APLICAÇÃO DE FERRAMENTAS PARA COLETA E ANÁLISE DE DADOS EM LINGUÍSTICA

AN APPLICATION OF TOOLS FOR DATA COLLECTION AND ANALYSIS IN LINGUISTICS

Roberlei Alves Bertucci*
bertucci@utfpr.edu.br

O uso de tecnologias influencia diferentes campos de saber e atividade humanos, inclusive a linguagem. Nesse sentido, as potencialidades verificadas nos ambientes digitais para circulação de dados linguísticos precisam ser exploradas. Para isso, este trabalho descreve três ferramentas digitais relacionadas à coleta e análise de dados: primeiro, o aplicativo *Netvizz*, integrado ao *Facebook*, que auxilia na montagem de *corpora* com dados dessa rede; segundo, o *software Tropes*, capaz de analisar textos a partir do processamento lexical, indicando elementos como o estilo e frequência das categorias lexicais; finalmente, o programa *Linguakit*, que seleciona palavras-chave, apresenta a frequência de palavras e realiza análise de sentimentos, entre outras tarefas. Para testar as ferramentas, selecionamos um conjunto de dados, retirado da página do *El País Brasil* por ocasião da prisão do ex-presidente Lula. Após a coleta de comentários, a aplicação no *Tropes* mostrou uma ocorrência alta de conectivos e modalizadores, além de itens lexicais referentes à situação ("universo de referência"). Já a análise no *Linguakit* apontou, além da alta frequência de termos específicos na situação, elementos típicos da Comunicação Mediada por Computador (como abreviações), bem como um sentimento mais negativo associado aos comentários.

Palavras-chave: Linguagem e tecnologia. Coleta de dados. Análise automática de textos. Rede social.

Technologies modify different human activities and knowledge fields, including language. Thus, some potentialities verified on digital environments for linguistic data interaction must be explored. In order to discuss how to do that, this paper describes three digital tools related to data collection and analysis: firstly, *Netvizz* App, on Facebook, which contributes to organize *corpora* by using data from this social network; secondly, *Tropes software*, which analyses texts from a lexical process, describing elements like text style and word frequency; thirdly, *Linguakit* program, which selects keyword, presents word frequency and does some sentiment analysis, among other tasks. To show how these tools work, we collect data from

* Universidade Tecnológica Federal do Paraná, Brasil.

El País Brasil on Facebook, at the day which former president Lula has arrested. After this collection, *Tropes* analysis presented a higher frequency of both connectives and modal expressions, besides lexical items related to that situation. In turn, *Linguakit* analysis described current elements of Computer Mediated Communication (like abbreviations), besides a higher frequency of specific situation expressions, as well as a negative sentiment related to those comments.

Keywords: Language and technology. Data collection. Automatic text analysis. Social network.

1. Introdução

Se a reflexão sobre o papel da língua numa comunidade perpassa boa parte da nossa história, especialmente a partir da Era clássica, sua relação com a tecnologia é uma discussão relativamente recente. Autores como Auroux (2014), defendem que a escrita pode ser considerada como a primeira grande revolução tecnolinguística: é a técnica, repleta de reflexão sobre a sua formação e viabilidade (tecnologia), que coloca a língua como item fundamental do processo de conhecimento. Continuando, o autor refere como segunda revolução técnico-linguística a gramatização das línguas, cujo processo de reflexão sobre as possibilidades de organização e interpretação gerou produtos como gramáticas e dicionários. A gramatização assume um papel fundamental no processo de contato entre línguas, sobretudo na época das expansões de povos, a partir do século XV. Sem a escrita, no entanto, a gramatização não poderia ocorrer, pois apenas com o registro de uma língua se pode conhecê-la e estudá-la de modo científico.

Os estudos sobre as línguas, desenvolvidos de modo mais acentuado a partir do século XX, foram possibilitados por inúmeras tecnologias, para além do aparecimento da escrita e da gramatização massiva. A invenção do gravador, por exemplo, foi um passo decisivo no desenvolvimento de trabalhos em fonética, fonologia e variação (entre várias outras áreas), já que o registro da fala permitiu o estudo de uma língua com recurso à criação de bancos de dados. O computador, por sua vez, transformou-se tanto em uma ferramenta de (re) produção e/ou análise da linguagem, como em uma ferramenta de organização de dados (da fala e da escrita). É neste contexto que o presente trabalho pretende descrever o apoio computacional para coleta e análise de dados.

A possibilidade de uso das ferramentas que aqui trazemos tem relação direta com o Manifesto das Humanidades Digitais (2012): à medida que a sociedade se vai configurando no ambiente tecnológico e as opções recaem sobre o digital, mudam as condições de produção e divulgação dos conhecimentos podendo, sem dúvida, retirar-se benefícios de pesquisas com qualidade que contribuem para o enriquecimento do saber – tal como se verificou com o aparecimento da escrita. Nesse sentido, importa mostrar como muito dos saberes e das potencialidades que envolvem ambientes digitais podem estar na mão dos próprios usuários-pesquisadores: participando em uma rede social, um linguista pode coletar dados suscetíveis de fundamentar teorias sobre como a linguagem funciona numa determinada comunidade (inclusive numa comunidade 'digital'). Trata-se de uma justificativa que sustenta a elaboração do presente trabalho, já que, no caso acadêmico, se pode partir do princípio de que há uma relação bem estabelecida entre os estudantes e as tecnologias digitais, utilizadas para atividades cotidianas por meio da linguagem. Afinal, como afirma Coscarelli (2016, p. 11),

> as tecnologias digitais, disponíveis agora nos celulares e amplamente utilizadas por todas as camadas sociais como meio de comunicação, produção e disseminação de saberes, precisam ser estudadas e compreendidas. Os mais diversos contextos escolares precisam discutir e se apropriar dessas tecnologias para que os alunos também incorporem em suas vidas as inúmeras possibilidades oferecidas por equipamentos e aplicativos.

Deste modo, esperamos também contribuir de forma efetiva para a discussão, apropriação e, quiçá, incorporação da relação entre linguagem e tecnologia. Aliás, na área de estudos de *corpora* tem-se destacado que, mais do que apenas processar mais rapidamente informações, o computador tem sido um valioso aliado dos estudos linguísticos, tanto do ponto de vista de coleta e análise de dados, quanto do ponto de vista de desenvolvimento de ferramentas vinculadas à linguagem.

Note-se que recursos de análise da língua portuguesa com recurso a ferramentas gratuitas não é uma novidade (Sardinha 2005). No entanto, o presente artigo procura mostrar que há a possibilidade tanto de coleta de dados em ambiente de redes sociais, como a possibilidade de processamentos desses *corpora* com ferramentas específicas, o que constitui uma novidade.

Por outro lado, é importante destacar que este trabalho não se insere necessariamente no âmbito de novas teorias para Linguística de Corpus, tendo uma intenção relativamente modesta: descrever como é possível formar corpora e analisá-los em ambientes computacionais. Assim, não

pretendemos discutir como um dado corpus *deve* ser formado, que critérios estão subjacentes à sua formação, mas sim, como ele *pode* ser construído nesses ambientes. Na descrição que se pretende fazer das ferramentas, algumas perguntas foram cruciais para traçar o plano de trabalho: 1) que possibilidades oferecem as ferramentas no que respeita ao estudo da linguagem? 2) como podem essas ferramentas ser utilizadas por pesquisadores interessados em dados de rede social ou em análises computacionais de textos?

Para encontrar as respostas, descrevemos uma coleta de dados realizada na rede social Facebook, por meio do aplicativo *Netvizz* nela integrado. Em seguida, apresentamos a análise desse corpus por meio dos *softwares* Linguakit e Tropes. Como se verá adiante, quer pelos resultados da coleta quer pelas análises, há uma grande facilidade de acesso por parte do pesquisador da linguagem que deseja tomar o ambiente virtual como espaço para coleta e pesquisa com finalidades linguísticas.

O presente artigo está dividido da seguinte forma: na Seção 1, analisamos a relação entre linguagem e tecnologia de forma mais aprofundada, apresentando a importância dos *corpora* para as análises em linguística; na Seção 2, descrevemos as ferramentas utilizadas, bem como as análises realizadas por meio delas; em seguida, apresentamos as considerações finais.

2. Linguagem, tecnologia e corpus

2.1. Da relação entre tecnologia e linguagem

Toda sociedade se modifica naturalmente, dadas as condições de desenvolvimento e estruturação que a envolvem e, por isso, é inegável que o conhecimento tecnológico modifica o modo de agir e de pensar das culturas: à medida que equipamentos são desenvolvidos (veja-se o caso do microscópio,), novos objetos de estudo passam a fazer parte do cotidiano científico daquele povo (o estudo de micro-organismos, por exemplo). Por isso, Cupani (2016) revela que a tecnologia pode, sim, influenciar o modo de pensar e os resultados das pesquisas em diferentes ambientes. Sem dúvida, a tecnologia é um tema e uma área que envolve a vida de todas as sociedades pois, como afirma Cupani (2016, p. 9), "a tecnologia nos afeta e desafia qualquer que seja nossa atividade".

Por outro lado, nesse ambiente de afetação, a própria tecnologia passa a ser objeto de investigação: quando o ser humano encara a vida como um problema torna-se também capaz de procurar alternativas para o resolver (Ortega y Gasset 1939). Por isso, a inter-relação entre linguagem e tecnologia pode trazer perguntas (problemas) como: por que e para quê servem os

diferentes dispositivos criados? Ou ainda antes disso: o que permite que o homem desenvolva tais artefatos?

Uma possível resposta para esta última pergunta parece ser a relação entre tecnologia e planificação: como uma criação tem um objetivo, o artefato precisa de ser refletido. Por isso, a noção de planificação ganha destaque nessa perspectiva. De modo similar, Vieira Pinto (2005) argumenta que o planejamento é inerente à tecnologia; por outro lado, pode-se acrescentar o fato de que é a linguagem a facilitadora desse simbolismo (abstração) específico, capaz de permitir o planejamento e, consequentemente, a realização do artefato. Tal capacidade é, para Cassirer (1979, p. 49) a chave para se entender o próprio ser humano:

> Entre o sistema receptor e o sistema de reação, que se encontram em todas as espécies animais, encontramos no homem um terceiro elo, que podemos descrever como o *sistema simbólico*; esta nova aquisição transforma toda a vida humana. Em confronto com os outros animais, o homem não vive apenas numa realidade mais vasta; vive, por assim dizer, numa nova dimensão da realidade. (grifos do autor)

Como se lê acima, o sistema simbólico, ofereceu ao homem condições claras de diferenciação dos outros animais, mas sobretudo, de superar dimensões mais próximas e desejar, planejar e até mesmo realizar outras, trazendo-as à dimensão real. Nesse sentido, portanto, a linguagem exerceu um papel decisivo, já que assume o papel de modeladora dessa capacidade.

Pode-se dizer que é graças a essa capacidade que as inovações modificam o ambiente humano. Barton e Lee (2015) consideram que toda mudança tecnológica causa alguma mudança na vida social; ou seja, à medida que novas técnicas, processos e produtos aparecem, a vida das pessoas se modifica. Como a linguagem é parte essencial no processo, não podemos desconsiderar a relação plena entre texto/linguagem e tecnologias: linguagem modificando práticas sociais (e tecnológicas) e tecnologias modificando/influenciando práticas linguísticas.

Portanto, para esses autores, "o mundo está cada vez mais mediado pelo texto e a web é parte essencial dessa mediação" (Barton & Lee 2015, p. 29). Sendo assim, não apenas se constata um espaço de circulação textual mais abrangente e mais ubíquo, como também se verifica a necessidade de um estudo mais específico sobre a mobilização dos recursos linguísticos utilizados nas práticas linguísticas nesse ambiente, o que é, sem dúvida, um desafio para os estudos sobre linguagens.

Igualmente importante é o fato de esse ambiente, que criou gêneros textuais relativamente instáveis ou rapidamente mutáveis (como os memes ou menes), ter permitido uma concentração de dados capaz de proporcionar

aos pesquisadores da linguagem uma boa fonte de compilação para suas pesquisas linguísticas. Dessa forma, ao mesmo tempo em que a Web fornece espaço para a criação textual, novos dispositivos de compilação de dados para sua análise são desenvolvidos em diferentes instituições.

Deste modo, é possível falar sempre de uma valoração da tecnologia, seus processos e produtos. Consequentemente, Cupani (2016, p. 12) argumenta que "aquilo que denominamos tecnologia se apresenta, pois, como uma realidade polifacetada: não apenas em forma de objetos e conjuntos de objetos, mas também como sistemas, como processos, como modo de proceder, como uma certa localidade". Em outras palavras, se o mundo está cada vez mais permeado de tecnologia e a linguagem é a grande mediadora da sociedade, perante esta realidade lança-se o desafio aos estudiosos da área de identificarem objetos, processos e modos proceder na coleta e análise de dados.

Por conseguinte, procuramos descrever algumas possibilidades de utilização de ferramentas de coleta e análise textual, levando em conta tanto ambientes de grande circulação, como as redes sociais, como outros mais restritos, como os de atividades didáticas em sala. Mais especificamente, nossa intenção é descrever como se podem coletar dados de uma rede social por um aplicativo específico (*Netvizz*), bem como analisar tais dados a partir de outros aplicativos (*Linguakit* e *Tropes*).

Inegavelmente, a área de Linguística de Corpus tem se beneficiado do desenvolvimento de ferramentas capazes de gravar, armazenar e analisar dados de línguas naturais. A disponibilização dos dados eletrônicos bem como de ferramentas que permitem seu tratamento é um ponto fundamental para os resultados atingidos por pesquisadores da área. Como afirmam Raso e Melo (2012, p. 33):

> A maior parte dos corpora produzidos no Brasil são escritos, principalmente com material pertinente a jornais e gêneros acadêmico-científicos, utilizados, sobretudo, por grupos de pesquisa voltados para os estudos do léxico e desenvolvimento de ferramentas computacionais para o tratamento da linguagem natural.

Das palavras dos autores depreende-se, portanto, que os corpora orais representam o português brasileiro em menor número, como também os corpora de textos menos monitorados são igualmente raros. Obviamente, aspectos operacionais para a construção e manutenção de banco de dados são fatores decisivos dessa realidade. No entanto, pode-se também falar de uma falta de interesse por "gêneros escritos menores", na medida em que estes se situam no limiar entre a fala (a língua natural) e a escrita (a língua a ser

aprendida). Aqui, queremos assumir que o estudo de ambientes de escrita menos monitorados pode revelar-se muito produtivo para todas as áreas de estudos que tomam a linguagem (as relações sociais, portanto) como relevantes. Há, compreendemos, aspectos linguísticos bastante reveladores, não só no uso da estrutura da língua, como em sua força expressiva (ou discursiva).

Ainda que possa haver projetos a respeito de textos escritos em situações menos formais (textos de alunos de escola básica, por exemplo), pretendemos aqui destacar a possibilidade de estudos relativos a *corpora* com propósitos específicos, como é o caso dos textos produzidos em redes sociais, especialmente os comentários. Diferentes trabalhos têm dado o seu contributo com análises relevantes de fatores linguísticos em ambientes de interação *on-line* (Araújo & Leffa 2016; Barton & Lee 2015; Coscarelli 2016, entre outros), ainda que mais direcionados para as possibilidades pedagógicas.

Acreditamos, tal como ocorre com outros estudos que recorrem a *corpora*, que a utilização de textos produzidos em ambientes de interação digital pode fornecer elementos importantes para o estudo da língua, como seja a análise do discurso, o estudo de texto/gêneros, a variação linguística, entre outros. Paiva e Paredes-Silva (2012), por exemplo, discutem aspectos de variação e mudança observáveis na escrita, a partir de dados de textos jornalísticos. Ora, considerando que esse ambiente é bastante monitorado, poderíamos questionar se as descrições de variação aí observadas serão também encontradas em gêneros menos formais. Vale a pena referir, que a compilação de dados permanentes de tais gêneros não parece necessária, uma vez que não exige uma logística complexa, como no caso de gravações de textos orais. Por isso, sugerimos o uso de aplicativos, como o *Netvizz*, como grandes facilitadores do processo de compilação.

Importa salientar o uso de ferramentas computacionais em ambientes de ensino e/ou pesquisa, especialmente de materiais em língua. Finatto (2017) realça que, inegavelmente, houve uma série de avanços na compilação e análise de dados com apoio computacional em português nos últimos anos. Outro ponto fundamental é a importância das relações interdisciplinares geradas (ou exigidas) por tais contextos, já que especialistas dos sistemas de comunicação/computação estão em diálogo com linguistas.

Igualmente, Finatto (2017) destaca a importância do apoio computacional quer para a formação de *corpora* de diferentes modos e fontes, quer para sua análise textual, relevando a descrição de gêneros textuais/discursivos. Concordamos com a autora, pois tal apoio, especialmente para os linguistas, mostra que a tecnologia é um meio que ajuda a explicar diferentes fenômenos da língua natural.

O presente trabalho pretende contribuir para destacar o apoio dado por recursos tecnológicos na área da linguagem: a partir da coleta de dados em rede social, com o aplicativo *Netvizz*, procederemos à análise do *corpus* por meio de dois recursos: o *Linguakit* e o *Tropes*. Queremos descrever como um pesquisador pode encontrar nessas ferramentas um aliado na análise de *corpora* mais extensos.[1] A partir dos dados selecionados, indicaremos alguns recursos linguísticos recorrentes apontados pelos analisadores automáticos.

2.2. Linguística de Corpus

Apesar do presente trabalho, como dito, não fazer parte propriamente das pesquisas em Linguística de *Corpus*, o fato de descrever um modo de coleta de dados para análises linguísticas faz com que discutamos a relevância da montagem de corpora para pesquisas na área.

Em geral, a grande contribuição das pesquisas com *corpora* é permitir que os pesquisadores façam uma análise empírica, ou seja, de dados reais e, para isso, o uso de instrumentos de coleta e armazenamento de dados é essencial. Sardinha (2000, p. 325) destaca a importância dos recursos computacionais para a formação de *corpora* linguísticos.

> A Linguística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.

Como se vê, a própria concepção dos *corpora* está permeada de questões teóricas, já que o critério de formação do banco de dados é algo essencial tanto para a estrutura do *corpus* quanto para o acesso aos dados. Por isso, o recurso a meios computacionais é imprescindível. Na relação entre linguagem e tecnologia, o corpus é, portanto, um produto artificial (tecnológico), composto de uma série de etapas de reflexões e técnicas, cuja finalidade é contribuir para a pesquisa com dados empíricos sobre as línguas naturais.

---

1 Alguém poderia questionar se a limitação de caracteres de processamento de alguns *softwares*, como o *Tropes* e o *Linguakit*, não impediria análises mais extensas. Argumentamos que essa limitação não é tão grande: o *Tropes* processa até 32 mil caracteres e o *Linguakit* até 20 mil, o que é um número considerável, especialmente quando se trata de comentários on-line. Agradecemos a um parecerista que chamou a atenção para esse fato.

Dessa forma, para que seja possível a análise, é preciso que se estabeleçam alguns critérios. Sardinha (2000, p. 340–341) apresenta alguns dos principais pontos a serem levados em conta na concepção de um corpus, elencados e explicados no Quadro 1, a seguir.

Critério	Tipo	Composição
Modo	Falado	porções de fala transcritas.
	Escrito	textos escritos, impressos ou não
Tempo	Sincrônico	compreende um período específico
	Diacrônico	compreende vários períodos
	Contemporâneo	representa o período corrente
	Histórico	representa um período do passado
Seleção	Amostragem (*sample corpus*)	porções de textos ou de variedades textuais, planejado para ser uma amostra finita da linguagem como um todo
	Monitor	reciclada para refletir o estado atual de uma língua; opõe-se a corpora de amostragem
	Dinâmico ou orgânico	crescimento e diminuição são permitidos; qualifica o corpus monitor
	Estático	oposto de dinâmico; caracteriza o corpus de amostragem
	Equilibrado (*balanced*)	os componentes (gêneros, textos, etc.) são distribuídos em quantidades semelhantes (por exemplo, mesmo número de textos por gênero)
Conteúdo	Especializado	tipos específicos (em geral gêneros ou registros definidos)
	Regional ou dialetal	uma ou mais variedades sociolingüísticas específicas
	Multilíngue	Inclui idiomas diferentes
Autoria	Aprendiz	Os autores dos textos não são falantes nativos
	Língua nativa	Os autores são falantes nativos
Disposição interna	Paralelo	Os textos são comparáveis (p.ex. original e tradução)
	Alinhado	As traduções aparecem abaixo de cada linha do original
Finalidade	Estudo	O corpus que se pretende descrever
	Referência	Usado para fins de contraste com o corpus de estudo
	Treinamento ou teste	Construído para permitir o desenvolvimento de aplicações e ferramentas de análise

Quadro 1. Critérios de formação de *corpus*. Adaptado de Sardinha (2000, pp. 340-341)

Na Seção 2, em que apresentamos a formação e análise do corpus, apontamos os itens específicos referidos no Quadro 2 que melhor se enquadram na perspectiva.

Do ponto de vista da representatividade, Sardinha (2000) alega que um *corpus*, independe do tamanho, mas sim que, tais conjuntos de dados, agrupados conforme critérios do pesquisador, precisam ser representativos do uso linguístico em alguma circunstância. Assim, não se pode definir uma extensão mínima para de um *corpus*, mas é essencial que seja entendido como suficientemente representativo. Além disso, embora não tenha extensão pré-definida, a representatividade é, logicamente, melhor caracterizada quanto maior for o *corpus*. Quando específicos (e não abertos), os *corpora* são de acesso exclusivo do pesquisador, que o produz para uma finalidade específica, não sendo disponíveis para outros pesquisadores e, consequentemente, acabam não sendo "verificáveis, o que compromete a pesquisa em termos de sua replicabilidade e generabilidade" (Sardinha 2000, p. 348).

No entanto, vamos mostrar, neste trabalho, que o *Netvizz*, sendo um aplicativo que coleta dados a partir da rede social pela conta do próprio usuário, torna essa especificidade não mais um obstáculo. Em outras palavras, um mesmo *corpus* pode ser usado por diferentes pesquisadores, ainda que não estejam armazenados num único local, nem mesmo sejam coletados no mesmo dia. Se os critérios forem idênticos, o *corpus* ficará acessível de forma ubíqua.

Na próxima seção, descrevemos com mais detalhes os recursos computacionais utilizados na pesquisa, bem como os resultados que fornecem a partir da compilação dos dados retirados do *Facebook*.

3. Ferramentas tecnológicas: compilação e análise de dados

Nesta seção, apresentamos algumas ferramentas computacionais que consideramos relevantes para o trabalho com *corpora*, especialmente aqueles que provêm das próprias redes sociais. Como apontado anteriormente, pretendemos verificar as potencialidades das ferramentas no que toca ao trabalho de pesquisa linguística, tanto do ponto de vista da coleta, quanto do ponto de vista da análise.

3.1. *Netvizz*: coleta de dados no *Facebook*

Para Araújo e Leffa (2016), o uso das redes sociais ou sua aplicação no ensino já vem sendo objeto de investigações graças a recursos que os promovem e à área que envolvem. Do ponto de vista das pesquisas sobre a língua, no entanto, aparece, ainda, haver carência de análises de dados relativos às suas variações/mudanças ou mesmo de aspectos já verificados em análises de outros textos orais e escritos.

Isso justifica o presente artigo, já que a forma como ocorrem as práticas de linguagem nesse ambiente é diferente da conversa presencial. Nesse sentido, se, como afirma Recuero (2012, p. 28), "num diálogo, tudo é informação: elementos prosódicos (como o tom da voz, a entonação e as pausas da fala), elementos gestuais e, evidentemente, as palavras", no ambiente virtual todos os demais itens semiotizados precisam de ser interpretados, para que o leitor construa os sentidos ali presentes. Por isso, a navegação e a interação no meio digital, pela complexidade da convergência de diferentes semioses, distinguem-se da interação exclusivamente escrita. Essas questões abrem espaço inúmeras pesquisas, que podem envolver as reações numa publicação no Facebook, a forma de pontuação que ocorre em textos ali escritos, a construção de identidades no espaço virtual, ou mesmo a utilização de uma ferramenta de apreciação de um comentário, tudo em busca de compreender melhor como se dão as práticas e trocas sociais ali estabelecidas por meio da linguagem (Bertucci & Nunes 2017).

Embora o trabalho com redes sociais seja algo promissor nos estudos de linguagens, obter os dados dos materiais ali presentes, como postagens, reações, e comentários é algo bastante desafiador. Daí a relevância da apresentação do aplicativo *Netvizz* como meio facilitador dessa tarefa. Segundo a pesquisa realizada em abril de 2018 no Portal de Periódicos da Capes, o que melhor concentra trabalhos acadêmicos no Brasil, encontraram-se apenas três trabalhos que faziam referência ao aplicativo, dos quais apenas um na área de Linguística, que traçava uma relação próxima entre reações de raiva e divergência de opinião na rede (Bertucci & Nunes 2017). O desconhecimento do dispositivo justifica o presente trabalho, já que pode ajudar outros pesquisadores a coletarem dados da rede.

O *Netvizz* é um aplicativo integrado na rede social *Facebook*, disponibilizado a todos os usuários. Sua função é extrair dados de páginas e grupos que podem servir para pesquisas em diferentes áreas. Depois de extraídos, os dados precisam de ser consolidados em uma planilha específica, ou até sujeitos a análise por ferramentas de análise linguística. Embora esteja para

além do foco do nosso trabalho mostrar o passo a passo para sua utilização, apontamos os elementos necessários para a constituição de um *corpus* por meio desse aplicativo, sendo eles:

- a seleção da página fonte de dados;
- a seleção do período ou da quantidade de publicações da página;
- a consolidação em planilha (*e.g.* Excel);
- e a seleção/filtragem dos tipos de dados para análise.[2]

Para a compilação dos dados, o usuário precisa, em primeiro lugar, selecionar uma página ou grupo público de onde deseja extrair os dados. No nosso caso, extraímos dados da página do *El País Brasil*[3], um periódico de notícias bastante atuante na rede. Dali, entendemos selecionar uma publicação de alto engajamento e comentários, por sua vez selecionando alguns deles para análise em outras ferramentas. Na sequência, acessamos o aplicativo *Netvizz*.

Depois, selecionamos o período: optamos pelos dias 7 e 8 de abril de 2018, data em que ocorreu a prisão do ex-presidente Lula da Silva. Consideramos que o clima de apreensão e tensão (e acirramento dos ânimos políticos) que envolveu o Brasil naquele período propício à produção de comentários com características linguísticas menos monitoradas. Embora nossa proposta de análise não seja especificamente de variação linguística, optamos por valorizar produções mais espontâneas, o que ocorre mais facilmente em situações de envolvimento emocional.

Em seguida importamos os dados gerados pelo aplicativo em uma planilha Excel, o que nos permitiu filtrar e selecionar a publicação com maior número de comentários (1.150). A Figura 1, que se segue, mostra um resumo da tabela consolidada gerada a partir do *Netvizz*, com as dez publicações mais comentadas na página nos dias 7 e 8 de abril (de um total de 31 publicações). Como se pode ver, os dados revelam um alto engajamento nas publicações (da menor, com 872, à maior, com 10.211); além disso, há um grande envolvimento dos usuários nessas postagens por meio de comentários (de 187 a 1.150).

---

2 Para uma descrição detalhada do aplicativo *Netvizz*, sugerimos o trabalho de Rieder (2013).

3 Disponível em: www.facebook.com/elpaisbrasil. Consultado em: 26 abr. 2018.

	E	M	N	O	P	Q
1	post_message	likes_count_fb	comments_count_fb	reactions_count_fb	shares_count_fb	engagement_fb
2	Após a mobilização da militância diante da iminente prisão do ex-presidente o PT ainda não	1979	1150	3381	1217	5748
3	O Lula que volta à prisão 38 anos depois ainda conserva muitas coisas do ousado sindicalist	1731	504	2147	450	3101
4	Há apenas uma década tudo era muito diferente. Em 2008 enquanto a Europa e os EUA mer	4219	369	5864	2040	8273
5	A segregação espacial pontua o dia a dia do Condomínio Laranjeiras onde um exército de se	1756	308	2991	1302	4601
6	O líder que protagonizou três décadas de política brasileira dedicou boa parte do discurso a	1612	280	2174	164	2618
7	Editorial | O Brasil deve realizar eleições num clima de estabilidade. Políticos juízes e milita	343	269	656	42	967
8	Segundo os dados coletados pelo grupo de Piketty a fatia do 1% mais rico de brasileiros fica	3786	241	5298	4672	10211
9	Repúdio ou silêncio. A prisão do ex-presidente Luiz Inácio Lula da Silva que se entregou à p	1904	196	2255	344	2795
10	Ex-presidente fez um discurso por quase uma hora a apoiadores diante Sindicato dos Metalu	2261	188	2811	335	3334
11	Tudo começou em março de 2014 quando em uma operação rotineira sobre crimes financei	419	187	550	135	872

Figura 1. Tabela consolidada – dados do *Netvizz*/Facebook
Fonte: O autor

Isso feito, decidimos submeter uma parte dos comentários da postagem selecionada (149) para análise nas ferramentas *Tropes* e *Linguakit*. Nossa intenção era verificar quais elementos linguísticos poderiam ser destacados desse *corpus* com 31 publicações.

Antes de passarmos à análise, propriamente dita, poderíamos classificar o corpus que montamos a partir da tipologia apresentada.

Critério	Tipo	Netvizz/*Corpus* atual	Explicação
Modo	Falado	escrito	Textos escritos (comentários) de rede social.
	Escrito		
Tempo	Sincrônico	contemporâneo	Representa o período corrente (07-08/04/2018)
	Diacrônico		
	Contemporâneo		
	Histórico		
Seleção	Amostragem (*sample corpus*)	amostragem	Com porções de textos ou de variedades textuais do Facebook, foi planejado para ser uma amostra finita da linguagem como um todo naquele ambiente.
	Monitor		
	Dinâmico ou orgânico		
	Estático		
	Equilibrado (*balanced*)		
Conteúdo	Especializado	especializado (só comentários)	tipos específicos (em geral gêneros ou registros definidos)
	Regional ou dialetal		
	Multilíngüe		
Autoria	aprendiz	língua nativa	Os autores dos textos não são falantes nativos, não havendo identificação de falantes de outras línguas)
	língua nativa		

Critério	Tipo	Netvizz/*Corpus* atual	Explicação
Disposição interna	Paralelo	Não se aplica	
	Alinhado		
Finalidade	estudo	treinamento/teste	Foi construído para permitir o desenvolvimento de aplicações e ferramentas de análise.
	referência		
	treinamento ou teste		

Quadro 2. Formação de *corpus* com Netvizz
Fonte: o Autor (adaptado de Sardinha, 2000)

Como nosso objetivo, primeiro, aqui, é descrever o funcionamento das ferramentas, a representatividade do corpus é limitada: seu papel é servir de teste para mostrar como dispositivos computacionais analisam textos escritos em português brasileiro. Além disso, preferimos focar no gênero *comentários* porque, como dissemos na escolha do período, deve revelar um menor monitoramento linguístico por parte dos usuários, em grande parte, adeptos de polos opostos sobre a questão da prisão do ex-presidente. A seguir, antes da apresentação das demais ferramentas e da indicação da análise, exemplificamos o *corpus* com os 10 primeiros comentários dos 149 coletados e analisados aqui.

Vc está justificando violência ? No caso dos tiros está em investigação ,ja agressão de ontem é nítida de onde veio ... Mas vc está querendo justificar ... Ai é doença mesmo
As duas violências foram feitas por fanáticos covardes.
Os vídeos são incontestáveis. Não era um "opressor". Foi um cidadão agredido de forma completamente desproporcional e violenta. Lamentável.
Quem mandou bater em todo mundo nos protestos de 2013?
Vitor Costa Isso justifica o quê? Guerra?
A esquerda está apanhando calada há muito tempo, está sendo vilipendiada em tudo. Todo mundo sabe que este indivíduo infelizmente foi lá provocar num momento muito sensível. Provocaram a esquerda demais, a esquerda vem sofrendo todo tipo de agressão e infelizmente vai piorar para o lado da esquerda mais uma vez.
A primeira coisa que eu pensei quando li este post ! Muito cinismo desta página!
Ai, lá vem o Cosimo Lowbike de novo, tá me perseguindo? Onde você leu eu defendendo violência? Me espantei com a chamada da reportagem falando de casos de violência dos últimos dias, como assim? E você Cosimo me viu falando contra a violência em outro post há uma hora atrás. Me poupe!

Injustificável! Mas como apareceu no JN.....soy contra. Fantoche é o c.!
Ricardo Oliveira..a esquerda é culpada por tudo isso ,aliás por onde a esquerda passa causa destruição etc .. Resultado ta ai ,vc mesmo justificando violência e se viabilizando etc ..

Quadro 3. Apresentação de 10 dos 149 comentários analisados, entre os 1.150 coletados
Fonte: Netvizz/Facebook.

Com a coleta finalizada, o próximo passo é a análise dos dados. Para isso, vamos utilizar igualmente ferramentas computacionais que podem auxiliar o pesquisador, especialmente com *corpora* digitais.

3.2. *Tropes:* análise de dados

Criado nos anos 90, pela *Semantic-Knowledge* (Acetic), com sede em Paris, o *Tropes* é uma das ferramentas que a empresa desenvolveu para análise textual. Por meio da incorporação de conhecimentos advindos da área de Processamento de Linguagem Natural, o *software* foi criado servir áreas diversas, tais como os Sistemas de Informação, a Sociologia e a Linguística. O intuito da ferramenta focou-se em quatro aspectos na análise dos textos, apresentados na Figura 2, a seguir: resumo (*summarization*), classificação semântica (*semantic classification*), análise quantitativa (*quantitative analysis*) e descoberta de conhecimento (*knowlege discovery*).[4]

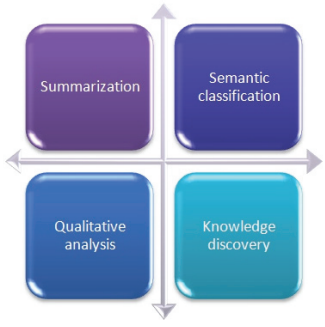

Figura 2. Análises textuais possíveis com o *Tropes*
Fonte: Tropes.

4 Mantivemos a figura em inglês pela fidelidade à fonte: embora o *software* tenha versão em português, a página está em inglês.

Tal como fizemos com o *Netvizz*, buscámos trabalhos no portal de periódicos da Capes e encontrámos apenas o trabalho de Araújo (2017) sobre o uso desse aplicativo em centros de pesquisa brasileiros. Ainda assim, o trabalho discutia a caracterização do gênero entrevista na língua espanhola. Em seu texto, Araújo (2017, p. 300) descreve que:

> [este] *software* destaca-se pelo processamento semântico de textos em línguas naturais. Para descrever as características dos enunciados em análise, o Tropes 7.2.3 vale-se de critérios linguísticos pré-programados e os associa às estruturas linguísticas encontradas nos textos processados.

Por suas características de análise lexical, o programa faz um processamento refinado do qual decorrem informações tais como o "estilo textual", o contexto básico relativo ao texto, denominado ali de "universo de referências" (sendo uma espécie de mapeamento do repertório mobilizado no texto); e dados quantitativos referentes a classes lexicais.

Para isso, é preciso que o pesquisador configure os documentos em formatação textual, preferencialmente em extensão de página web filtrada. Feito isso, ao acessar o aplicativo, o pesquisador poderá solicitar ao *software* o processamento dos textos salvos em uma pasta, individualmente ou de forma conjunta. A Figura 3 apresenta a tela de abertura da ferramenta.

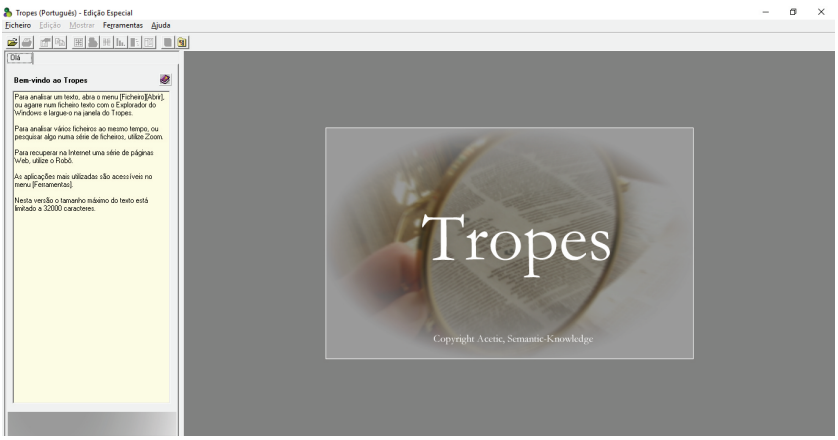

Figura 3. Tela inicial do *Tropes*
Fonte: Tropes

Com a aplicação no *corpus* de treinamento aqui apresentado, a intenção era descrever suas características textuais a partir desse *software*. Assim,

elementos como "estilo textual", "universo de referência" e algumas "categorias lexicais" foram selecionadas para apresentação neste trabalho. Embora a publicação contasse com 1.150 comentários, a restrição a 149 foi necessária devido ao número reduzido de caracteres para análise que possibilita o programa. O resultado é apresentado pela ferramenta em seções específicas. Começamos pelo estilo (Figura 4).

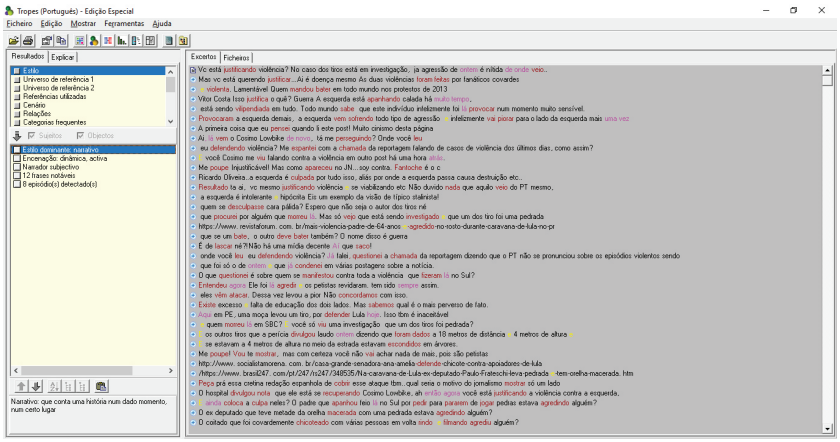

Figura 4. Estilo no *Tropes*
Fonte: Tropes

Dos diferentes estilos categorizados no aplicativo (narrativo, descritivo, argumentativo e enunciativo), o *corpus* com 149 comentários foi analisado como predominantemente narrativo. Ainda que o esperado fosse um estilo argumentativo, percebe-se que o aplicativo justifica a escolha do estilo marcando no texto os elementos que o levaram a isso, entre eles, verbos de ação (em oposição a estativos), advérbios de tempo ("ontem" e "agora", por exemplo), uso recorrente do conectivo "e". Aqui, fica claro que essa predominância tem relação direta com o evento em si, denotando uma sequência de fatos que culminaram na prisão do ex-presidente.

Com relação ao "universo de referência", o aplicativo apontou 28 agrupamentos de categorias de substantivos (referências), tais como "vida humana" (144 ocorrências), em itens como *violência* e *pessoas*; "conceitos gerais" (106 ocorrências), em casos como *esquerda* e *manifestações*; e "comunicação e mídia" (26 ocorrências), em exemplos como *vídeo* e *jornal.* Os casos sublinhados aqui apresentam uma noção do tema da publicação e dos eventos que estavam envolvidos no período da prisão, sobretudo aqueles envolvendo

manifestações contrárias e favoráveis a Lula, bem como episódios de violência que se registaram na ocasião. O resultado é apresentado na Figura 5.

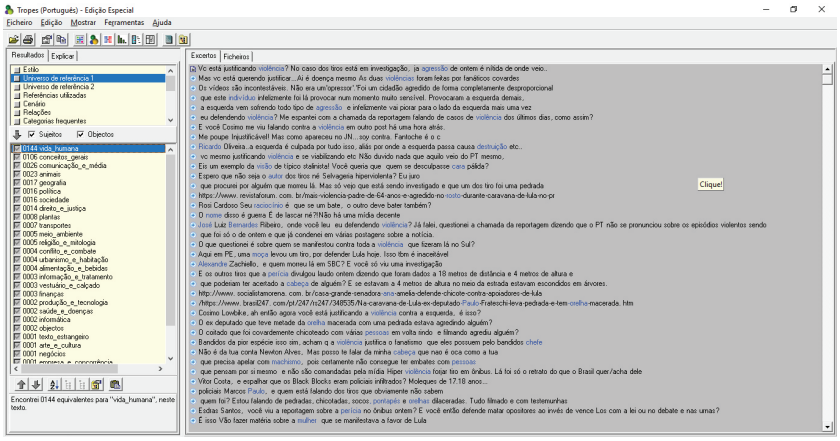

Figura 5. Universo de referência no Tropes
Fonte: Tropes

Em seguida, selecionamos as categorias lexicais com maior frequência: verbos factivos (61%); conectivo de adição (56,3%); modalização de negação (22.4%) e de tempo (21.8%); adjetivos subjetivos (56.1%); pronomes de primeira e segunda pessoa (15.1% e 24.5%, respectivamente). O pesquisador pode acessar todas as categorias analisadas pelo Tropes, assim como, ao clicar sobre cada uma, observar as expressões que a exemplificam no *corpus*. As Figuras 6 e 7, que se seguem, apresentam os dados completos.[5]

5 Embora as Figuras 6 e 7 apresentem muitos dados e pudessem ser apresentadas em tabelas, preferimos as figuras para que o leitor veja claramente como o aplicativo apresentada os textos analisados.

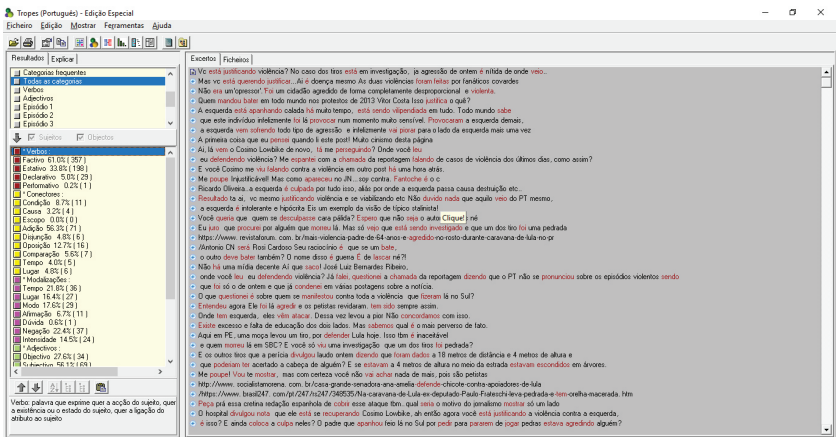

Figura 6. Categorias lexicais no Tropes – parte 1
Fonte: Tropes

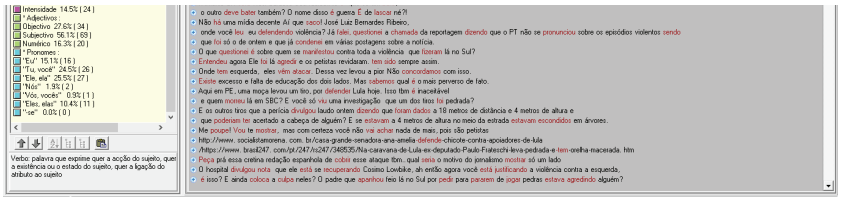

Figura 7. Categorias lexicais no Tropes - parte 2
Fonte: Tropes

Alguns destaques poderiam ser feitos nesse *corpus*, do ponto de vista linguístico: além do número produtivo de verbos factivos (de ação), o que representa bem o episódio da prisão, como dito antes, o baixo índice de primeira pessoa do plural (*nós, nos*) associado ao alto índice de outros pronomes representam a polaridade que tem dominado o cenário de discussões políticas no País. Por outro lado, a vasta ocorrência de itens de conexão e modalização fornecem dados importantes quanto às estratégias textuais de coesão e posicionamento nos comentários *on-line*, que poderiam ser comparados com *corpora* distintos. Tudo isso justifica o trabalho de coleta e análise por meio de ferramentas computacionais.

A seguir, debruçamo-nos sobre o *Linguakit* e mostramos de que forma este portal pode contribuir para a análise linguística.

3.3. *Linguakit*

Desenvolvido pelo *Cilenis Language Technology*, da Universidade de Santiago de Compostela, o *Linguakit* é um *site* multilíngue com diversas ferramentas de uso linguístico, baseadas em Processamento de Linguagem Natural, tais como resumidor, analisador de sentimentos ou de frequência de palavras entre muitas outras (Figura 8). A maior parte dessas ferramentas é de uso gratuito e apresenta resultados interessantes no que diz respeito à análise de dados.

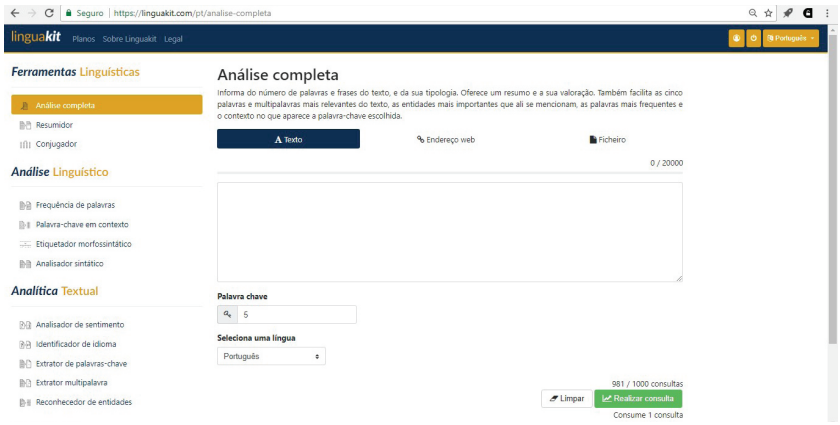

Figura 8. Página inicial do Linguakit
Fonte: Linguakit.

Para apresentá-la, neste trabalho, tomamos 149 textos da postagem mais comentada no *El País Brasil*, entre 7 e 8 de abril de 2018, conforme descrevemos na subseção anterior. Embora a publicação tivesse 1.150 comentários, foi feita uma seleção de 149 por restrição do programa (conferir a primeira nota de rodapé). Começamos a descrição da análise do *Linguakit* pelo extrator de palavras-chave (Figura 9).

Figura 9. Nuvem de palavras do Linguakit
Fonte: O autor.

Nesse caso, o resultado nos permite observar os elementos mais proeminentes nos 149 comentários: percebemos que "Lula", "petistas" e "selvageria" foram os elementos mais recorrentes nesse *corpus*. Assim, o extrator aponta para o contexto sobre o qual os comentários versavam.

Cabe observar que itens como "q" ou "vc" foram igualmente bem citados nos textos; nesse sentido, vale destacar que o aplicativo pode contribuir para pesquisas sobre elementos próprios da linguagem *on-line*, já que mostra sua frequência. Como defendem alguns autores (Recuero, 2012; Coulmas, 2014), a Comunicação Mediada por Computador (CMC) é uma nova e importante área de pesquisa sobre as modificações linguísticas que ali podem ocorrer (tanto na fala como na escrita): não só uma oralização é comum (uso de emoticons ou repetição de letras na tentativa de indicar a entonação ou ação, por exemplo), como a presença de abreviações e inovações podem ser descritas. A tabela a seguir mostra a frequência de alguns desses itens.

Tabela 1. Frequência de itens CMC do Linguakit

frequência	item
13	vc
7	q
4	tbm
4	vcs
4	post
3	pq
2	tb
1	uhhhhh
1	uéeeeee
1	heim
1	kkkkkkk

Fonte: O autor.

O resultado acima nos leva a algumas conclusões: primeiro, que a abreviação parece ser o tipo de item da CMC mais recorrente; depois, que entre elas, "vc" e "q" estão praticamente estabelecidas. Naturalmente, isso deve ser confrontado com outros dados, especialmente levando em conta os tipos de páginas e postagens realizadas na rede social. De qualquer modo, conseguimos mostrar que a ferramenta aponta para elementos importantes no que diz respeito a essa estratégia de escrita em ambiente digital. Sem dúvida, tanto essas questões como aquelas apontadas pelo Tropes podem servir para descrição do gênero *comentário on-line*, sendo igualmente possível a busca por elementos de variação, especialmente na escrita, tal como nos apontam Paredes e Silva (2012) ser um ponto fundamental nas pesquisas da área.

O último item que desejamos apresentar no *Linguakit* é sua ferramenta de análise de sentimentos. Segundo Benevenuto *et al.* (2015, p. 2),

> o principal objetivo da análise de sentimentos é definir técnicas automáticas capazes de extrair informações subjetivas de textos em linguagem natural, como opiniões e sentimentos, a fim de criar conhecimento estruturado que possa ser utilizado por um sistema de apoio ou tomador de decisão.

Para fazer a classificação, o software é programado para medir a polaridade da frase, sendo ela classificada, no *Linguakit*, de forma ternária, a

saber: positiva, negativa ou neutra. Novamente, Benevenuto *et al.* (2015, p. 3) nos explicam a diferença:

> (...) a frase "Como você está bonita hoje" é *positiva* e a frase "Hoje é um péssimo dia" é *negativa*, já a frase "Hoje é 21 de Outubro" não possui polaridade e normalmente é classificada como *neutra*. (grifos dos autores)

Nesse sentido, a polaridade tem relação direta com os itens que, na sentença, podem ser mais associados com questões subjetivas, apontadas, por exemplo, por adjetivos.

Considerando o tema dos comentários analisados no presente trabalho, bem como os resultados prévios a respeito das palavras mais frequentes no contexto, poderíamos levantar a hipótese de que os comentários tenderão a se encontrar no polo negativo, já que palavras como "selvageria", "violência" e "bandido" apareceram com frequência na nuvem de palavras. A Figura 10, a seguir, apresenta os resultados do *Linguakit*.

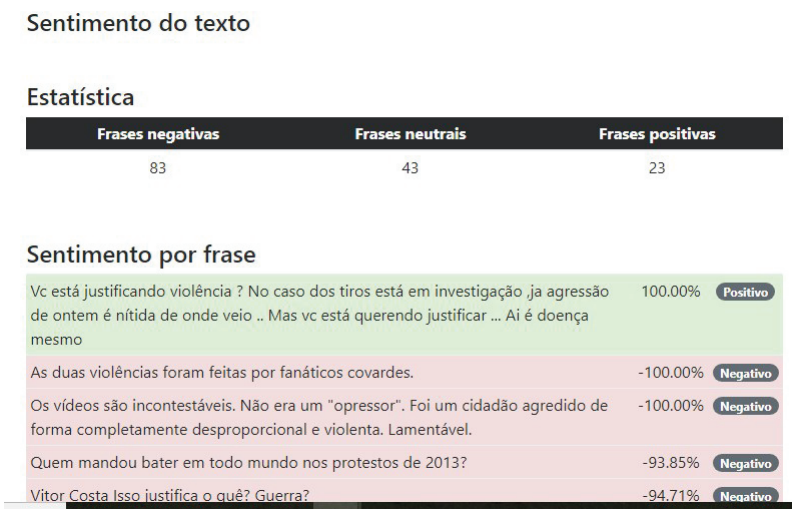

Figura 10. Análise de sentimento do *Linguakit*
Fonte: O autor

De fato, como se esperava, houve uma predominância de frases analisadas como negativas no contexto dos 149 comentários selecionados no *El País Brasil*. Embora essa área seja menos explorada por linguistas, não deixa de ser importante chamar a atenção para essa funcionalidade do *Linguakit*.

Nesta seção, apresentamos algumas das principais funcionalidades do *Tropes* e do *Linguakit*, ferramentas de análise textual digitais que podem contribuir para o trabalho dos linguistas.

4. Considerações finais

Neste trabalho, descrevemos tomamos um corpus de rede social (*Facebook*), por meio do aplicativo *Netvizz* nela integrado e o submetemos a uma análise automática de textos, por meio dos aplicativos *Tropes* e *Linguakit*. O primeiro software indicou questões textuais importantes, como o estilo geral do corpus, o contexto de referência e as categorias frequentes. Com tais dados, sugerimos que pesquisas sobre tais tópicos poderiam ser facilitadas com a ferramenta, que faz todo o trabalho de mapear as ocorrências no texto. A segunda ferramenta apontou para análises importantes, como o índice de palavras-chave, que indicam a frequência mais alta de termos no corpus, como também a ocorrência de termos próprios da CMC (abreviações, inovações e repetições de letras). Finalmente, mostramos a funcionalidade "análise de sentimentos" e, a partir das questões lexicais apontadas pelas ferramentas, ligadas a violência, sugerimos que os comentários seriam analisados, em sua maioria, como negativos, o que de fato ocorreu: o aplicativo apontou um total de 83 frases negativas, 43 neutras e apenas 23 positivas.

Em suma, as análises apresentadas têm como intenção chamar a atenção de pesquisadores para novas estratégias de coleta e análise de dados, especialmente em redes sociais. Sua importância vem sendo discutida por diferentes autores. Sendo um espaço de engajamento, é ali também que ocorrem manifestações linguísticas relevantes. Nesse sentido, Benevenuto *et al.* (2015, p. 2) argumentam que

> as redes sociais são a criação de uma revolução digital, permitindo a expressão e difusão das emoções e opiniões através da rede. De fato, redes sociais são locais onde as pessoas discutem sobre tudo expressando opiniões políticas, religiosas ou mesmo sobre marcas, produtos e serviços.

Para os linguistas, dedicados ao estudo da linguagem em suas diferentes manifestações, parece fundamental esse olhar voltado a um ambiente tão importante como o digital. Esperamos que este trabalho tenha contribuído para isso.

Referências

Araújo, J. Leffa, V. (Ed.) (2016). *Redes sociais e ensino de línguas: o que temos de aprender?*. (1ª ed.) São Paulo: Parábola Editorial.

Araújo, L. S. de. (2017). O gênero entrevista radiofônica em comunidades hispânicas: um aporte da Análise Textual Automática. *Domínios de Lingu@gem*, *11*(2), 289–312. Disponível em: DOI: 10.14393/DL29-v11n2a2017-2. Consultado em: 23. abr. 2018.

Auroux, S. (2014). *A revolução tecnológica da gramatização*. (3ª ed.) Campinas, Brasil: Editora da UNICAMP.

Barton, D. & Lee, C. (2015). *Linguagem online: Textos e práticas digitais*. M. C. Mota (Trad.). (1ª ed.) São Paulo: Parábola Editorial.

Benevenuto, F., Ribeiro, F. & Araújo, M. (2015). *Métodos para Análise de Sentimentos em Mídias Sociais.* Short course in the Brazilian Symposium on Multimedia and the Web (Webmedia). Manaus. Disponível em: <http://homepages.dcc.ufmg.br/~fabricio/download/webmedia-short-course.pdf>. Consultado em: 26 abr. 2018.

Bertucci, R. A. & Nunes, P. A. (2017). Interação em rede social: Das reações às características do gênero comentário. *Domínios de Lingu@gem*, *11*(2), 313–338. DOI: 10.14393/DL29-v11n2a2017-3.

Cassirer, E. (1979). *Antropologia filosófica: Ensaio sobre o homem*. São Paulo: Mestre Jou.

Coscarelli, C. V. (2016). Navegar e ler na rota do aprender. In C. V. Coscarelli (Ed.), *Tecnologias para aprender*. São Paulo: Parábola.

Coulmas, F. *Escrita e Sociedade*. (2014). São Paulo: Parábola Editorial.

Cupani, A. (2016). *Filosofia da tecnologia*: um convite. Florianópolis: Editora da UFSC.

Finatto, M. J. B. (2017). Descrição de gêneros textuais/discursivos com apoio computacional. *Domínios de Lingu@gem*, *11*(2), 282–288. doi: 10.14393/DL29-v11n2a2017-1.

Linguakit. http://linguakit.com/pt/. Consultado em: 26 abr. 2018

Manifesto das Humanidades Digitais. THATCamp Paris. 2012. Disponível em: <https://humanidadesdigitais.org/manifesto-das-humanidades-digitais> Consultado em: 26 abr. 2018.

Netvizz. https://apps.facebook.com/netvizz/. Consultado em: 26 abr. 2018

Ortega y Gasset, J. (1965). *Meditación de la técnica*. Madrid: Espasa-Calpe, (orig. 1939).

Paiva, C. & Paredes-Silva, V. L. (2012). Cumprindo uma pauta de trabalho: Contribuições recentes do PEUL. *Alfa*, São Paulo, *56*(3), 739–770.

Portal de Periódicos da Capes. *www.periodicos.capes.gov.br*. Consultado em: 26 abr. 2018.

Raso, T. & Mello, H. (Eds.) (2012). *C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG.

Recuero, R. (2012). *A conversação em rede: comunicação mediada pelo computador e redes sociais na Internet*. Porto Alegre: Sulina.

Rieder. B. (2013). Studying Facebook via data extraction: The Netvizz application. *Proceedings of the 5th Annual ACM Web Science Conference*, 346–355.

Sardinha, T. B. (2000). Corpus Linguistics: History and problematization. *DELTA*, *16*(2), 323–367. Disponível em: <http://dx.doi.org/10.1590/S0102-44502000000200005> Consultado em: 26 abr. 2018.

Sardinha, T. B. (2005). Trazendo a língua portuguesa para o computador. In T. B. Sardinha (Ed.), *A Língua Portuguesa no computador* (pp. 269–295). Campinas: Mercado de Letras & Fapesp.

Tropes. *http://www.semantic-knowledge.com/tropes.htm*. Consultado em: 26 abr. 2018.

Vieira Pinto, Á. (2005). *O conceito de tecnologia*. Rio de Janeiro: Contraponto.

[recebido em 26 de abril de 2018 e aceite para publicação em 30 de março de 2019]