

**IDIOMA E PERSPECTIVA: UMA ANÁLISE DA VISÃO DE MUNDO
A PARTIR DA ESCOLHA DE PALAVRAS USADAS EM NOTÍCIAS
SOBRE INDÍGENAS NO BRASIL E NO PARAGUAI**

Idioma y Perspectiva: un Análisis de la Visión de Mundo desde la Elección de Palabras Utilizadas en Noticias Acerca de los Indígenas en Brasil y en Paraguay

Tiele dos Santos KAWARLEVSKI
Universidade Federal do Rio Grande do Sul
tiele.kaw@gmail.com
<https://orcid.org/0000-0001-8408-4106>

RESUMO: O artigo propõe uma análise das palavras mais utilizadas em notícias sobre os povos indígenas, nos portais informativos mais acessados no Brasil e no Paraguai. Dado que o Paraguai, diferentemente do Brasil, tenha uma língua de origem indígena, o guarani, como um dos idiomas oficiais, a hipótese é de que naquele país seja perceptível a diferença na forma como os não indígenas se referem a tais povos, devido ao reconhecimento e possível aproximação manifestados na língua. As notícias, compiladas no período de janeiro de 2016 a dezembro de 2017, foram analisadas a partir da Metodologia da Linguística de *Corpus*. Usou-se o programa AntConc para a extração de informações linguísticas; sobretudo, a função *lista de palavras*, que possibilitou a identificação das unidades lexicais de maior uso. Os resultados da amostra apontam uma confirmação da hipótese, ilustrada nas listas de palavras mais frequentes.

PALAVRAS-CHAVE: Léxico; Notícias; Indígenas.

RESUMEN: Este trabajo propone analizar las palabras más utilizadas en las noticias sobre los pueblos indígenas en los periódicos en línea con mayores accesos en Brasil y en Paraguay. Dada la diferencia de Brasil, hay en Paraguay una lengua de origen indígena como idioma oficial, se sustenta como hipótesis que en aquel país hay una diferencia perceptible en el modo cómo los no indígenas se refieren a esos pueblos, debido aparentemente a que la visión de mundo del guaraní se proyecta sobre los hablantes de español en aquel país. Se analizan noticias recopiladas entre enero de 2016 y diciembre de 2017 usando como metodología la Lingüística de *Corpus*. Se utilizó el programa AntConc para extraer informaciones lingüísticas; se empleó sobre todo la función *lista de palabras*, que permite la identificación de las unidades léxicas más frecuentes. Los resultados confirman la hipótesis ilustrada en la lista de palabras más frecuentes.

PALABRAS CLAVE: Léxico; Noticias; Indígenas.

INTRODUÇÃO

A forma como os povos indígenas estão representados diante de comunidades não indígenas é uma preocupação de diferentes áreas de estudo, especialmente das que se ocupam em analisar o comportamento humano em sociedade. No âmbito das ciências sociais, antropólogos e sociólogos observam a evolução dessa sociedade e analisam os indígenas em meio a tal evolução (DORTIER, 2010). No âmbito dos estudos linguísticos, entre outras áreas que podem ter seus estudos associados a essa temática, existe a área da Etnolinguística, que busca compreender os efeitos do contato entre língua e sociedade – principalmente, quando se fala nas estruturas de uma e de outra (DURANTI, 2000 *apud* TEILLIER et al., 2016); e a área da Linguística Cognitiva, que relaciona os conhecimentos linguísticos a outros processos da mente humana (CUENCA; HILFERTY, 1999).

Embora essas áreas de estudo tenham um importante ponto de convergência entre elas, isso é, o interesse em analisar as conexões entre cultura e sociedade, propõe-se, aqui, apresentar um estudo que está intimamente ligado às duas últimas áreas citadas no parágrafo anterior. Ainda assim, não deixa de ser uma preocupação social ocupar-se de qualquer que seja das problemáticas que envolvem o contato entre indígenas e não indígenas. Todo estudo que vise avaliar essa relação pode vir a causar impacto não somente nas ciências da linguagem, mas pode, também, constituir uma contribuição para as políticas linguísticas – uma política pública urgente atualmente.

Norteado por essa noção, o objetivo deste artigo é apresentar uma análise das palavras utilizadas em notícias que mencionem indígenas no Brasil e no Paraguai. Ambos os países se caracterizam por contar com uma expressiva população indígena; como diferença destacável, o Paraguai, país oficialmente bilíngue, conta com a presença constante do idioma guarani. Já o Brasil, embora possua alguns municípios bilíngues e seja, também, um país aparentemente multilíngue e multicultural, tem, em sua Constituição Nacional, a indicação de apenas uma língua oficial, a língua portuguesa. Tendo em vista esses pontos de convergência e divergência entre ambos os países, almeja-se apresentar as principais diferenças e semelhanças na forma como os povos nativos são representados nas notícias publicadas, a fim de analisar como a presença ou ausência de um segundo idioma, pré-colonial, conforma o modo de ver e falar sobre esse Outro (TODOROV, 1991).

A CHEGADA DOS EUROPEUS

A forma como os espanhóis e portugueses agiram ao chegar na América Latina, em relação aos diversos povos que aqui encontraram, indica, de acordo com Todorov (1991), a não observação das características comunitárias e individuais de tais povos. O autor aponta alguns dos escritos de Colombo em suas cartas, que podem confirmar essa suspeita:

As pessoas eram como os outros já mencionados, de mesmas condições, igualmente desnudos e da mesma estatura (17.10.1492). Vieram muitas dessas pessoas, semelhantes aos outros das outras ilhas, igualmente pintados (22.10.1492). Essa gente [...] é de mesma qualidade e mesmos costumes dos outros encontrados (1.11.1492). Eles são gente como os outros que encontrei – disse o almirante –, com a mesma crença (3.12.1492)¹ (TODOROV, 1991, p. 45)².

O fato de não perceber a singularidade desse Outro também está explícito na forma como denomina esses povos. Por um equívoco sociogeográfico (D'ANGELIS, 2017), Colombo pensa estar na Índia, lugar em que almejaria chegar por novos caminhos marítimos. Como prova do desconhecimento dos povos de lá, assim como dos daqui, e de que não viam os indígenas, de maneira extensiva, chamou-os de *índios*, independentemente de suas diferenças, seja em pinturas corporais, seja em idiomas e crenças (D'ANGELIS, 2017).

Colombo, nesse primeiro momento, desconsiderou totalmente a existência de características que distinguissem as comunidades encontradas. Eles eram todos iguais e por, aparentemente, não verem naqueles estrangeiros inimigos potenciais, observando as riquezas naturais de suas terras, foram vistos como ingênuos, comparáveis a animais ou, ainda, facilmente colocáveis na situação de escravizados.

Ao chamar os aborígenes americanos de *índios*, Colombo demonstra uma tentativa de simplificar a denominação desses povos, ignorando suas diferenças (D'ANGELIS,

¹ As citações em inglês e espanhol mencionadas neste artigo foram traduzidas pela autora e as originais se encontram em notas de rodapé.

² “La gente toda era una con los otros ya dichos, de las mismas condiciones, y así desnudos y de la misma estatura” (17.10.1492). Vinieron muchos de esta gente, semejantes a los otros de las otras islas, así desnudos y así pintados (22.10.1492). Esta gente [...] es de la misma calidad y costumbre de los otros hablados (1.11.1492). Ellos son gente como los otros que he hallado – dice el Almirante –, y de la misma creencia (3.12.1492)”.

2017). A conotação dessa palavra, no entanto, acabou reduzindo as comunidades a algo muito menor do que eram, deixando transparecer, naqueles que recém-chegavam, o desconhecimento em relação aos habitantes originários da Índia e em relação às comunidades aqui encontradas. Essa análise de D'Angelis (2017) permite presumir que, em certa medida, a palavra que usamos para denominar determinado sujeito ou objeto pode interferir na forma como o enxergamos. Embora esse efeito também possa se dar na direção inversa, ou seja: a forma como enxergamos o sujeito ou objeto pode determinar a escolha de sua denominação.

Para analisar esse encontro entre diferentes visões de mundo, tem-se, em um dos pontos de partida possíveis para a Linguística, a Hipótese de Sapir-Whorf, também conhecida como Princípio do Relativismo Linguístico.

AS CONTRIBUIÇÕES DE HUMBOLDT E A HIPÓTESE DE SAPIR-WHORF

Sobre esse tema, Humboldt realizou uma importante contribuição para a Linguística ao trazer à luz a noção de que a linguagem revela a visão de mundo de seu falante: “[...] é apenas na língua que se mostra por completo o caráter, e é nela também, enquanto veículo geral de entendimento do povo, que sucumbem as individualidades particulares, em prol da maior visibilidade do geral” (HEIDERMANN; WEININGER, 2006, p. 3).

O entendimento de que a linguagem não é um simples meio transmissor do pensamento, mas que forma parte da sua constituição, contribuiu para a elaboração de diversos pressupostos que, um pouco mais tarde, seriam importantes fatores para as ideias iniciais do que conhecemos como Princípio do Relativismo Linguístico; além de contribuir e fundamentar os estudos etnolinguísticos. São hipóteses sobre os efeitos que o idioma materno produz no falante e na sua forma de observar acontecimentos no mundo; entre essas, cita-se a Hipótese de Sapir-Whorf.

Os seres humanos não vivem só no mundo objetivo... mas estão muito à mercê da linguagem particular que se tornou o meio de expressão de sua sociedade. Os mundos nos quais diferentes sociedades vivem são mundos distintos, não apenas o mesmo mundo com *rótulos* diferentes

(SAPIR, 1963 *apud* WERNER, 1997, p. 77)³.

A partir das constatações de Sapir e da análise da língua dos indígenas Hopi, Whorf concluiu que um idioma funciona como lentes que modelam a forma como seus falantes veem e pensam o mundo. A hipótese de Sapir-Whorf tem duas versões. A versão considerada *forte*, também chamada de determinismo linguístico, é a menos aceita por linguistas em geral, porque indica que a língua determina a forma de pensar do falante; seria dificilmente aplicável aos bilíngues, por exemplo (MOLINA MARTÍNEZ, 2006). Já a versão *fraca*, também chamada de relativismo linguístico, indica que a linguagem não determina, mas exerce uma forte influência sobre a forma de pensar (HATIM; MASON, 1995 [1990] *apud* MOLINA MARTÍNEZ, 2006).

Por ser menos radical, a versão *fraca* da hipótese se tornou mais plausível. Alguns estudiosos têm se empenhado, recentemente, em pesquisar tais efeitos em diferentes idiomas, a fim de verificar o nível de influência perceptível entre linguagem e pensamento. Além disso, os pesquisadores buscam entender se é a forma como pensamos que influencia nossa maneira de falar ou se nossa fala influencia nosso pensamento. Até o momento, percebe-se que ambas as direções devem ser consideradas.

RESULTADOS EMPÍRICOS E PESQUISAS MAIS RECENTES

Em publicação mais recente, Boroditsky (2013) reuniu alguns resultados de investigações que têm buscado comprovar como os diferentes idiomas afetam as percepções de mundo. Um deles demonstra as distintas representações físicas de tempo e revela que falantes do inglês, por exemplo, quando pensam no futuro, estão mais inclinados a fazer um movimento que aponte para sua frente, assim como pensam no passado como algo que está para trás (MILES, 2010 *apud* BORODITSKY, 2013). Já os falantes do Aimará, na Cordilheira dos Andes, tendem a pensar o futuro como algo desconhecido, sendo assim, indicam-no com sinais que apontem para trás, como aquilo que não se pode ver, e o passado como algo que esteve visível aos seus olhos, portanto, à frente (NÚÑEZ, 2006 *apud* BORODITSKY, 2013).

³ “Human beings do not live in the objective world alone... but are very much at the mercy of the particular language which has become the medium of expression for their society. The worlds in which different societies live are distinct worlds, not merely the same world with different *labels* attached”.

Em outro estudo, foi possível verificar que a linguagem pode apontar o nível de simpatia ou empatia por alguém ou por um grupo étnico. Ogunnaike *et al.* (2010 *apud* BORODITSKY, 2013) e Danziger *et al.* (2010 *apud* BORODITSKY, 2013) entrevistaram falantes bilíngues árabe-francês em Marrocos, espanhol-inglês nos Estados Unidos e árabe-hebraico em Israel. Em cada um dos pares de idiomas, foram testadas as tendências implícitas dos participantes.

Na primeira fase, solicitava-se às pessoas que pressionassem a tecla *M* cada vez que aparecesse na tela um nome hebreu, como *Yair*, ou uma característica positiva, como *bom* ou *forte*. E que pressionassem a tecla *X* sempre que vissem na tela um nome de origem árabe, como *Ahmed*, ou uma característica negativa, como *mesquinho* ou *fraco*. Na segunda fase, a paridade era revertida e os nomes judaicos partilhavam da mesma tecla que aspectos negativos, enquanto os nomes árabes compartilhavam tecla com aspectos positivos. Foi medida a velocidade de resposta nas duas condições; método que tem sido amplamente utilizado para medir respostas involuntárias ou automáticas. Como resultado, constatou-se que “os bilíngues em árabe e hebraico mostraram atitudes implícitas mais positivas em relação aos judeus quando testados em hebraico, do que quando testados em árabe” (BORODITSKY, 2013, p. 19).

Em aparente contraposição à noção de que linguagem e pensamento andam juntos, Pinker associa a linguagem a uma qualidade inata do ser humano:

[...] as pessoas sabem falar mais ou menos da mesma maneira que as aranhas sabem tecer teias. A capacidade de tecer teias não foi inventada por alguma aranha genial não reconhecida e não depende de receber a educação adequada ou de ter aptidão para arquitetura ou negócios imobiliários. As aranhas tecem teias porque têm cérebro de aranha, o que as impele a tecer e lhes dá competência para fazê-lo com sucesso (PINKER, 2002, p. 9).

Pinker (2002) defende que o desenvolvimento da linguagem humana independe da vivência em sociedade. Além disso, faz uma análise comparativa entre o conceito de *Novilíngua*, da obra literária distópica de Orwell (1949), intitulada *1984*, e o relativismo linguístico, acusando a impossibilidade de comprovar a existência de ambos os conceitos, ainda que eles não sejam comparáveis (CASASANTO, 2008).

A ideia de que “o pensamento é o mesmo que linguagem” (como sugeriu Orwell) não deve ser confundida com a ideia de linguagem temporal e pensamento, “diferenças entre idiomas, causam diferenças nos pensamentos de seus falantes” (como sugeriu Whorf). Orwell e Whorf levantaram duas questões distintas: pensamos na linguagem? e, a linguagem molda o pensamento? É possível (e eu argumentarei) que a linguagem possa moldar a maneira como as pessoas pensam, mesmo que o façam sem pensar na linguagem (CASASANTO, 2008, p. 64-65)⁴.

A proposta de Pinker, originalmente publicada em 1994 e apelidada, pelo próprio autor, de *obituário da hipótese Whorfiana* (PINKER, 2007 *apud* CASASANTO, 2008), contém erros de lógica, na análise de Casasanto (2008). De acordo com o autor, os argumentos utilizados por Pinker para a comparação e anulação conjunta da ideia contida na ficção de Orwell e na teoria de Whorf são ineficientes. Na criação fictícia da *Novilíngua* (ORWELL, 1949 *apud* CASASANTO, 2008), os falantes seriam educados para pensar na sociedade a partir do idioma – no qual algumas palavras não existiriam para que seus falantes não pensassem no que elas representam, como, por exemplo, a palavra *livre* usada em contexto de liberdade intelectual.

De acordo com Casasanto (2008, p. 66), “Pinker adota uma abordagem indiscriminada para creditar a equação Orwelliana da linguagem com o pensamento, começando com apelos à intuição”⁵ e força uma aproximação de situações nem sempre cotidianas, tal como associar a nossa dificuldade de expressar em palavras algo complexo que sentimos, à suposição de que há diferença entre o que dizemos e o que queremos dizer.

A ideia Orwelliana de que as pessoas pensam (principal ou inteiramente) no meio de linguagem natural e que, portanto, essa linguagem pode ser equiparada ao pensamento, não é suportada empiricamente e, além disso, é problemática, em princípio, dado o que se sabe sobre linguagem e pensamento. Por outro lado, a ideia Whorfiana de que diferenças linguísticas podem fazer com que falantes de diferentes idiomas pensem diferentemente não enfrenta desafios a princípio. Argumentos padrões, adequadamente interpretados, não soam um sinal de morte para a

⁴ “The idea that ‘thought is the same as language’ (as Orwell suggested) should not be conflated with the idea that ‘differences among languages cause differences in the thoughts of their speakers’ (as Whorf suggested). Orwell and Whorf raised two distinct questions: Do we think in language? and Does language shape thought? It is possible (and, I will argue, true) that language can shape the way people think even if they do not think in language”.

⁵ “Pinker takes a shotgun approach to discrediting the Orwellian equation of language with thought, starting with appeals to intuition”.

hipótese Whorfiana, mas um apelo a pesquisas mais rigorosas (CASASANTO, 2008, p. 66)⁶.

Além das questões já expostas, acredita-se que as propostas de Pinker, embora válidas para o avanço dos estudos da Linguística Cognitiva, não condizem com a análise que se propõe neste trabalho. A razão para que se opte pela hipótese do Relativismo Linguístico está na necessidade de discutir não a origem da linguagem e seus efeitos, mas sim a forma como a linguagem influencia na vida em sociedade e o quão representativa ela pode ser do modo como os indivíduos dessa sociedade enxergam o mundo que os cerca.

Uma das possibilidades de analisar essa visão compartilhada por determinada população, em um espaço de tempo específico, encontra-se em textos escritos⁷ no período que se busca investigar. Alguns exemplos de materiais úteis para esse tipo de pesquisa são os textos literários e os textos informativos, como os diferentes gêneros oferecidos à população pela imprensa, em revistas e jornais.

O GÊNERO NOTÍCIA COMO INSTRUMENTO DE ANÁLISE

Entre esses diferentes gêneros publicados em jornais e revistas, há um especificamente comprometido com o relato informativo de fatos recentes ou históricos: a notícia. Vivaldi (1993 apud PERUZZO, 2007) recomenda que a notícia apresente relatos de diferentes fontes sobre um mesmo tema, em linguagem culta e acessível, que permita manter sua objetividade; esta seria sua principal função sociocomunicativa:

Delimitamos como função sociocomunicativa da notícia: estabelecer a comunicação entre os membros da comunidade discursiva jornalística e leitores de jornais e revistas, através da divulgação de fatos e acontecimentos novos ou mais remotos (tanto no âmbito regional quanto nacional e mundial), informando a população sobre algo ou alguém; refletindo, assim, a ação social dos jornalistas que se refere ao

⁶ “The Orwellian idea that people think (mostly or entirely) in the medium of natural language, and therefore that language can be equated with thought, is unsupported empirically and is also problematic in principle, given what is known about language and about thought. By contrast, the Whorfian idea that linguistic differences can cause speakers of different languages to think differently faces no such in-principle challenges. When standard arguments are properly interpreted, they do not sound a death knell for the Whorfian hypothesis, but rather a call for more rigorous research”.

⁷ Além do material em texto, pode-se usar vídeos e áudios, mediante transcrição.

compromisso ético e profissional de transmitir informações (SILVA, 2011, p. 6).

De acordo com o *Manual de Redação da Folha de São Paulo*⁸, amplamente utilizado em redações jornalísticas, a notícia é o “puro registro dos fatos sem opinião. A exatidão é o elemento-chave da notícia [...]. Por essa característica de gênero mais voltado à informação, sem a pretensão de opinar sobre o assunto tratado, assume-se que a notícia seja o texto menos subjetivo do jornal e, por isso, possa ser analisada exclusivamente pelo seu valor informativo⁹. Embora se saiba que todo jornal tem um editorial, que define as políticas da empresa, parte-se do princípio de que a notícia aponta para o compromisso ético da neutralidade.

Por essa razão, além do fato de sua maior facilidade de acesso e compilação e da sua capacidade de alcance a um considerável número de leitores, optou-se por analisar em notícias de periódicos *on-line* a forma como os indígenas estão representados.

Para definir quais dos vários portais jornalísticos contemplariam os *corpora* de análise, com um método que fosse aplicável a ambos os países, Paraguai e Brasil, optou-se por recorrer à ferramenta *Alexa*, disponibilizada pela Companhia Amazon. Essa ferramenta permite verificar quais são os *sites* mais acessados no mundo inteiro e classificá-los por país. Desse modo, foi possível determinar que as notícias que deveriam compor os *corpora* brasileiro e paraguaio, nos idiomas português-BR e espanhol, respectivamente, seriam compiladas a partir dos portais informativos UOL e G1, do Brasil, e ABC Color e Última Hora, do Paraguai.

POR QUE PREOCUPAR-SE COM OS POVOS INDÍGENAS?

A chegada dos europeus à América acarretou mudanças radicais às rotinas e culturas dos povos originários americanos, mudanças provocadas principalmente pelas diferentes cosmovisões dos envolvidos, como visto, anteriormente, em Todorov (1991). Para ilustrar a ameaça real vivida por muitos idiomas, o *Atlas Sociolinguístico da UNICEF* aponta que “no Brasil, o país com a maior diversidade do continente, junto aos Estados

⁸ Disponível em: <https://www1.folha.uol.com.br/folha/circulo/manual_como_consultar.htm>. Acesso em: 12 ago. 2020.

⁹ Cabe lembrar que este trabalho justamente pretende provar que tudo que um indivíduo diz/escreve está

Unidos, as 180 [línguas] de hoje são as sobreviventes das 1.200 de cinco séculos atrás” (UNICEF, 2009, p. 241)¹⁰.

É possível verificar no relatório *Os povos indígenas na América Latina*, publicado pela ONU (2015), que as pequenas conquistas e os direitos alcançados por esses povos nos últimos 30 anos se devem, em muito, pela força de luta e resistência apresentada por eles próprios. A ONU (2015) apresenta dados comparativos, dos censos de 2000 e 2010 do Centro Latino-Americano e Caribenho de Demografia, que indicam um crescimento populacional de indígenas na América Latina. Conforme o relatório, o censo da Comissão Econômica para a América Latina e o Caribe (CEPAL), de 2007, estimava que, em 2000, existiam 30 milhões de indígenas em território latino. O censo de 2010, mais recente, atualizou esse número para 45 milhões; um aumento de 49,3% em 10 anos.

Esse aumento, no entanto, não representa o fortalecimento das culturas, tampouco a manutenção dos diferentes idiomas existentes. O próprio relatório apresenta a preocupação mencionada:

[...] a maioria dos países da região incluiu perguntas a esse respeito junto com as de autoidentificação. Alguns resultados permitem constatar que em muitos povos a preservação das línguas indígenas é muito intensa dentro dos territórios, mas fora deles uma pequena proporção de jovens a mantém. Em outros casos, a perda do idioma é um fato grave, tanto dentro como fora dos espaços territoriais próprios, e alguns povos indígenas estão experimentando uma tendência à rápida extinção de suas línguas (ONU, 2015, p. 98).

Embora não somem os mais altos números de pessoas que se identificam como indígena, Brasil e Paraguai chamam atenção por duas características destacáveis. De um lado, está o Brasil, país com apenas um idioma oficial, segundo sua Constituição, e com o maior número em variedade de línguas indígenas da América Latina – 186 idiomas diferentes (UNICEF, 2009). Do outro lado, está o Paraguai com sua característica tão peculiar de ser o único país no mundo, segundo as informações que se tem até o momento, em que uma maioria de pessoas não indígenas se comunica pelo uso de um idioma de origem indígena, o guarani (ONU, 2019). Além disso, considera-se, para os fins

sustentado por uma visão de mundo própria, portanto, é subjetivo.

¹⁰ “En Brasil, el país con la mayor diversidad del continente, junto con Estados Unidos, las 180 [lenguas] de hoy son las sobrevivientes de unas 1.200 hace cinco siglos”.

comparativos desta análise, o fato de serem países vizinhos e compartilharem fronteira no principal espaço ocupado por indígenas da etnia guarani.

A partir da noção de que as palavras usadas em determinado idioma podem refletir parte do pensamento de seus falantes, conforme apresentado e defendido por Humboldt, importa observar como esses povos estão representados nas notícias. E, tão significativo quanto observar as palavras usadas nas notícias brasileiras, é poder compará-las com as palavras usadas em um país vizinho que, tal como o Paraguai, oferece condições similares de vida a esses povos, por um lado, mas, por outro, apresenta diferenças marcantes, como o fato de ter, entre os idiomas oficiais, uma língua de origem indígena.

LINGUÍSTICA DE *CORPUS*

Para alcançar os objetivos deste trabalho, optou-se por utilizar a Linguística de *Corpus* como metodologia.

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 (SARDINHA, 2000, p. 325).

Cabe ressaltar neste ponto que o presente trabalho é fruto de uma pesquisa em andamento e que os resultados formam parte de um estudo inicial, em que se buscou testar a metodologia e conferir sua aplicabilidade, em consonância com a teoria, neste modelo de análise empírica. Tendo em vista que os resultados foram considerados positivos, ou seja, aplicáveis, atualmente os *corpora* estão sendo ampliados, a fim de possibilitar a coleta de um maior número de informações linguísticas dentro do mesmo tema. Este estudo inicial está composto por notícias brasileiras e paraguaias, publicadas no período compreendido entre janeiro de 2016 e dezembro de 2017.

Dentro do período indicado, foi possível compilar um total de 400 notícias. Destas, 193 são brasileiras e 207 são paraguaias. O *corpus* em língua portuguesa soma um total de 48.740 *tokens* (i.e., total de palavras, incluindo as repetições) e 9.659 *types* (i.e., cada item ou palavra, descartadas as repetições). Já o *corpus* de língua espanhola tem um total de

33.465 *tokens* e 9.122 *types*. Ainda que se trate de um *corpus* bem pequeno, acredita-se que sirva aos propósitos deste breve estudo, por se tratar de uma compilação inicial e conter uma variação de fontes informativas que possibilite a análise da representação desses povos nas notícias dos portais mais acessados pela população em ambos os países.

Sendo assim, pode-se dizer que, durante a compilação dos *corpora*, foi observada uma lista de critérios, também citados por Sardinha (2000) como pré-requisitos de um *corpus* destinado a um estudo computadorizado. Tal lista está composta por: origem ou autenticidade dos dados retirados de textos escritos em linguagem natural; propósito ou finalidade do estudo a que se destina a compilação; composição, que diz respeito ao critério de escolha do conteúdo a ser compilado para análise; formatação, para que os textos sejam legíveis pelo *software*; representatividade, que, no caso específico desta análise, foi encontrada em diferentes portais de notícias de cada idioma; além da extensão ou número de textos compilados, para que seu conjunto dê origem a um *corpus* minimamente representativo.

Para a coleta de notícias, utilizou-se a barra de pesquisa do Google, onde foram digitadas as palavras entre aspas “indígena OU indígenas OU índio OU índios”; no botão *ferramentas*, foi selecionada a opção *período personalizado*, onde foi colocado o ano que se desejava pesquisar, como 01/01/2016 a 31/12/2016; buscou-se por *todos os resultados*. No botão *preferências* foi feita a buscar por *ferramentas avançadas*, onde foi possível digitar o domínio ou o *site* dos portais, mencionados anteriormente, de modo individual. Feito isso, a guia *notícias* foi selecionada, para que fossem filtrados os textos desse gênero.

Os arquivos foram organizados em pastas por país de origem e ano de publicação. Após a limpeza do texto, que consiste na retirada de informações não textuais ou irrelevantes, como imagens, data e nome do autor, cada arquivo recebeu um nome com a sigla do país e a data da publicação, AAAA/MM/DD. Assim, o nome de um arquivo com notícia publicada no Brasil no dia vinte de maio de 2016, por exemplo, é *BR20160520*. Nos casos em que se encontrou mais de uma notícia no mesmo dia, a data foi repetida e entre parênteses colocou-se o número da ordem em que foi encontrada; por exemplo *BR20160520(2)*, *BR20160520(3)*. Após esses passos, os arquivos foram convertidos para a extensão .txt, para que fosse possível realizar o *upload* deles no

programa AntConc. O AntConc “é um concordanciador utilizado para listar as ocorrências de uma determinada palavra ou frase em uma quantidade definida de contextos” (KADER; RICHTER, 2013, p. 13). É uma ferramenta que facilita a extração de informação linguística de textos já digitalizados.

Com o AntConc, foi possível gerar a lista de palavras mais frequentes em ordem decrescente, ou seja, das que mais aparecem até as que contam com apenas uma aparição nos textos. Além disso, foi possível confirmar que o *corpus* está composto por textos com um tema específico, já que a primeira palavra de cada lista é indígenas, seguida pelo seu singular indígena e por outras palavras cujas presenças colaboram com a elucidação da temática dos textos, conforme demonstrará a lista apresentada no capítulo a seguir, no qual será realizada a análise das palavras mais encontradas. Por se tratar de uma análise comparativa, o mesmo processo foi realizado separadamente para ambos os *corpora*, a fim de gerar duas listas de palavras mais frequentes, uma em português, outra em espanhol.

ANÁLISE

De acordo com Humboldt (*apud* HEIDERMANN; WEININGER, 2006), quando se observa uma comunidade, geralmente, as individualidades de cada falante sucumbem e prevalece o modo geral do grupo, ou seja, ao emitir uma mensagem, o emissor subentende que aquela visão de mundo, compartilhada, permitirá a total compreensão da sua mensagem por parte dos receptores. Isso também permite identificar o que, possivelmente, não seria compreendido por esses receptores e, portanto, necessita de maiores explicações.

Considerando essa percepção na linguagem jornalística, em que o texto precisa ser escrito de forma clara a fim de que seja entendido por leitores muito ou pouco letRADOS, mais ou menos eruditos, especialistas ou leigos (VIVALDI, 1993 *apud* PERUZZO, 2007), percebe-se que o jornalista necessita seguir um estilo específico de escrita.

Desse modo, cabe pensar que as palavras constantes em uma notícia de difusão nacional representam, em certa medida, a forma como as pessoas daquele país percebem determinado tema. Isso pode ser presumido por se tratar de um gênero cujo objetivo é informar, logo, entende-se que a notícia precisa estar composta por uma maioria de palavras que sejam de uso comum da população.

Partindo dessa noção, observa-se abaixo as 10 palavras mais utilizadas nas notícias

de portais mais acessados no Brasil e no Paraguai que informam sobre os povos indígenas no período analisado. O número de repetições está especificado entre parênteses e segue a ordem decrescente. Esta tabela apresenta o número de *types*, ou seja, estão somadas as variantes plural e singular, como *indígena* + *indígenas*.

Tabela 1: *Types* - ordem decrescente

	Brasil - 9.659 <i>types</i>	Paraguai - 9.122 <i>types</i>
1	<i>indígena(s)</i> (1.424)	<i>indígena(s)</i> (781)
2	<i>índio(s)</i> (806)	<i>comunidad(es)</i> (418)
3	<i>terra(s)</i> (434)	<i>tierra(s)</i> (198)
4	<i>funai</i> (324)	<i>pueblo(s)</i> (186)
5	<i>aldeia(s)</i> (312)	<i>nativo(s)</i> (160)
6	<i>federal/federais</i> (307)	<i>nacional(es)</i> (139)
7	<i>área(s)</i> (287)	<i>guaraní</i> (135)
8	<i>polícia(s)</i> (254)	<i>zona(s)</i> (124)
9	<i>região/regiões</i> (225)	<i>indi</i> (106)
10	<i>estado(s)</i> (213)	<i>lugar</i> (104)

Fonte: Criada pela autora a partir dos resultados revelados pelo AntConc.

Na lista de *types* das notícias do Brasil, é possível identificar que *índio(s)* aparece como sinônimo de *indígena(s)*, tal como nas frases “Índios de MT com carência nutricional devem receber 2,1 mil cestas básicas” (BR20161229); “Confronto entre índios e polícia no Congresso tem mais de 50 flechas apreendidas” (BR20170425(2)), o que justifica sua alta frequência.

Já no Paraguai, a palavra *indio* é usada para referir-se a algo ou alguém proveniente da Índia, tal como no exemplo retirado do portal Última Hora em um texto de fevereiro de 2016: “Gobierno regional indio bloquea Internet para evitar que se copie en exámenes”. Por isso, essa palavra não aparece no *corpus* de língua espanhola. Observa-se que, para substituir a palavra *indígena* ou evitar sua repetição, as notícias paraguaias

contêm as palavras *nativos* ou o nome da etnia a qual se referem, como *aché* e *guaraní*: “Ernesto Vera, de la comunidad **Ava Guaraní** [...]” (PY20160414); “Cinco nativos pertenecientes al pueblo **Avá Guaraní** [...]” (PY20160416); “Los **aché** denuncian que el diputado José “Pakova” Ledesma respalda a usurpadores de las tierras privadas que pertenecen a su comunidad, en Canindeyú” (PY20160310).

A frequência da palavra *índio* em notícias brasileiras, usada como sinônimo de *indígena*, expõe uma questão que desperta curiosidade e tem sido discutida atualmente. Antes mesmo de tratar dessa existência, cabe lembrar que ela tem certa visibilidade no Brasil desde a Constituição de 1988, em que aparece como sinônimo de *indígenas*, tal como no capítulo VIII intitulado *Dos índios* (BRASIL, 1988). Outro exemplo é a FUNAI, que ostenta a palavra em seu nome: Fundação Nacional do Índio; essa mesma instituição antes se chamava Serviço de Proteção aos Índios (SPI).

A questão já apresentada por Todorov (1991) de que o nome *índio* teria sido dado pelos espanhóis, pela crença de que haviam chegado às Índias por um novo percurso marítimo, é aprofundada por D’Angelis (2017), que acredita que os portugueses adotaram o mesmo nome a título de empréstimo linguístico, já que, em português, a nacionalidade de quem nasce na Índia é *indiano(a)*.

Por isso, de acordo com o autor, este termo não seria o mais adequado a se usar como sinônimo de *indígena*, que, do latim, significa *originário do lugar* [*inde + gena*] (D’ANGELIS, 2017). Aparentemente, poucos dicionários mencionam esse equívoco etnográfico; foi possível encontrá-lo apenas em Beaurepaire-Rohan (1889, s. v. *índio*): “nome que se aplica normalmente aos aborígenes da América, o que os confunde com os naturais das Índias Orientais. É um erro etnográfico que se cometeu desde a descoberta da América, pela crença em que ficara Colombo de ter chegado à Índia [...]”.

Quando se refere à variedade cultural, a palavra *pueblo(s)* costuma preceder *indígena(s)*, como em: “El pintor subrayó que el reconocimiento de las actividades culturales y costumbres indígenas es una parte del derecho de esos pueblos a su soberanía y a ser dueños de su tierra” (PY20160803); “Pueblos indígenas se unen para formar partido político” (PY20170730(2)); “[...] se debe principalmente a la victimización de los pueblos originarios en sus tierras ancestrales y explica el desalojo violento que sufrieron [...]” (PY20171212).

Ao mencionar o lugar onde vivem esses povos, as notícias brasileiras utilizam a palavra *aldeia(s)*: “[...] durante visitas à aldeia, a manifestação cultural passou a ser reanalisada [...]” (BR20170428); “[...] emitiu nota esclarecendo que negocia com os líderes indígenas com o intuito de liberar as máquinas e atender às demandas da aldeia” (BR20160212). Enquanto as paraguaias optam por usar *comunidad(es)*: “Caciques de otras comunidades también llevaron sus pedidos [...]” (PY20160130); “[...] comunidad indígena afectada por fumigación de soja les en Curuguaty” (PY20170808(2)).

Essa diferença atrai a atenção especialmente pelo fato de existir a palavra *comunidadade* também em língua portuguesa. O que se observa, porém, é que na prática jornalística deste idioma tal palavra é mais usada para denominar bairros populosos e/ou periféricos, tal como se pode verificar no seguinte trecho do portal *G1*¹¹, datado de setembro de 2017: “Equipes do Baep realizam operação em comunidades de Santos e São Vicente”.

O grande número de ocorrências da palavra *policía(s)* chama atenção no *corpus* de língua portuguesa. Alguns exemplos disso são: “A Polícia Federal (PF) investiga a morte do indígena” (BR20160708); “De acordo com a Polícia Civil, três índios, da etnia Rikbaktsa, estavam em uma motocicleta que foi atingida pelo caminhão” (BR20170614); “A Polícia Civil disse que o funcionário da Funai e os quatro indígenas foram presos com três armas, localizadas dentro de um carro” (BR20170511).

Já nas notícias paraguaias, a palavra *policía(s)* nem sequer aparece entre as 50 palavras mais usadas, embora apareçam algumas descrições de situações violentas: “El joven está sindicado por la Policía como el autor del crimen de dos miembros de la comunidad indígena” (PY20160703); “El supuesto autor iba detrás de ella, cuando vio a la Policía, el hombre también indígena, huyó” (PY20171205); “Además, la policía indicó que entre el grupo de atacantes se encontraba el anterior cacique de la comunidad [...]” (PY20170508).

Em ambos os países é frequente o uso da palavra *terra(s) / tierra(s)*; pode-se perceber, de acordo com os exemplos, que essa aparição está relacionada a uma

¹¹ O exemplo que segue não consta no *corpus* de análise; é resultado de uma pesquisa à parte, realizada a fim de se compreender o porquê da ausência dessa palavra nas notícias brasileiras sobre indígenas, e verificar com qual temática ela estaria mais associada. Disponível em: <<https://g1.globo.com/sp/santos-regiao/noticia/dois-homens-sao-mortos-em-confronto-com-policiais-militares-em-comunidades-no-litoral-de-sp.ghtml>>. Acesso em: 12 ago. 2020.

problemática comum entre esses povos, a da demarcação de terras indígenas: “Varios son los pueblos indígenas que se enfrentan a problemas, como la comunidad Yakye Axa que no podía asentarse en sus tierras [...]” (PY20160224); “Uno de los caciques, Eliseo Rivas, comentó que decidieron ocupar las tierras porque alguna vez pertenecieron a sus ancestros, por lo que, segúnél, les ‘corresponde’” (PY20171009); “Com a indicação positiva, são necessárias duas portarias: a que delimita a terra, a cargo da Funai; e a que declara a terra, do Ministério da Justiça” (BR20170419(2)); “Índios guarani-kaiowá entraram na fazenda no último domingo - eles reivindicam a área como terra tradicional indígena” (BR20160615).

Tanto nas notícias brasileiras como nas paraguaias, nota-se a frequência de menção aos órgãos responsáveis pela manutenção dos direitos indígenas, que respondem juridicamente por esses povos. O Instituto Paraguayo del Indígena (INDI) e a Fundação Nacional do Índio (FUNAI) têm suas siglas entre as palavras mais mencionadas, principalmente por serem representantes legais e alvos de entrevistas em um grande número de casos, como indicam os exemplos: “De acordo com a Funai, a floresta é o habitat de quatro povos indígenas reconhecidos e ainda possui cerca de três grupos de índios isolados que também vivem na área verde” (BR20170217); “Ante ola de frio, INDI habilita albergue para indígenas” (PY20160427).

Também aparecem entre as mais citadas, as palavras *federal* e *nacional*, que estão relacionadas às instituições, aos órgãos e a documentos públicos, como Congresso, Polícia, Governo, Constituição etc. Além disso, a lista ainda apresenta o ambiente em que ocorre o fato noticiado: *área(s)*, *região(ões)*, *estado(s)*, *zona(s)* e *lugar*.

CONCLUSÕES

A partir das listas de *types* aqui apresentadas e dos exemplos citados, é possível perceber algumas semelhanças temáticas entre as notícias do Brasil e do Paraguai, como a frequência compartilhada da palavra *terra(s)*. As principais diferenças, no entanto, estão na escolha de palavras utilizadas para mencionar os povos indígenas, nas notícias, e o ambiente em que eles vivem.

A fim de buscar uma possível explicação para a existência dessas diferenças, pode-se retomar um dos pontos expostos pela Hipótese do Relativismo Linguístico, que se

confirma nos exemplos encontrados, de que falantes de diferentes idiomas observam o mundo de forma diferente. Isso também inclui as escolhas dos falantes por algumas palavras específicas, ainda que, algumas vezes, existam outras opções em seu léxico. Um exemplo disso está na escolha da imprensa brasileira de chamar de *aldeia* o que a imprensa paraguaia prefere chamar de *comunidad*, ainda que a palavra *comunidadade* também exista na língua portuguesa, conforme visto na análise.

Já em relação ao uso da palavra *índios* pode-se concluir, neste estudo empírico, que Todorov (1991) tinha razão quando apresentava a noção de que os europeus, ao descobrir o continente, não se dedicaram a descobrir quem eram as pessoas que já viviam nele. A manutenção dessa palavra em nosso idioma poderia indicar, de certo modo, uma manutenção desse desconhecimento até a atualidade, diferentemente do caso paraguaio, em que esse uso não aparece.

Assim, conclui-se que os resultados apresentam diferenças visíveis entre a representação dos indígenas nos dois países analisados. Como informado anteriormente, os resultados apresentados neste trabalho formam parte de uma pesquisa em andamento. Tem-se então, como próximo objetivo, explorar estes resultados de forma qualitativa, fazendo uso da prosódia semântica.

REFERÊNCIAS

- AMAZON COMPANY. *Alexa: the top 500 sites on the web*. Disponível em: <<https://www.alexa.com/topsites/countries>>. Acesso em: 12 ago. 2020.
- BEAUREPAIRE-ROHAN, H. *Dicionário de vocábulos brasileiros*. 2.ed. Bahia: Progresso Editora, 1956.
- BORODITSKY, L. *Como a linguagem modela o pensamento*. 2013. Disponível em: <https://issuu.com/ed_moderna/docs/aulaaberta14>. Acesso em: 12 ago. 2020.
- BRASIL. Constituição (1988). *Constituição da República Federativa do Brasil*. Disponível em: <<https://www2.camara.leg.br/legin/fed/consti/1988/constituicao-1988-5-outubro-1988-322142-publicacaooriginal-1-pl.html>>. Acesso em: 12 ago. 2020.
- CASASANTO, D. Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought. *Language Learning*. University of Michigan, 58: Suppl, dec. 2008, p. 63-79.

CUENCA, M. J.; HILFERTY. *Introducción a la lingüística cognitiva*. Barcelona: Ed. Ariel, 1999.

D'ANGELIS, W. R. *No Brasil ainda tem “índio”*. 2017. Disponível em: <<http://kamuri.org.br/kamuri/no-brasil-ainda-tem-indio/>>. Acesso em: 12 ago. 2020.

DORTIER, J. *Dicionário de Ciências Humanas*. São Paulo: Martins Fontes, 2010.

HEIDERMANN, W; WEININGER, J. M. (orgs). *Wilhelm von Humboldt. Linguagem, Literatura, Bildung*. Florianópolis: UFSC, 2006. (Edição Bilíngue).

KADER, C. C. C.; RICHTER, M. G. Linguística de *Corpus*: Possibilidades e avanços. *Revista Instrumento - Revista de Estudo e Pesquisa em Educação*, v.15, n.1, 2013. Disponível: <<https://periodicos.ufjf.br/index.php/revistainstrumento/article/view/18855>>. Acesso em: 12 ago. 2020.

MOLINA MARTÍNEZ, L. *El otoño del pingüino*: Análisis descriptivo de la traducción de los culturemas. *Publicaciones de la Universitat Jaume I*, 2006, p. 21-24. Disponível em: <<https://bit.ly/2yCZFOL>>. Acesso em: 12 ago. 2020.

ONU, Organização das Nações Unidas (CEPAL). *Os povos indígenas na América Latina: avanços na última década e desafios pendentes para a garantia de seus direitos*. Nações Unidas: Chile, 2015.

ONU, Notícias. *La lengua guaraní, orgullo de un país*. (21 feb. 2019). Disponível em: <<https://news.un.org/es/story/2019/02/1451281>>. Acesso em: 12 ago. 2020.

PERUZZO, M. S. *Como lidar com os neologismos no texto jornalístico*. Dissertação de Mestrado. UFRGS, 2007. Disponível em: <<https://lume.ufrgs.br/handle/10183/10910>>. Acesso em: 12 ago. 2020.

PINKER, S. *O instinto da linguagem*: Como a mente cria a linguagem. Tradução: Cláudia Berliner. São Paulo: Martins Fontes, 2002.

SARDINHA, T. B. *Linguística de Corpus*: Histórico e Problemática. *DELTA*, v.16, n.2, 2000, p.323-367. Disponível em: <https://www.researchgate.net/publication/26361085_Linguistica_de_Corpus_historico_e_problematica>. Acesso em: 12 ago. 2020.

SILVA, P. H. Os gêneros jornalísticos e a notícia. *Simpósio Internacional de Letras e Linguística*. Anais do SILEL, v.2, n.2, Uberlândia: EDUFU, 2011. Disponível em: <http://www.ileel.ufu.br/anaisdosilel/pt/edicao_volume_2_numero_2.php>. Acesso em: 12 ago. 2018.

TEILLIER, F. *et al.* De qué hablamos cuando hablamos de Etnolingüística: bases teórico-metodológicas para un trabajo con el Mapunzugun. *Revista de Lingüística Teórica y Aplicada*. Concepción (Chile), 54 (2), 2016, p.137-161 Disponível em:

<https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-48832016000200007>. Acesso em: 12 ago. 2020.

TODOROV, T. *La conquista de América: el problema del otro*. Naucalpan de Juárez: National print, 1991.

UNICEF, Fondo de las Naciones Unidas para la Infancia. *Atlas Sociolingüístico de pueblos indígenas en América Latina*. 2009. Disponível em: <https://www.unicef.org/tomo_1_atlas.pdf>. Acesso em: 12 ago. 2020.

WERNER, O. Sapir-Whorf Hypothesis. In: LAMARQUE, P. V. (Ed.) *Concise Encyclopedia of Philosophy of Language*. Oxford/New York/Tokyo: Pergamon, 1997, p.76-83. Disponível em: <https://www.academia.edu/6229219/Concise_Encyclopedia_of_Philosophy_of_Language>. Acesso em: 12 ago. 2020.

Recebido em: 01 jul. 2020.

Aceito em: 21 ago. 2020.