ser observado na Figura 9 a seguir (foram apagadas a foto e o nome do colaborador da figura).

³¹ <https://www.tripadvisor.com/hc/pt-br/articles/204291188-Diretrizes-para-votos-uteis>

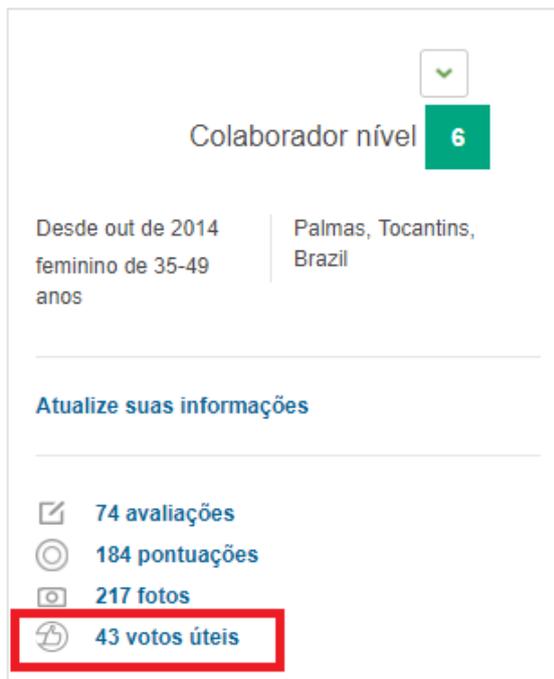


Figura 9: Perfil de colaborador com votos úteis (*likes*) (Fonte: Site TripAdvisor®)

O TripAdvisor® criou um conceito denominado *TripColaboradores* (Figura 10). Para tanto, utiliza um sistema de gamificação que pontua seus usuários de acordo com ações que este realiza no site. Assim, a publicação de comentários, de fotos, a indicação de pontuação, etc. delineiam sua experiência de consumo em viagens nos mais diversos destinos turísticos e agregam pontuação para atingir determinados níveis como *TripColaborador*. Segundo informações contidas no próprio TripAdvisor®³², “os pontos do TripColaboradores não têm valor monetário e não podem ser resgatados”.

 Avaliação	100 pontos	Nível 6	10,000 pontos
 Foto	30 pontos	Nível 5	5,000 pontos
 Vídeo	30 pontos	Nível 4	2,500 pontos
 Publicação no fórum	20 pontos	Nível 3	1,000 pontos
 Pontuação	5 pontos	Nível 2	500 pontos
		Nível 1	300 pontos

Figura 10: Composição do nível de *TripColaborador* (Fonte: Site TripAdvisor®)

³² <https://www.tripadvisor.com.br/TripCollective>

Apesar da quantidade de votos úteis não ser usada (segundo informações do site do *TripAdvisor*³³) como fonte de pontuação no sistema gamificado implantando no site, essa informação é exibida em cada comentário que recebeu o *like* e no perfil do colaborador (Figura 9). Mais recentemente, o site implementou um sistema de reconhecimento relacionado aos votos úteis³⁴. Assim, o sistema gamificado do *TripAdvisor*® concede ao colaborador um reconhecimento especial quando as suas avaliações são muito úteis para os viajantes (representada pelo *like* no comentário) e toda vez que um certo número de *likes* é atingido, há uma atualização do selo Votos Úteis, conforme apresentado na Figura 11.



Figura 11: Selos dos votos úteis do *TripAdvisor*® (Fonte: Site *TripAdvisor*®)

Nesse contexto, os selos e a informação de votos úteis no perfil do colaborador aumentam a probabilidade de ocorrência futura das respostas verbais, que decorrem do *status* alcançado na comunidade do *TripAdvisor*®. Assim, ambos funcionam como reforços informativos (“simbólicos”) de natureza social, ou seja, dependem e são mediados pelo comportamento dos outros *TripColaboradores* através de processos verbais (nesse caso, a emissão do *like* em um comentário). Pelo resultado apresentado na Tabela 15, foi evidenciado que esse sistema de pontos do site que define o nível de experiência de cada *TripColaborador* é um fator preponderante para

³³ <https://www.tripadvisor.com.br/pt-br/articles/204291188-Diretrizes-para-votos-uteis>

³⁴ <https://www.tripadvisor.com.br/TripCollectiveBadges>

explicar a diferença da quantidade de *likes* recebida pelos que estão nos níveis iniciais (que é bem inferior) daqueles usuários que já alcançaram os níveis superiores. Segundo Pohl, Oliveira-Castro, Bertoldi & Lourenzo (2006), representações da marca de um produto podem se tornar estímulos discriminativos sinalizadores do nível de reforço informativo. Nesse sentido, assim como avaliar restaurantes de São Paulo (SP), um destino turístico que está em primeiro lugar no *ranking* de avaliação geral pode provocar um impacto social diferente do que avaliar um restaurante em Bauru (SP), emitir um *like* em um comentário de um *TripColaborador* nível 6 e com selo representativo de votos úteis parece ser mais provável do que emitir essa resposta verbal em um comentário de um colaborador iniciante.

Além das avaliações dos comentários sobre os destinos turísticos oriundas do processo de AS da *SentimentALL*, que é o cerne desse estudo, foram analisadas as avaliações feitas por consumidores-turistas por meio do uso da escala *Likert* disponível no website. Na Tabela 18 pôde ser verificado que os coeficientes de correlação entre essas duas análises foram muito elevados e significativos. A partir disso, numa análise de contingências e relações de controle do responder verbal, é relevante o fato de que as respostas verbais “livres” dos comentários, tratadas de modo a que seus aspectos avaliativos positivos e negativos fossem extraídos, apresentaram a mesma tendência quantitativa das respostas a um instrumento estruturado (também verbal, porém estruturado). Sob as condições de verbalização livre (comentários), onde a rigor, não há perguntas (estímulos verbais) definidas a responder, as pessoas parecem falar do que as reforçam (avaliações positivas) ou punem (avaliações negativas) (vide proporção de avaliações positivas e avaliações negativas por comentário; Tabela 2). Sob as condições definidas por uma escala *Likert*, o responder verbal é diretamente controlado por perguntas e os modos de resposta formal e funcionalmente definidos. As palavras-estímulo designadoras dos níveis de uma escala *Likert* como a do *TripAdvisor*® (Horrível, Ruim, ..., Muito bom, Excelente) podem ser entendidas como “sintetizadoras” das centenas de palavras positivas ou negativas emissíveis na verbalização livre. O fato dos dois tipos de emissão se correlacionarem quantitativamente sugere compartilhamento ou equivalência, em algum nível, de processos de controle funcional. Assim, além dos testes realizados na *SentimentALL* para verificar seu desempenho na execução do objetivo proposto, a correlação apresentada na Tabela 18, que trata as duas formas

de análise, possibilitou a interpretação de que as avaliações analisadas pela técnica computacional, de certo modo, foram validadas pelas avaliações humanas.

DISCUSSÃO GERAL

Para atingir os objetivos propostos inicialmente, foram articulados dois estudos em duas áreas de conhecimento. Primeiro, um estudo computacional, com foco no processamento da linguagem natural (Haddi, 2015) e análise de sentimento (Liu, 2010) para comentários em língua portuguesa. Segundo, a interpretação dos dados obtidos com o processo computacional tendo como base a abordagem do *Behavioral Perspective Model (BPM)*; Foxall, 1990; 2005; 2007; 2010) na análise do comportamento do turista-consumidor, mais especificamente no processo de avaliar produtos turísticos em *websites*. O turismo foi o tema transversal que perpassou os dois estudos e, como um produto complexo, foi sistematizado a partir de seus componentes principais, serviços e produtos relacionados a acomodações, restaurantes e atrações (Candela & Figini, 2012; Acerenza, 2002).

Para Mercier, Schroeck e Shockley (2013), à medida em que o consumidor tem à disposição e adota maior número de tecnologias e acessa mais informações sobre produtos e serviços, passa a assumir maior controle sobre sua experiência de compra. Nesse contexto, as experiências de compra que utilizam uma série de canais se sobressaem: o consumidor pode pesquisar sobre o menor preço de passagem de uma viagem em um agregador disponível em um aplicativo que reúna os preços de produtos de várias empresas, depois usar o site da empresa (ou via aplicativo) para fazer a compra e, ainda, avaliar o produto em um *site* de turismo como o *TripAdvisor®*. Mas em muitas experiências, o início acontece justamente pelo site do turismo, quando as experiências de outros turistas-consumidores apresentadas como respostas verbais em forma de comentários funcionam como estímulos verbais para aqueles que buscaram informações sobre um determinado destino turístico (ou sobre um de seus componentes).

Para a extração e análise das avaliações verbais feitas pelos turistas-consumidores no site *TripAdvisor®*, iniciou-se um estudo de técnicas computacionais e voltou-se para orientações de trabalhos com foco na implementação de algoritmos (Norvig, 2016; Church & Hanks, 1990; Cruys, 2011) que pudessem ser capazes de extrair as opiniões dos comentários. Esses algoritmos, juntamente com bibliotecas de

linguagens de programação e léxicos de sentimentos foram fundamentais para o desenvolvimento da ferramenta. Para tanto, foram feitas duas versões da ferramenta com o objetivo de aperfeiçoamento/atualização das técnicas utilizadas, de forma a entregar os dados necessários ao Estudo 2, que são os quantitativos de aspectos polarizados tanto em suas totalidades gerais, quanto por componentes, na perspectiva de estudar o comportamento de avaliar do turista-consumidor.

Algumas questões técnicas da ferramenta não foram consideradas neste trabalho, por exemplo, a análise de desempenho não foi feita em relação ao tempo de processamento, pois a ferramenta não tinha caráter comercial, apesar de ter potencial para isso. Vários testes foram realizados para garantir maior confiabilidade em relação às opiniões/aspectos analisados, pois a eficiência no processo de AS era decisiva para a viabilidade do Estudo 2. Além disso, a correlação entre os aspectos polarizados e a avaliação do comentário como um todo a partir da escala *Likert* do *TripAdvisor*® (Tabela 18) corroborou com a garantia de eficiência da ferramenta na detecção dos aspectos e suas polaridades.

Ao considerar a análise do comportamento do consumidor, outros algoritmos, além daqueles utilizados no processo de AS podem ser explorados. Nessa direção, há algoritmos e técnicas de mineração de dados (Goebel & Gruenwald, 1999), como o algoritmo *Apriori* (Agrawal & Srikant, 1994), que tem como propósito identificar associações de itens em um conjunto de dados, por exemplo, contextualizando para o tema da tese, podem ser geradas associações como essa: se alguém opinar sobre comida e restaurante (positiva ou negativamente), então há grande probabilidade de opinar sobre atendimento. Com isso, a ferramenta expandiria suas possibilidades, já que, nessa tese, o aspecto em si, seu significado, não foi considerado, apenas como ele estava polarizado e de qual componente ele tinha sido extraído, além de outras informações sobre o turista-consumidor, autor do comentário. Na análise dos aspectos associados por polaridade, nessa estrutura de que dois aspectos implicariam um terceiro, com um certo grau de confiança (valor estatístico gerado pelo algoritmo), seria possível, com os dados já extraídos da *SentimentALL*, fazer uma análise dessas associações por componentes, como foram feitas a maioria das análises do Estudo 2.

Para ampliar o escopo de análise do Estudo 2, é possível adicionar aos dados extraídos da *SentimentALL*, alguns dados secundários presentes nos arquivos de dados abertos que existem em grande quantidade no setor do turismo. Nesse caso, mais variáveis podem promover uma interpretação mais minuciosa dos estímulos

ambientais que foram sinalizadores das contingências reforçadoras ou punitivas. Uma exemplificação disso foi apresentada na seção anterior, ao incluir um dado secundário, a demanda doméstica do destino, em um cenário de análise. Outros dados podem ser associados às variáveis primárias e secundárias da tese para verificar o comportamento das respostas verbais por tipo de destino turístico (metrópoles, cidades litorâneas, capitais, polo de agronegócios etc.). Além disso, com os dados da tese, pode ser realizada uma análise por região do país, para verificar se o comportamento do turista-consumidor muda a partir dessa sistematização.

Um dos módulos desenvolvidos na SentimentALL foi um *dashboard* para a visualização das 197 variáveis primárias. A Figura 12 mostra a relação de votos úteis por nível de *TripColaborador*. Na análise realizada no Estudo 2, foram feitas comparações entre a quantidade de *likes* emitidos por *TripColaboradores* nos níveis 1 e 2 (que estão no intervalo de 300 a 500 pontos) e nos níveis 5 e 6 (entre 5.000 e 10.000 pontos), concluindo que são dados mais votos úteis (*likes*) aos comentários dos usuários mais experientes (níveis 5 e 6). Mas, ao considerar todos os níveis, o componente ACO é diferenciado, especialmente pelo fato dos *TripColaboradores* iniciantes (que talvez estejam pela primeira vez no site) terem o maior índice de voto útil desse componente. Essas e outras situações diferenciadas que foram detectadas nesse componente geram elementos para estudos futuros voltados para interpretar essas especificidades.

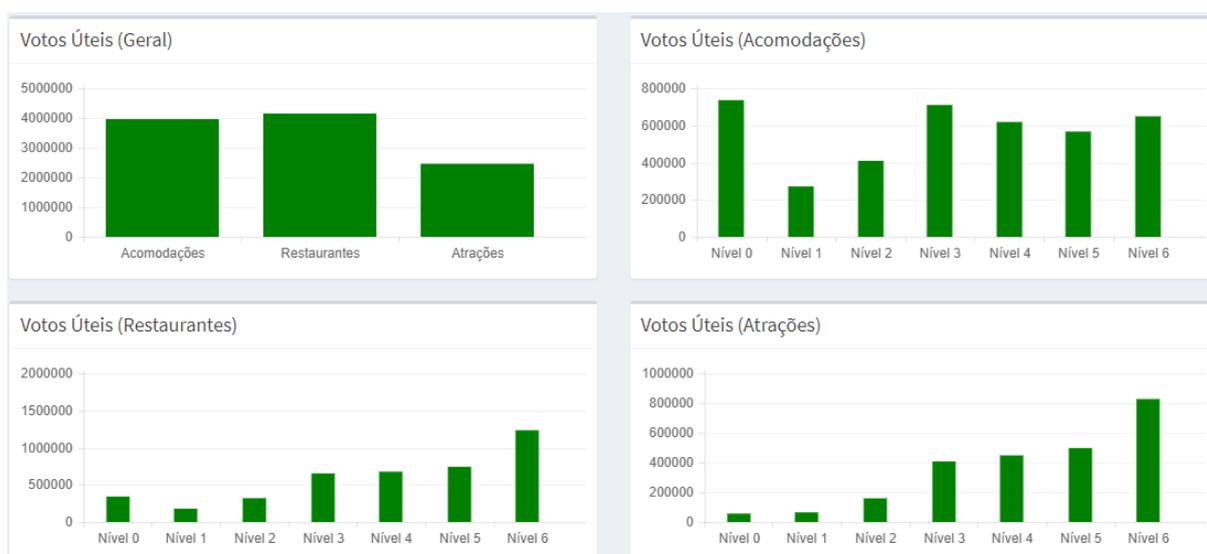


Figura 12: Número total geral de votos úteis (*likes* ) para comentários sobre Acomodações, Restaurantes e Atrações; número total de votos úteis por ACO, RES, ATR feitos por usuários de cada nível.

Entender o comportamento do consumidor tem se tornado o foco de algumas empresas. Segundo uma pesquisa realizada por Hawkins (2012), os maiores varejistas do mercado americano estão empregando o marketing personalizado. Para eles, o ponto principal para o estabelecimento dessa estratégia na relação com o cliente começou com dados obtidos nas transações *offline* e nos dados de clientes das bases de dados da empresa, o que ainda constitui um aspecto importante para o entendimento deste. Mas, acrescenta ainda, que agora as possibilidades de análises aumentaram porque são considerados, por exemplo, dados demográficos, monitoramento de vídeo da loja, dados de localização baseados em dispositivos móveis de dentro e fora da loja, *feeds* de mídia social em tempo real, anexos de dados de terceiros, tempo. Ou seja, o varejo entrou definitivamente na era do Big Data. Para Schroeck et al. (2012), isso representa uma mudança cultural na forma como os varejistas se conectam aos consumidores, acrescenta ainda que os dados importantes em um determinado contexto são especialmente promissores e diferenciadores para um produto/serviço.

Enquanto empresas como Amazon, Uber, Netflix usam os dados extraídos nas relações com os clientes para impulsionar seu negócio, seja no desenvolvimento de novos produtos e serviços, ou na antecipação do desejo do cliente e na recomendação de serviços/mercadorias, o varejo tradicional ainda está tentando se adaptar a essa nova realidade. Cada transação comercial gera um excesso de dados e ainda há pouco conhecimento técnico para lidar com isso. Mas, a inovação tecnológica intensa dos últimos 20 anos mostra que o custo para adquirir tecnologias computacionais tende a diminuir (Anderson, 2006, 2009; Pries & Dunnigan, 2015), assim como a exigência de conhecimento técnico para lidar com esses sistemas.

De certa forma, o *background* para analisar dados está ficando mais acessível, mas ainda há poucos estudos na Psicologia que tenha esses dados como foco na interpretação do comportamento. Muitos desses dados podem ser especificadores de contingências, e o consumidor, nesse ínterim, está exposto a n contingências de consumo que, às vezes, são minuciosamente programados para que o “estímulo certo” (discriminável e discriminativo) acompanhe a “consequência desejável” (reforçadora positiva ou negativa, utilitária ou informativa). No ambiente do *TripAdvisor*®, por exemplo, o sistema gamificado é uma forma de especificação e controle de contingências naquele ambiente, de maneira que os *TripColaboradores*

mantenham o ambiente ativo, dinâmico e gerando interesse em outros turistas-consumidores e, especialmente, em empresas da indústria turística. Na presente tese, foram explorados dados que mostram algumas possibilidades de análise do comportamento de consumo, especialmente, de avaliação no contexto do turismo. O trabalho desenvolvido não trouxe conclusões finais sobre esse assunto, mas possibilitou, com os esforços computacionais envolvidos, interpretar à luz do *BPM* variáveis que, interpretativamente, lançam luz aos potenciais reforços utilitários e informativos de destinos turísticos, e às contingências verbais sob as quais esses produtos são avaliados, como o fato de que 90% das avaliações foram positivas, as altas correlações entre diferentes comportamentos verbais, as contingências de reforço informativo para o comportamento de avaliar e comentar no ambiente social específico do *TripAdvisor*®. De caráter fundamentalmente exploratório, o esforço interpretativo sugeriu linhas profícuas de pesquisa, podendo ser considerado expressão da utilidade da integração entre conhecimentos computacionais e psicológicos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Acerenza, M. A. (2002). *Administração do turismo*. Bauru: EDUSC.
- Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 487-499.
- Akerkar, R. (2012). Big Data and tourism: to promote innovation and increase efficiency in the tourism sector, *TMRF Report 2011-12*, TMRF White paper, 4-12
- Alalwan, N., Zedan, H., Siewe, F. (2009). Generating OWL Ontology for Database Integration. Software Technology Research Laboratory. De Montfort University. Leicester, United Kingdom. *Third International Conference on Advances in Semantic Processing*. 2009.
- Albornoz, J. C., Plaza, L., & Gervás, P. (2010). A hybrid approach to emotional sentence polarity and intensity classification. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, 153 -161.
- Albornoz, J. C., Plaza, L., Gervás, P., & Díaz, A. (2011). A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating. In *Proceedings of the 33rd European Conference on Information Retrieval (ECIR 2011)*, 55-66.
- Alhadeff, D.A. (1982). *Microeconomics and human behavior: Towards a new synthesis of economics and psychology*, University of California Press, Berkeley, London.
- Alshurideh, M. T. (2014). A Qualitative Analysis of Customer Repeat Purchase Behaviour in the UK Mobile Phone Market, *Journal of Management Research*, 6(1), 109-125.
- Anderson, C. (2006). *A cauda longa*. Rio de Janeiro: Editora Campus.
- Anderson, C. (2009). *FREE – Grátis, O futuro dos preços*. Rio de Janeiro: Editora Campus.
- Andrade, J. V. (2000). *Turismo: Fundamentos e Dimensões*. São Paulo: Editora Ática.
- Araújo, L.G. A. (2017). *SENTIMENTALL VERSÃO 2: Desenvolvimento de Análise de Sentimentos em Python* (Trabalho de conclusão de curso). Centro Universitário Luterano de Palmas - CEULP/ULBRA, Palmas, TO, Brasil.
- Araújo, M., Gonçalves, P., & Benevenuto, F. (2013). Métodos para Análise de Sentimentos no Twitter. In *Simpósio Brasileiro de Sistemas Multimídia e Web (WEBMEDIA)*, 19, 2013, Salvador. *Proceedings...* Recuperado de: <http://homepages.dcc.ufmg.br/~fabricio/download/webmedia13.pdf>.
- Araújo, L. G. A., Marinho, D. S., & Brito, P. F. (2017). Otimizando Maltparser para Análise de Dependências na Língua Portuguesa. In *XVII Jornada de Iniciação Científica do CEULP/ULBRA*, Palmas-TO.

- Araújo, L. G. A., Brito, P. F. (2018). Módulo de Pré-Processamento da Ferramenta *SentimentALL*. In *Congresso de Computação e Tecnologias da Informação*, 20, Palmas, TO.
- Arendit, E. J. (2002). *Introdução à economia do turismo*. São Paulo: Alínea.
- Ashworth, G. J. (1989). Urban tourism: an imbalance in intention. In C. P. Cooper (Ed.), *Progress in tourism, recreation and hospitality management*. London: Belhaven.
- Baker, A. (2008). *Computer Aided Invariant Feature Selection*. (Doctoral Dissertation). University of Florida. Recuperado de <http://ufdc.ufl.edu/UFE0022870>
- Balage Filho, P. P., Pardo, T. A. S., & Aluísio, S. M. (2013). An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. In *Proceedings of The 9th Brazilian Symposium In Information and Human Language Technology - STIL*, 2013, Fortaleza, 215 - 219.
- Basant Agarwal, Namita Mittal, Pooja Bansal, & Sonal Garg. (2015). *Sentiment Analysis Using Common-Sense and Context Information*, Computational Intelligence and Neuroscience, vol. 2015, Article ID 715730, 9 pages. doi:10.1155/2015/715730
- Baum, W. M. (2006). *Compreender o behaviorismo: comportamento, cultura e evolução*. Tradução Maria Teresa Araújo Silva... [et. al.]. 2.ed., rev. e ampl. Porto Alegre: Artmed.
- Benckendorff, P. J., Sheldon, P. J., & Fesenmaier, D. R. (2014). *Tourism information technology*. Wallingford: Cab International.
- Beni, M. C. (2001). *Análise estrutural do turismo*. 5ª ed, São Paulo: SENAC.
- Beyer, M. (2011). *Gartner says solving big data challenge involves more than just managing volumes of data* [Gartner]. Recuperado de: <https://www.gartner.com/newsroom/id/1731916>
- Blackwell, R. D., Miniard, P. W. & Engel, J. F. (2000). *Comportamento do consumidor*. São Paulo: LTC.
- Blythe, A. (2008). *Consumer behavior*. London: Thompson Learning.
- Boos, R. A. S., Prestes, K. V., & Villavicencio, A. (2014). Identification of Multiword Expressions in the brWaC. In *Language Resources and Evaluation Conference (LREC)*, Reykjavik (iceland). Recuperado de: http://www.lrec-conf.org/proceedings/lrec2014/pdf/518_Paper.pdf.
- Brito, P. F., Oliveira, W. C. C., & Souza, J. G. (2015). *SentimentALL*. Fábrica de Software: CEULP/ULBRA.
- Briscoe, T. (2013). *Introduction to Linguistics for Natural Language Processing*. Computer Laboratory. University of Cambridge. Michaelmas Term 2013. Recuperado de: <https://www.cl.cam.ac.uk/teaching/1314/L100/introling.pdf>

- Brooke, J. (2009). *A semantic approach to automated text sentiment analysis*. (PhD thesis), Simon Fraser University.
- Bunn, M. D. (1993). Taxonomy of buying decision approaches. *Journal of Marketing*, 57(1), 38–56.
- Candela, G., & Figini, P. (2012). *The Economics of Tourism Destinations*, Springer, Heidelberg.
- Cardoso, M. (2012). *Determinação empírica dos níveis reforçadores utilitário e informativo de destinos turísticos*. 149 f. (Dissertação de Mestrado). Programa de Pós-Graduação Stricto Sensu em Psicologia da Pontifícia Universidade Católica de Goiás, Goiás.
- Catania, A. C. (1999). *Aprendizagem: comportamento, linguagem e cognição*. Porto Alegre: Artmed.
- Chen, B. (2011). *Topic Oriented Evolution and Sentiment Analysis*. 137 f. (PhD Tese). The Pennsylvania State University, Pennsylvania.
- Chen, M., Mao, S., & Liu, Y. (2014). *Big data: a survey, Mobile Networks and Applications*, 19 (2), 171-209.
- Chen, Z., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting domain knowledge in aspect extraction. In *EMNLP*, pages 1655–1667.
- Church, K. W., & Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, 16(1). MIT Press Cambridge, MA, EUA, p. 22-29.
- Constantinides, E. (2004). *Influencing the Online Consumer's Behavior: the Web Experience*. *Internet Research*, 14, 111-126.
- Cooper, C., Fletcher, J., Fyall, A., Gilbert, D., & Wanhill, S. (2008). *Tourism: Principles and Practice* (4a. ed.). Essex: Pearson Education Limited.
- Cruys, T. V. (2011). Two multivariate generalizations of pointwise mutual information. In *Workshop on Distributional Semantics and Compositionality Disco, Proceedings*. Stroudsburg, Pa, EUA: Association for Computational Linguistics, 16 - 20.
- Davenport, T.H. (2013). *At the Big Data Crossroads: turning towards a smarter travel experience*. Recuperado de <http://amadeusblog.com/wp-content/uploads/f-Big-Data-Report.pdf>
- Esuli, A., & Sebastiani, F. (2006). Determining term subjectivity and term orientation for opinion mining. In *Proc. of EACL*, 93-200.
- Evans M., Jamal A., & Foxall G. (2006). *Consumer Behaviour*. Sussex: Wiley.
- EY (2014). *Big data - Changing the way businesses compete and operate*. EYGM Limited. Recuperado de http://www.ey.com/Publication/vwLUAssets/EY_-

[_Big_data:_changing_the_way_businesses_operate/%24FILE/EY-Insights-on-GRC-Big-data.pdf](#)

- Foxall, G. R. (1990). *Consumer psychology in behavioral perspective*. New York: Routledge.
- Foxall, G. R. (1993). A behaviouristic perspective on purchase and consumption. *European Journal of Marketing*, 27, 07-16.
- Foxall, G. R. (1996). *Consumers in context: The BPM research program*. London: Routledge.
- Foxall, G. R. (1997). *Marketing psychology: The paradigm in the wings*. London: Macmillan.
- Foxall, G. R. (2001). Foundations of consumer behaviour analysis. *Marketing Theory*, 1, 165-199.
- Foxall, G. R. (2002). *Consumer behavior analysis: Critical perspectives*. London and New York: Routledge.
- Foxall, G. R. (2004). *Context and cognition: interpreting complex behavior*. Reno, NV: Context Press.
- Foxall, G. R. (2005). *Understanding consumer choice*. New York: Palgrave Macmillan.
- Foxall, G. (2007). Explaining consumer choice: coming to terms with intentionality. *Behavioural Processes*, 75, 129-145.
- Foxall, G. (2010). *Interpreting consumer choice: the behavioral perspective model*. New York: Routledge.
- Foxall, G. R. (Ed.). (2016a). *The Routledge companion to consumer behavior analysis*. Abingdon: Routledge.
- Foxall, G. R. (2016b). Consumer behavior analysis comes to age. In Foxall, G. R. *The Routledge companion to consumer behavior analysis* (pp. 3-22). Abingdon: Routledge.
- Foxall, G. R., Yani-de-Soriano, M. (2011). Influence of reinforcement contingencies and cognitive styles on affective responses: an examination of Rolls' theory of emotion in the context of consumer choice. *Journal of Applied Social Psychology*, 41 (10), 2508-2535.
- Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations - A case from Sweden. *Journal of Destination Marketing and Management*, 3(4), 198-209.
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iView*, 1-12.

- Goebel, M., & Gruenwald, L. (1999). *A Survey of Data Mining and Knowledge Discovery Software Tools. Sigkdd Explorations*, [s.l.], v. 1, n. 1, p.20-33, jun. 1999. Semestral. Recuperado de: <http://www.kdd.org/exploration_files/survey.pdf>.
- Goeldner, C. R., Ritchie, J. R. B. (2002). *Tourism: Principles, practices, philosophies* (9th ed.). New Jersey: John Wiley & Sons.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907--928.
- Haddi, E. (2015). *Sentiment Analysis: text pre-processing, reader views and cross domains* (Tese de Doutorado). Brunel University London, Londres, Inglaterra. Recuperado de <http://bura.brunel.ac.uk/handle/2438/11196>.
- Harrison-Walker, L. J. (2001). E-complaining: a content analysis of an Internet complaint forum. *Journal of Service Marketing*, 15(5), 397-412.
- Hatzivassiloglou, V., & Mckeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the Eighth Conference on European Chapter of The Association for Computational Linguistics*, 97, Madrid, Spain, 174-181.
- Hawkins, G. (2012). Will Big Data Kill All But The Biggest Retailers? In *The Promise And Challenge Of Big Data*, *Harvard Business Review*.
- Heerschap, N., Ortega, S., Priem, A., & Offermans, M. (2014) Innovation of tourism statistics through the use of new big data sources, *12th Global Forum on Tourism Statistics*, Prague, 15-16.
- Hong, J-Y & Lee, W-N (2008). Consumer complaint behavior in the online environment. In Clarke, S. (Org.) *End-user computing: concepts, methodologies, tools, and applications*. Hershey: Information Science Reference, 1607-1619.
- Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. *Proceedings of the 19th International Conference on Artificial Intelligence (AAAI '04)*, 755-760.
- Irudeen, R. & Samaraweera, S. (2013). Big data solution for Sri Lankan development: A case study from travel and tourism. In *2013 International Conference on Advances in ICT for Emerging Regions*, ICTer 2, Colombo.
- Jaccard, J., & Jacoby, J. (2010). *Theory construction and model building skills: A practical guide for social scientists*. (Methodology in the social sciences). New York: Guilford Press.
- James, J. (2016) *Data Never Sleeps 4.0*. Recuperado de: <https://www.domo.com/blog/data-never-sleeps-4-0/>
- Jeong, M., & Jeon, M. M. (2008). Customer reviews of hotel experiences through consumer generated media (CGM). *Journal of Hospitality & Leisure Marketing*, 17(1-2), 121-138.

- Justeson, J. S., & Katz, S. M. (1995). Technical terminology: some linguistic properties and an algorithm for identification in text. *Natural Language Engineering*, 1, 9-27. doi:10.1017/S1351324900000048.
- Kaiquan Xu, Stephen Shaoyi Liao, Jiexun Li, and Yuxia Song. 2011. *Mining comparative opinions from customer reviews for Competitive Intelligence*. *Decision support systems* 50, 4 (2011), 743–754.
- Kim, S.M., & Hovy, E. (2005). Automatic detection of opinion bearing words and sentences. In *Companion Volume to the Proceeding of the International Joint Conference on Natural Language Processing*.
- Kim, S.M., & Hovy, E. (2004). Determining the sentiment of opinions. In *Proc. of COLING*, 1367-1373.
- Yaakub, R. M., Li, & Feng, Y., 2011. Integration of Opinion into Customer Analysis Model in *proceedings of Eighth IEEE International Conference on e-Business Engineering*, pp. 90-95
- Laney, D. (2001). 3D data management: controlling data volume, velocity and variety. *META Group Research Note*, 6(70).
- Leal, M.R., Oliveira, W. C. C., & Brito, P. F. (2017). Avaliação de Desempenho de uma Ferramenta de Análise de Sentimentos baseada em Aspectos. In *Congresso de Computação e Tecnologias da Informação*, 19., Palmas, TO.
- Leng, Y., Noriega, A., Pentland, A. S., Winder, I., Lutz, N., & Alonso, L. (2016). Analysis of Tourism Dynamics and Special Events through Mobile Phone Metadata. *Proceedings of Data for Good Exchange (D4GX)*. New York, NY.
- Liddy, E.D. (2001). *Natural Language Processing*. In *Encyclopedia of Library and Information Science*, 2nd Ed. NY. Marcel Decker, Inc. Recuperado de <https://surface.syr.edu/cnlp/11/>
- Limberger, P.F., Anjos, F.A., Meiria, J.V.S., & Anjos, S.J.G. (2014). Satisfaction in hospitality on TripAdvisor.com: an analysis of the correlation between evaluation criteria and overall satisfaction. *Tourism & Management Studies*, 10(1), 59-65.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Liu, B. (2010). Sentiment analysis and subjectivity. In Indurkha, N. Damerau, F. J. *Handbook of Natural Language Processing*. 2. ed. Boca Raton, FL: Chapman And Hall/crc, Cap. 26, 627-666.
- Marko, K. (2015). *Using big data and machine learning to enrich customer experiences*. Recuperado de: http://www.forbes.com/sites/kurtmarko/2015/04/08/big-data-machine-learning_customer-experience/
- Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of*

the Association for Computational Linguistics: Human Language Technologies- Volume 1, pages 142–150. Association for Computational Linguistics, 2011.

- Mathieson, A., & Wall, G. (1982). *Tourism: Economic, Physical and Social Impacts*. New York: Longman House.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review*, 90 (10), 60-68
- Meijer, E. (2011). The world according to linq. *Communications of the ACM* 54(10), 45-51
- Mercier, K., Schroeck, M., & Shockley R. (2013). Analytics: The real-world use of big data in retail. *IBM Global Business Services*, Executive report.
- Miah, S., Vu, H., Gammack, J., & McGrath, M. (2016). A Big Data Analytics Method for Tourist Behaviour Analysis, *Information & Management*, ScienceDirect, EBSCOhost.
- Miguéns, J., Baggio, R., & Costa, C. (2008). Social media and Tourism Destination: TripAdvisor Case Study. *Proceedings of the IASK Advances in Tourism Research 2008 (ATR2008)*, Aveiro, Portugal, 194-199.
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4), 235-312.
- Nalini, L. E. G., Cardoso, M. M., & Cunha, S. R. (2013). Comportamento do consumidor: uma introdução ao Behavioral Perspective Model (BPM). *Fragmentos de Cultura*, 23 (4), 489-505.
- Nivre, J. (2005). *Dependency Grammar and Dependency Parsing*. Recuperado de: <http://stp.lingfil.uu.se/~nivre/docs/05133.pdf>
- Nivre, J., Hall, J., & Nilsson, J. (2006). Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of LREC*, v. 6, 2216-2219.
- Norvig, P. (2016). *How to Write a Spelling Corrector*. Recuperado de: <http://norvig.com/spell-correct.html>
- Noy, N. F., Mcguinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University. Stanford, California.
- Oliveira, W. C. C. (2015). *SentimentALL: Módulo para Análise de Sentimentos em Português*. 86 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Centro Universitário Luterano de Palmas, Palmas.
- Oliveira, W. C. C., & Brito, P. F. (2015). Utilização do Pointwise Mutual Information na Identificação de Expressões Multipalavras. In *XVII ENCOINFO – Congresso de Computação e Sistemas de Informação*, 17, Palmas, 25-34.

- Oliveira-Castro J. M., & Foxall G. R. (2005). Análise do comportamento do consumidor. In Abreu-Rodrigues J. A., Ribeiro, M. R. *Análise do Comportamento: pesquisa, teoria e aplicação*. São Paulo: Artmed.
- Padilla, O.T. (1992). *El turismo, fenómeno social*. México, Fondo de Cultura Económica.
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. of ACL*, 271-278.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using Machine Learning techniques. In *Proc. of CoRR*.
- Paraskevas A. (2005). The Impact of Technological Innovation in Managing Global Value Chains in the Tourism Industry, *OECD-Korea Conference on Global Tourism Growth: A Challenge for SMEs*, Gwangju.
- Pohl, Roberta H. B. F., Oliveira-Castro, Jorge M., Bertoldi, Lílian, & Lourenço, Ana Lúcia. (2006). Efeitos do nível de benefício utilitário sobre a duração do comportamento de procura por produtos. *Revista Psicologia Organizações e Trabalho*, 6(1), 233-264. Recuperado em 14 de junho de 2018, de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1984-66572006000100009&lng=pt&tlng=pt..
- Popescu, A., & Etzioni, O. (2005). Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 339–346. Association for Computational Linguistics.
- Pries, K. H., & Dunnigan, R. (2015). *Big data analytics: A practical guide for managers*. Boca Raton, FL: CRC Press.
- Regalado, A. (2013). The Data Made Me Do It. *MIT Technology Review*. Recuperado de: <https://www.technologyreview.com/s/514346/the-data-made-me-do-it/>
- Schroek, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P. (2012) *Analytics: The real-world use of big data. How innovative enterprises extract value from uncertain data IBM Institute for Business Value*. Recuperado de http://www-03.ibm.com/systems/hu/resources/the_real_word_use_of_big_data.pdf
- Shafiee, S., & Ghatari, A. R. (2016). Big Data in Tourism. *10th International Conference on E-Commerce with Focus on E-Tourism*, 1-7. doi: 10.1109/ECDC.2016.7492979
- Sigurdsson, V., Foxall, G. R., Saevarsson H. (2010). In-store experimental approach to pricing and consumer behavior. *Journal of Organizational Behavior Management*, 30 (3), pp. 234-246.
- Sigurdsson, V., Menon, R. G.V., Sigurdarson, J. P., Kristjansson, J. S., & Foxall, G. R. (2013). A test of a behavioral perspective model in the context of an e-mail marketing experiment. *The Psychological Record*, 63(2), 295-308. doi:10.11133/j.tpr.2013.63.2.005, 2013.

- Silva, M. J., Carvalho, P., & Sarmiento, L. (2012). Building a Sentiment Lexicon for Social Judgement Mining. In *International Conference on Computational Processing of Portuguese (PROPOR), Lecture Notes in Computer Science (LNCS)*, Coimbra.
- Simms, A. (2012). Online user-generated content for travel planning: different for different kinds of trips? *E-review of Tourism Research*, 10(3), 76-85.
- Sinclair, M. T., & Stabler, M. (1997). *The economics of tourism*. London: Routledge.
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1945/1984). The operational analysis of psychological terms. *The behavioral and Brain Sciences*, n. 7, pp. 547-553.
- Skinner, B. F. (1953). *Science and Human Behavior*. New York: Macmillan.
- Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1969). *Contingencies of reinforcement: a theoretical analysis*. Nova York: Appleton-Century-Crofts.
- Solomon, M. R. (2008). *O comportamento do consumidor: comprando, possuindo e sendo*. São Paulo: Bookman.
- Song, H., & Liu, H. (2017). Predicting Tourist Demand Using Big Data. *Analytics in Smart Tourism Design. Springer International Publishing*, 13-29.
- Souza, F. R. M. (2017). *Implementação de um Dashboard para a ferramenta SentimentALL* (Trabalho de Conclusão de Curso). Centro Universitário Luterano de Palmas - CEULP/ULBRA, Palmas, TO, Brasil.
- Souza, M., Vieira, R., Buseti, D., Chishman, R., & Alves, I. M. (2012). Construction of a Portuguese Opinion Lexicon from multiple resources. In *8th Brazilian Symposium in Information and Human Language Technology*.
- Staab, S., & Studer, R. (2004), *Handbook on Ontologies*. Steffen Staab (Editor), Rudi Studer (Editor). Berlin, Germany: *Springer-Verlag*.
- Stringam, B. B., Gerdes Jr. J., & Vanleeuwen, D. M. (2010). Assessing the importance and relationships of ratings on user-generated traveler reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 11(2), 73-92.
- Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications*, 36(7):10760–10773.
- Täckström, O., & McDonald, R. (2011). Semi-supervised latent variable models for sentence-level sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers* -Volume 2, pages 569–574. Association for Computational Linguistics.

- Telnarova, Z. (2010). Relational Database as a Source of Ontology Creation. University of Ostrava. Czech Republic. *Proceedings of the International Multiconference on Computer Science and Information Technology*. p 135-139.
- Todorov, J.C. (1985). *O conceito de contingência tríplice na análise do comportamento humano*. *Psicologia: Teoria e Pesquisa*, 1, 75-88.
- Tucker, P. (2013). Has Big Data Made Anonymity Impossible? *MIT Technology Review*. Recuperado de: <https://www.technologyreview.com/s/514351/has-big-data-made-anonymity-impossible/>
- Turney, P.D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL*, 417-424.
- OMT (1995). *UNWTO. Technical Manual No. 2: Collection for Tourism Expenditure Statistics*. Recuperado de: <https://pub.unwto.org/WebRoot/Store/Shops/Infoshop/Products/1034/1034-1.pdf>
- OMT (2010). *UNWTO. Tourism highlights*. Madrid: UNWTO. 2010. Recuperado de: http://www.unwto.org/facts/eng/pdf/highlights/UNWTO_Highlights10_en_HR.pdf
- Qiu, G., Liu, B., Bu, J., Chen, C. (2009). Expanding Domain Sentiment Lexicon through Double Propagation. In: *International Joint Conference On Artificial Intelligence*, 21., 2009, [s.l.]. Proceedings. 2009. p. 1199 - 1024.
- Qiu, G., Liu, B., Bu, J., Chen, C. (2011). *Opinion Word Expansion and Target Extraction through Double Propagation*. *Association For Computational Linguistics*. [s.l.], p. 9-27. 2011.
- Vinerean, S., Cetina, J., Dumitrescu, L., and Tichindelean, M. (2013). The effects of Social Media Marketing on Online Consumer Behaviour. *International Journal of Business and Management* Vol. 8, No. 14 , 66 - 79
- Wiebe, J.M., Bruce, R. F. & O'hara, T.P. (1999). Development and use of a gold-standard data set for subjectivity classification. In *Proc. of ACL*, 246-253.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3).
- Yu, H. & Hatzivassiloglou, V. (2003) Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the conference on Empirical methods in natural language processing*, pages 129–136. EMNLP-2003.
- Zephoria (2016), The top 20 valuable Facebook statistics, *Zephoria Digital Marketing*, Recuperado de: <https://zephoria.com/top-15-valuable-facebook-statistics/>
- Zikopoulos, P. & Eaton, C. (2011) *Understanding big data: analytics for enterprise class hadoop and streaming data*. New York: McGraw-Hill Osborne Media.