



II Seminário
— Viagens da Língua

LÍNGUA E TECNOLOGIAS

DOSSIÊ

 Museu
da Língua
Portuguesa

II Seminário
— Viagens da Língua

LÍNGUA E TECNOLOGIAS



Museu
da Língua
Portuguesa

SUMÁRIO

- 9 Apresentação**
Centro de Referência do
Museu da Língua Portuguesa
- 13 Conferência de abertura: “Língua como tecnologia de criação de mundos”**
Ana Rüsche
- 20 Viagens da língua:
a convergência da inteligência artificial com as línguas indígenas**
Nina da Hora
Raymundo Baniwa
- 25 Linguagem simples e complexa:
metacognição epilinguística para o letramento e cidadania plenos**
Marcus Maia
Gisele Rodrigues
- 42 A linguagem simples e as instituições culturais**
José de Souza Muniz Jr.
- 49 Mapeamento conceitual do domínio da informação tecnológica a partir de *corpus* linguístico**
Maria Carolina Coradini
Eder Cavalcanti Coimbra
Gabriela Wick-Pedro
Milena Ambrosio Telles
Isaac Souza de Miranda Junior

58 Tradução e tecnologia para as línguas do Brasil: o trabalho do grupo de tradução do par sateré-português

Adriano Clayton da Silva
Osinara da Paz Batista

66 CAPIVARA: como relacionar imagens e textos em português utilizando inteligência artificial

Jhessica Silva, Gabriel Santos, Diego Moreira, Alef Ferreira, Luiz Pereira, Pedro Bueno, Thiago Sousa, Helena Maia, Nádia da Silva, Esther Colombini, Helio Pedrini, Sandra Avila

74 Tecnologias para o registro de línguas de sinais: um panorama histórico

Luis dos Santos Miguel
Fernanda Henriques

79 O que as ferramentas de pedra podem nos contar sobre o surgimento da linguagem

Vitor Nóbrega

89 Gamificação de conteúdos e jogos digitais para o ensino e o aprendizado da língua portuguesa e da educação ambiental

Maria de Oliveira Rodrigues

96 Linguística computacional para professores da educação básica: perspectivas e contribuições

Fábio Rezende de Souza
Ana Rosa Frazão Paiva

105 A exposição virtual “A criança na língua: passo a passo”. Uma estratégia para a democratização do acesso ao conhecimento científico na área de aquisição da linguagem

Rosângela Nogarini Hilário
Alessandra Del Ré

112 Afrofuturismo e linguagem: um olhar sobre os usos da linguagem como ferramenta de construção de futuros dissidentes

Pétala Tainá de Oliveira de Souza
Isabella Raíssa Oliveira de Souza

119 O racismo algorítmico em funcionamento em textos gerados pelo ChatGPT

Gustavo Bonil da Silva
Simone Tiemi Hashiguti

129 Humanidades digitais: perspectivas para o estudo da linguagem

Rodrigo de Lima Lopes

135 Relações entre jogo e cultura na evolução da linguagem

Ernane Guimarães Neto

142 Aplicativos para o ensino-aprendizagem de línguas indígenas de povos do Brasil

Suellen Tobler

150 A metamorfose digital dos códigos lúdicos

Gilson Schwartz

160 Inteligência Artificial e linguagem

João Carlos Pereira da Silva



APRESENTAÇÃO

A língua e a tecnologia se atravessam em diversos momentos: nos inúmeros sistemas de escrita criados para gravar as ideias em suporte material, quando se desenha com um pedaço de carvão em uma pedra, ao se escrever uma carta, em duas latas ligadas por um barbante, nas conversas por meio de aplicativos...

Para muito além de aparatos eletrônicos, como o atual uso da palavra evoca, o saber tecnológico faz parte das mais diversas esferas da vida, e há muito tempo. Tecnologia pode ser entendida como um sistema – de técnicas ou de processos – por meio do qual uma sociedade satisfaz as necessidades e desejos de seus membros. Portanto, trata-se de um sistema que pode envolver equipamentos e programas, mas também pessoas, processos, organização e diversas finalidades de propósito.

Entendemos que língua, tecnologia, cognição e cultura são inseparáveis, e abordar essa relação também é falar da história e das dinâmicas sociais de seus falantes. Assim, do mesmo modo que o Museu trata a língua como parte crucial da natureza humana, também as interações tecnológicas fazem parte de nossa história.

O Museu da Língua Portuguesa reabriu ao público em 2021 fortalecendo sua proposta original de tratar da língua portuguesa como objeto cultural, lançando mão de um uso extenso de aparatos tecnológicos, com o intuito de dar conta do dinamismo e constante movimento inerente à língua. Entretanto, recorrer a ferramentas tecnológicas não é algo recente, e muitas vezes, assim como a linguagem, é fator de diferenciação da espécie humana.

Considerando a pertinência do debate sobre esses temas, o Centro de Referência do Museu da Língua Portuguesa promoveu de 6 a 8 de dezembro de 2023 a segunda edição do Seminário Viagens da Língua, dessa vez voltado a discutir as relações entre língua e as tecnologias, considerando desde as possibilidades técnicas até as dimensões sociais e culturais. O evento reuniu pesquisadores, docentes, artistas e escritores para compartilhar estudos científicos e experiências que contribuem para a reflexão e o aprofundamento desse tema que é fundamental ao Museu.

Em sua concepção, os debates foram organizados em duas frentes. A primeira, “Língua, pesquisa e tecnologias”, abarcou debates que contemplam a relação entre as pesquisas feitas sobre e por meio de língua com as diferentes tecnologias já criadas, bem como ferramentas atuais de documentação linguística e de preservação de outros acervos, explorando como a pesquisa sobre línguas se beneficia dos avanços tecnológicos. Se, por um lado, existe uma relação estreita entre linguagem, cognição e as ferramentas que os seres humanos usam há milênios, por outro, as ferramentas de registro e reprodução desenvolvidas nas últimas décadas trouxeram um salto para a pesquisa sobre as línguas. O conhecimento sobre a língua ajuda a construir dispositivos que contribuem para a preservação de línguas.

A segunda frente, “Língua, interação e tecnologias”, focou nos impactos (e nas contribuições) que os avanços tecnológicos causam no ensino e na aprendizagem. Ainda que a primeira e principal forma de comunicação humana seja a conversa face a face, vimos como as tecnologias podem reorganizar nossas práticas comunicativas, diminuindo ou criando barreiras à interação humana.

Reunimos neste Dossiê textos oriundos das frutíferas trocas realizadas nas mesas-redondas, rodas de conversa e cursos produzidos ao longo desses três dias. Abre a publicação a participação da escritora e pesquisadora Ana Rüsche na conferência de abertura “Língua como tecnologia de criação de mundos”. Além do mote da conferência, que Ana explorou principalmente por meio da ficção científica, a escritora também falou dos estereótipos de ciência e tecnologia no imaginário popular, criado na produção audiovisual no século XX, e trouxe a dimensão das tecnologias do Sul Global e dos conhecimentos ancestrais para esse universo. A conferência de abertura também contou com a cientista da computação e *hacker* antirracista Nina da Hora, que participa deste Dossiê com um texto em coautoria com Raymundo Baniwa sobre a convergência da Inteligência Artificial com as línguas indígenas.

O II Seminário Viagens da Língua: Língua e tecnologias reafirma o Museu da Língua Portuguesa, por meio de seu Centro de Referência, como ponto de encontro, um espaço de debate e reflexão de temas pertinentes à esfera dos estudos sobre línguas, sempre por meio de linguagens variadas e promovendo o diálogo entre públicos diversos.

Boa leitura!

CENTRO DE REFERÊNCIA DO
MUSEU DA LÍNGUA PORTUGUESA



CONFERÊNCIA DE ABERTURA: “LÍNGUA COMO TECNOLOGIA DE CRIAÇÃO DE MUNDOS”¹

ANA RÜSCHE

Quero começar agradecendo o convite do Centro de Referência do Museu da Língua Portuguesa, nas pessoas da Cecília e Camila, e também agradecer ao pessoal da produção, na pessoa do Antônio, à equipe de intérpretes e, claro, à Nina da Hora. Foi um prazer imenso escutá-la. É muito bom estar aqui com vocês.

Cheguei mais cedo para rever o museu, revisitá-lo, e fiquei muito emocionada circulando por aqui. Para começar esta nossa conversa, peço que vocês imaginem uma palavra de que gostam muito. Eu já tenho uma anotada no meu coração, mas vamos usá-la no final da minha fala, porque vamos conversar a respeito da língua, ou das línguas, como tecnologias. Meu primeiro convite a vocês, e que acho que tem muito a ver com o que a Nina nos trouxe, é pensar

[1] Este texto é a transcrição da palestra proferida por Ana Rüsche na Conferência de Abertura do II Seminário Viagens da Língua: Língua e tecnologias, em 6 dez. 2023.

um pouco em qual conceito de tecnologia a palavra “tecnologia” implica, principalmente no contexto em que a estamos utilizando, o contexto do Sul Global.

Quando pensamos em tecnologia, na palavra tecnologia – poderíamos inclusive falar de *techne*, mas hoje não vamos tão longe –, temos uma ideia de que ela se refere a um laboratório todo branco com luzes, pessoas de jaleco etc. Isso para nós é tecnologia porque o cinema, em especial o cinema de ficção científica produzido nas décadas de 1960 e 1970, reforçou essa visão. O que é o espaço? Como espécie, quase nunca fomos para o espaço, mas sabemos bem o que ele é, aquela coisa branca. Demorou muito para as naves espaciais perderem esse aspecto de higienização que Stanley Kubrick propôs em seus filmes.

Mas tecnologia é mais do que isso. A língua, por exemplo, é uma tecnologia. É muito importante que nos lembremos de que tecnologia nada mais é do que um saber humano colocado em uma aplicação prática, um saber que pode ser reproduzido, repassado a outras gerações e, assim, aprimorado.

Por exemplo, fazer um chá com determinada erva que nossas avós usavam em diferentes lugares é uma tecnologia. Pode até ser que essa erva não produza o efeito de acalmar uma pessoa específica, mas o gesto vai mais longe, e entendemos o que significa preparar um chá de erva para alguém, oferecer conforto emocional. Outro exemplo de tecnologia que, penso eu, não funciona muito para mim, mas que pode funcionar para outras pessoas, agora ou no futuro, é o ato de acender uma vela para algo ou alguém. Dependendo do que você acredita, da sua fé, esse gesto pode ou não produzir um efeito. Esses são exemplos de tecnologias espirituais, ancestrais, de diferentes tradições e que existem em diferentes culturas ao redor do planeta.

Dito isso, quando pensamos na palavra “tecnologia” no Sul Global, precisamos, antes de tudo, trazer esse conceito para um lugar que não é visto como produtor de tecnologia. Nós somos vistos como povos que consomem tecnologia pronta do Norte Global. É claro que existe, na palavra tecnologia, uma identificação

com a produção técnica, com uma produção científica, de foguetes, de novos componentes de computadores etc., mas, ao trazer essa palavra para nós, podemos trazer com ela outras coisas, como a língua, por exemplo. Com a tecnologia da língua, podemos construir outros tipos de sentidos e outras redes entre os seres humanos.

Para começarmos a tratar dos aspectos e capacidades das línguas, quero destacar que toda língua é uma bela tecnologia. Primeiro, porque é histórica – mas não vou me alongar nessa questão porque aqui no Museu vocês podem ver tudo isso de perto. Há muitas coisas que cada palavra nos traz. No início da minha fala, pedi que vocês pensassem em uma palavra. A minha, por exemplo, eu sei que vem do latim, mas não sei se foi no latim que ela surgiu. É algo que acho, mas não tenho certeza. E vocês, conhecem a raiz da palavra que imaginaram? Não importa se não conhecerem, mas é uma pergunta interessante de nos fazermos: quem falou essa palavra antes de nós? É muito bonito isso, pensar que palavras são uma tecnologia que atravessam os séculos.

Isso nos leva a outro aspecto da língua, que é sua capacidade de criar uma comunidade. Vou dar um exemplo bem comum: os adolescentes usam muitas gírias. Qual é a beleza das gírias? As gírias estabelecem comunidades de falantes, criando uma identificação entre esses falantes e excluindo outros. Isso permite que um adolescente diga: “Esta é minha turma”. Aliás, “turma” é uma palavra que está bem longe de ser uma gíria contemporânea. Mas quero comentar também um exemplo literário. Imagino que boa parte de vocês já tenha ouvido falar do livro *Laranja mecânica*, de Anthony Burgess, publicado em 1962. É um livro bastante violento, que foi adaptado para o cinema por Stanley Kubrick. Esse livro é interessante porque usa a língua inglesa no seu limite para emular a fala de um delinquente juvenil. Em português, o livro foi traduzido por Fábio Fernandes e foi publicado pela Editora Aleph, e tem essa questão muito interessante de levar a língua a um limite que só pode ser compreendido de fato por aqueles que fazem parte daquele grupo social.

Outra característica das línguas, e que talvez seja uma das mais relevantes de sinalizar, é que toda língua é adaptável. Por isso

é que a linguagem, e a língua especificamente, é tão poderosa: porque conseguimos fazer com que as palavras mudem de sentido de acordo com as situações. Vou dar mais exemplos literários para ilustrar. A palavra “Deus” aparece tanto em *Floresta, o nome do mundo* (1972), de Ursula Le Guin, quanto em *Parábola do semeador* (1983), de Octavia Butler. No livro de Le Guin, existe uma construção bastante longa para mostrar como um tradutor pode ser um deus, é um momento auge da novela, em que o tradutor aparece como uma pessoa que consegue transitar entre mundos. Já no livro de Octavia Butler, a protagonista diz: “Deus é mudança”. E por que ela constrói essa ideia de Deus como mudança? Justamente porque se trata de uma distopia em que há várias coisas terríveis acontecendo, e a capacidade de mudança e adaptabilidade é vista como algo muito positivo. Portanto, temos aqui duas formas muito diferentes de usar a palavra “Deus”.

Agora, gostaria de pensar com vocês mais precisamente na capacidade criativa das línguas. Há vários livros e obras artísticas que criam línguas imaginárias. É famoso, por exemplo, o élfico criado por J. R. R. Tolkien, autor de *O Senhor dos Anéis* e de *O Hobbit*, mas temos também o klingon de *Star Trek*, a língua do P, que algumas crianças aprendem quando são pequenas, e até mesmo o sonho do esperanto, como um idioma que poderia ser comum a todos os seres humanos. Nessa linha, um livro muito interessante é *Língua nativa*, publicado em 1984 por Suzete Elgin e traduzido recentemente no Brasil por Jana Bianchi, em edição da Aleph. Elgin, além de escritora, é linguista, e ela não só cria uma língua (como nos exemplos que citei ou como o votrak de *Game of Thrones*), mas direciona a criação dessa língua para a emancipação das mulheres. Na distopia de seu livro, a língua fictícia criada por ela é usada para que pessoas consigam se comunicar em um código específico e efetivamente fazer um movimento de resistência. É óbvio que essa linguista está fazendo alusão a movimentos reais de emancipação que são realizados por meio de línguas, as quais são faladas por comunidades específicas de falantes. Na história, temos vários exemplos disso.

Faço então uma pergunta engraçada: seria a língua uma espécie de telepatia? Hoje de manhã, eu estava um lendo um livro chamado *Ypabuçu: a vida nas lagoas*, escrito pelo biólogo Marcos Rodrigues. Esse livro, premiado da Editora da Unicamp, faz uma descrição meticulosa de uma lagoa durante vários meses. Eu não conheço essa lagoa, que fica no estado de Minas Gerais, mas lendo o livro dele eu fico entendendo como é essa lagoa. Às vezes, ouvindo um *podcast* passo a conhecer algo que não conheço, como um espaço físico ao qual não tenho acesso. Por meio de uma língua, posso também experienciar épocas que não vivi, coisas que não vi. Quer dizer, além da questão tecnológica de transmitir um conhecimento, uma língua também permite que tenhamos uma imagem mental muito nítida de algo ou de uma pessoa. Não é bonito pensarmos nos poderes que temos ao usar uma língua?

Bem, separei três palavras que foram criadas pela ficção científica, mas há muitas outras que podem ser inventadas. Trouxe essas porque são muito famosas: robô, robótica e ciberespaço [no telão, lê-se: “Invente a próxima”]. Foi o escritor tcheco Karel Capek quem cunhou, em sua peça *R.U.R.* (1920), o termo “robotá”, que usamos como “robô” em português. Depois, Isaac Asimov partiu dele e criou “robótica”, uma palavra que usou para se referir a uma ciência que à época não existia. Palavras como essas, inventadas pela literatura, expressavam um sentimento tão presente em sua época que se tornaram vocábulos cotidianos, a partir dos quais se desenvolveu toda uma etimologia. Hoje, temos faculdade de robótica, e criamos um novo termo, “bot”, para denominar algo diferente que surgiu em nossa época. É fascinante observar que, para designar um fenômeno, a literatura cria algo que ainda não existe, mas que depois ajuda a comunidade científica a nomear fenômenos técnicos.

Eu também selecionei a palavra “ciberespaço”. Ela foi criada por William Gibson, está presente em seu livro *Neuromancer*, de 1984. Gibson a usou para designar o que hoje chamamos de internet, esse espaço que seria criado de uma forma “cibernética”, que era o termo da época. *Neuromancer* é um livro muito crítico do avanço tecnológico e do que estava acontecendo; aliás, quase todos

os livros de ficção científica são. Em *Eu, Robô*, de Isaac Asimov, os possíveis problemas gerados pelos robôs estão sempre em questão, assim como em *R.U.R.*, de Capek, em que a palavra “robô” é uma metáfora crítica para as pessoas que trabalham demais. Por isso coloquei no telão a sugestão “Invente a próxima”. É um fato: novas palavras e novos termos sempre serão inventados pela literatura, pelo cinema etc.

O último ponto que quero abordar hoje trata da relação entre língua, inovação e poder. É fundamental que não nos esqueçamos de que toda língua sofre formas de policiamento, não podemos usar algumas palavras que são consideradas inadequadas em determinadas situações. Vocês podem imaginar uma série de palavras que não seria conveniente eu usar aqui, por exemplo. E o controle linguístico carrega em si também um controle ideológico muito forte. Saber falar, saber como falar ou falar “bem” marcam uma posição sociológica de quem fala. Em alguns idiomas, como no inglês ou mesmo no espanhol, a distância entre a língua falada e a escrita não é tão grande; em outros, como no português, ela é enorme. Aliás, quanto a isso, acho que a internet tem colaborado para diminuir um pouco a distância entre o português brasileiro falado e escrito, e – fazendo um exercício de futurologia, enquanto pesquisadora de ficção científica que está sempre pensando no futuro – acredito que, provavelmente, o uso muito recorrente de ferramentas de ditado e transcrição também vão atuar para o encurtamento dessa distância. Podemos nos perguntar por que a distância entre língua falada e escrita é tão grande no português brasileiro. Acredito que a resposta tenha muito a ver com a questão colonizadora. Historicamente, o ensino do português escrito ficou restrito a um grupo pequeno de pessoas. Em outros contextos culturais, o ensino da língua escrita alcançou mais camadas sociais há muito mais tempo. Assim, talvez em um futuro não muito distante, nós também tenhamos uma aproximação maior entre a língua que falamos e a língua que escrevemos.


As mudanças linguísticas, que acontecem o tempo todo e são inevitáveis também podem afetar as pessoas subjetivamente. Eu,

por exemplo, já tenho mais de 40 anos, e às vezes vejo as pessoas usando palavras que eu não sei usar. Nesse tipo de situação, podemos ter uma sensação de perda de controle ou de que não estamos mais no centro de alguma coisa. E não podemos esquecer que nossa forma de falar também diz muito. Ela dá muitas informações sobre nossa origem geográfica, nossa origem social e, às vezes, até mesmo sobre a cor de nossa pele, nossa orientação sexual ou religiosidade.

Chegamos agora na palavra “poder”. Precisamos fazer um exercício constante de nos assenhormos de nosso próprio idioma, de lembrar que existem, sim, instâncias que regulam a língua e que nos regulam. É uma tecnologia que temos e que não vai sair do nosso corpo durante muito tempo. A língua é um patrimônio, e faz muita diferença termos algumas palavras para nós.

Gostaria que vocês se recordassem agora da palavra que imaginaram no início da nossa conversa. Eu pensei na palavra “água”, acho uma palavra muito bonita, ela mimetiza o som de beber água. E “água” se fala para tantas coisas. Talvez seja algo tão destruidor quanto precioso. E que nos define tanto também. Outra palavra em que pensei, enquanto fazia alguns ajustes na minha fala antes de começarmos, foi “maritaca”, porque aqui no Museu fica cheio de maritacas no entardecer. Eu gosto muito das maritacas, onde moro também tem muitas, e é quase como se elas estivessem falando: “Ei, você está em casa?”. É engraçado, a palavra “maritaca” é já um pouco do ruído desse pássaro, parece que estão batendo um papo, fofocando. Não sei o que se passa na cabeça das maritacas, é minha interpretação humana do que elas falam, mas palavras como essa me trazem um sentimento de estar no meu lugar. Eu crio um lugar a partir disso, e isso não é pouco.

Quando pensarem em língua e tecnologia, pensem nessas possibilidades, pensem em como é potente lembrar que somos isso: língua, palavras. Esse é o convite que deixo.

A decorative graphic consisting of several blue circles of varying sizes connected by thin blue lines. One circle on the right is larger and contains several concentric blue circles. The lines and circles are scattered across the page, with some overlapping the text area.

VIAGENS DA LÍNGUA: A CONVERGÊNCIA DA INTELIGÊNCIA ARTIFICIAL COM AS LÍNGUAS INDÍGENAS

NINA DA HORA

*Instituto da Hora e
Universidade Estadual de Campinas*

RAYMUNDO BANIWA

Rede Wayuri e Instituto da Hora

A preservação das línguas indígenas é uma questão crucial de direitos culturais e identidade. Com o avanço da Inteligência Artificial, novas ferramentas surgem como potenciais aliadas na luta contra o esquecimento linguístico. Este texto explora como a IA pode auxiliar na revitalização de línguas indígenas e quais são os desafios éticos e técnicos envolvidos nesse processo.

Na interseção entre tecnologia e preservação cultural, a Inteligência Artificial (IA) surge como uma ferramenta promissora para auxiliar na compreensão e na tradução de línguas indígenas. Processadores de linguagem natural e sistemas de tradução automática estão sendo cada vez mais adaptados para lidar com idiomas que, por muito tempo, foram negligenciados por tecnologias digitais. Projetos de documentação linguística, por exemplo, utilizam IA para criar extensos bancos de dados de idiomas ameaçados de extinção, facilitando o acesso e o estudo dessas línguas.

Um exemplo prático do potencial dessas tecnologias é visto em aplicativos de aprendizado de línguas que incorporam dialetos indígenas. Tais plataformas não apenas aumentam a visibilidade dessas línguas, mas também permitem que pessoas de todo o mundo tenham a oportunidade de aprendê-las e praticá-las, promovendo maior inclusão linguística e cultural.

No entanto, a integração da IA com línguas indígenas não está isenta de desafios. Do ponto de vista técnico, a escassez de grandes conjuntos de dados linguísticos, que são essenciais para treinar algoritmos de maneira eficaz, é um obstáculo significativo. Muitas línguas indígenas, por terem sido tradicionalmente transmitidas de forma oral e tendo um número limitado de falantes, carecem de dados escritos suficientes para o desenvolvimento de tecnologias robustas de IA.

Além dos desafios técnicos, existem questões éticas importantes que precisam ser abordadas. A coleta de dados linguísticos levanta preocupações sobre a apropriação cultural e a necessidade de garantir que as comunidades indígenas sejam envolvidas ativamente em todas as etapas do desenvolvimento dessas tecnologias. É crucial que esses projetos sejam conduzidos de maneira ética, com transparência e respeito pelas culturas que estão sendo representadas e auxiliadas.

Do ponto de vista cultural e social, a IA pode funcionar como uma ferramenta de empoderamento para comunidades indígenas, ajudando a fortalecer suas identidades culturais e a promover maior reconhecimento e respeito por essas culturas no cenário global. Contudo, existe o risco de homogeneização e perda de nuances culturais únicas quando a IA é usada para interpretar e traduzir línguas indígenas. Isso pode levar a uma simplificação excessiva de idiomas complexos, resultando em perda de significado e de conexão cultural.

Assim, enquanto a IA apresenta oportunidades significativas para a revitalização e a preservação de línguas indígenas, é fundamental que seu desenvolvimento e sua aplicação sejam feitos com uma reflexão cuidadosa sobre os impactos técnicos, éticos e culturais envolvidos.

A IA tem um potencial significativo para auxiliar na preservação e na revitalização de línguas indígenas, oferecendo recursos que podem aumentar a acessibilidade e o engajamento linguístico. No entanto, é vital que a implementação dessas tecnologias seja conduzida com um forte compromisso ético e colaborativo, assegurando que as comunidades indígenas sejam as principais beneficiárias e participantes ativas no processo tecnológico. A “Viagem da Língua” é, portanto, tanto uma jornada tecnológica quanto uma jornada de reconhecimento e respeito pelas vozes indígenas.

LINGUAGEM SIMPLES E COMPLEXA: METACOGNIÇÃO EPILINGUÍSTICA PARA O LETRAMENTO E CIDADANIA PLENOS¹

MARCUS MAIA

Universidade Federal do Rio de Janeiro

GISELE RODRIGUES

Universidade de Brasília

I – Introdução

Procuramos exercitar, no presente capítulo, uma reflexão sobre uma questão prática que opõe, de um lado, a complexidade estrutural natural da linguagem humana e seu desenvolvimento na educação linguística e, de outro, a proposta de formulação de uma política nacional² que institui padrões gerais de uso da língua no escopo do que vem sendo chamado de “linguagem simples”, na forma de um conjunto de regras fixas a serem observadas pelo governo em todas as suas ações de comunicação dirigidas à população. Para tentar contribuir com esse debate, propomos usar o modelo de arquitetura da linguagem exposto por Chomsky (2005) em *Three Factors in Language Design* (Três fatores no design da linguagem).

O primeiro dos três fatores apresentados por Chomsky representa os princípios inatos da linguagem, “o que vem de dentro”. Este fator pode fundamentalmente ser observado na operação

Merge – juntar, que é a capacidade humana de formar estrutura a partir de dados linguísticos disponíveis. Ninguém ensina a criança a fazer estrutura; trata-se de um universal da cognição linguística humana que constrói binariamente estruturas potencialmente complexas, em relações hierárquicas que, ao serem computadas, permitem a representação mental, a compreensão e a produção da linguagem.

O segundo fator de arquitetura da linguagem seria, segundo Chomsky, “o que vem de fora”, ou seja, os estímulos da experiência sociocultural que nos envolve. Este fator não nos determinaria *behavioristicamente* moldando pelos estímulos externos uma mente vazia, do tipo *tábula rasa*, mas interagiria com o primeiro fator, ativando os princípios estruturantes inatos. Em outras palavras, a experiência com o mundo exterior atuaria, a partir da ativação da capacidade inata de formar estrutura, na fixação de parâmetros que definem o funcionamento da língua e viabilizam seu uso.

Finalmente, o terceiro fator do *design* da linguagem humana seria, de acordo com Chomsky, externo à própria linguagem, consistindo fundamentalmente de **princípios de economia e de simplicidade**³ e de seus epifenômenos, tais como a simetria e a localidade. Note-se que, em Trotzke, Bader & Frazier (2013), faz-se a proposta de que o processamento da linguagem humana, em si, possa ser concebido como o terceiro fator ao qual se refere Chomsky. De fato, os princípios de processamento estabelecidos por meio de vários estudos de natureza experimental baseiam-se todos em métricas de economia e de simplicidade. Por exemplo, tome-se o princípio conhecido como *Minimal Attachment* – Aposição Mínima, atuante no processamento de frases estruturalmente ambíguas, como a da notícia “Mãe suspeita de assassinato do filho morre”. Provavelmente, ao ler a frase, mesmo um leitor que já conheça essas teorias terá entrado naturalmente no que se chama de *garden-path*. Explicamos. A métrica reflexa e automática de economia de processamento rapidamente guia a decisão do leitor pela forma estruturalmente mais simples, o que o faz analisar imediatamente o termo “suspeita” como verbo flexionado em terceira pessoa do singular, no presente do indicativo, ou seja, a mãe suspeitaria de algo, seria aquela que

pratica uma ação. Entretanto, ao se chegar à forma verbal “morre”, esta análise *default*, baseada em simplicidade⁴, não pode ser mantida. Nesse momento, instancia-se o chamado *garden-path*. Embora a estratégia de processamento (terceiro fator) seja a de simplicidade, há o fator de complexidade estrutural natural (primeiro fator). O processador, então, deverá reanalisar a frase nos termos mais complexos em que ela está de fato formulada, em estrutura com mais nós sintáticos, da qual faz parte uma oração subordinada adjetiva onde o termo ambíguo (“suspeita”) ocorre na forma nominal – a mãe que é suspeita. Essa tensão natural entre os fatores **complexidade estrutural** e **simplicidade de processamento** pode, claro, ser mediada pelo fator “de fora”, que abrange desde o ambiente linguístico natural em que a criança se desenvolve até as pressões socioculturais, escolares e/ou de ordem normativa ou mesmo prescritiva, que têm, sem dúvida, efeitos sobre a aquisição e o uso da linguagem em diferentes níveis.

Não pretendemos discutir aqui de modo mais amplo a necessidade ou a pertinência de formulações prescritivas regulatórias do uso linguístico, senão apontar que (i) a complexidade estrutural nos é intrínseca; (ii) o processamento da linguagem é necessariamente baseado em métricas de simplicidade; (iii) regras normativas de simplicidade deveriam, no mínimo, ser pensadas à luz de princípios naturais de processamento, e não estipuladas arbitrariamente, “de fora para dentro”, como ora se tenta fazer por meio da prescrição de fórmulas de “linguagem simples” a serem seguidas pelo governo em sua comunicação com a sociedade. Nesse debate, os estudos psicolinguísticos teriam relevância direta para oferecer subsídios legítimos na definição de métricas de simplicidade natural, sendo questionável, portanto, o viés externalista de base prescritivista e sem fundamentação explícita adotado no texto legislativo que estabelece uma política nacional de “linguagem simples”.

Tendo em vista esse quadro analítico de três fatores – complexidade interna natural, pressão prescritivista externa e simplicidade natural de processamento –, o presente texto organiza-se da seguinte forma. Na seção II, discutimos o primeiro fator – a

complexidade natural – e seu desenvolvimento por meio de práticas metacognitivas epilinguísticas, ou seja, atividades ativas em que estruturas gramaticais são objeto de reflexão a partir de problemas e exemplos concretos, com base na análise de dados de leitura obtidos através da técnica de rastreamento ocular (*eye-tracking*). Na seção III, a respeito da criação da política nacional de “linguagem simples”, comentamos, mais especificamente, a determinação de se organizar qualquer texto de forma que “as informações mais importantes apareçam primeiro”. Nessa reflexão, ao demonstrarmos que atividades educacionais epilinguísticas têm possibilidade efetiva de promover o desenvolvimento da competência analítica e leitora, entendemos que pode ser considerada simplória a proposta de obrigar que textos apresentem a ideia principal logo em seu início, na medida em que o critério da linearidade desconsidera a complexidade estrutural natural da linguagem humana. Em outras palavras, propomos que o excesso de simplificação da linguagem em nada contribui para a literacia plena, para a vigilância epistêmica e para a autodefesa intelectual⁵. Por fim, na seção IV, oferecemos conclusões preliminares acerca da questão em análise no texto, argumentando que o viés prescritivista sem clara fundamentação da proposta que cria a política nacional de “linguagem simples” é, no mínimo, equivocado e ingênuo.

II – Rastreamento ocular da leitura de períodos – atividades metacognitivas

Em vários textos, o linguista Carlos Franchi avaliou que a educação linguística no Brasil costuma ser excessivamente metalinguística, devendo haver, segundo o autor, mais oportunidades para se exercitarem atividades epilinguísticas na sala de aula. Por exemplo, em *Mas o que é mesmo gramática?*, Franchi referia-se a essas atividades assim:

... chamamos de atividade epilinguística a essa prática que opera sobre a própria linguagem, compara as expressões, transforma-as, experimenta novos modos de construção canônicos ou não, brinca com a linguagem,

investe as formas linguísticas de novas significações.
(FRANCHI, 2006, p. 97)

Nessa perspectiva, entendemos que exercícios epilinguísticos podem ser usados como ferramenta para desenvolver o letramento científico e a capacidade de leitura dos alunos. Nossa proposta epilinguística beneficia-se, ademais, do uso da técnica de rastreamento ocular (*eye-tracking*), em que o Laboratório de Psicolinguística Experimental (LAPEX) da UFRJ foi pioneiro na pesquisa psicolinguística no Brasil. Além de gerar planilhas quantitativas detalhadas sobre diferentes métricas de inspeção visual, a técnica também permite a visualização qualitativa de imagens em mapas de calor (*heatmaps*) e de fixação estáticos e dinâmicos (*gazeplots*).

Sem maiores delongas, passemos às atividades epilinguísticas a partir de dados obtidos em estudos de rastreamento ocular (*eye-tracking*) de períodos. Observem-se, por exemplo, os mapas de calor da leitura de dois períodos, como indicado nas figuras 1 e 2:



Figura 1: Leitura de período
por aluno de Ensino Superior.
Fonte: primeiro autor.



Figura 2: Leitura de período por
aluno de Ensino Fundamental II.
Fonte: primeiro autor.

Temos levado imagens microscópicas de leitura de períodos e de visualização de imagens para oficinas com alunos e professores em aulas de português na Educação Básica e de Linguística na Educação Superior. Mesmo sem muitas informações prévias sobre a técnica de rastreamento ocular e seus produtos qualitativos, tais

como os mapas de calor acima, de modo geral os participantes das oficinas engajam-se animadamente na observação, comparação, contraste, formulação de hipóteses e nas conclusões sobre as diferentes leituras. Os mapas de calor são bastante intuitivos e nem é preciso dizer que as áreas mais vermelhas atraíram mais a atenção na leitura do que as amarelas e verdes. Concordam?

Então, o que aprendemos sobre essas leituras? Geralmente, testes mais tradicionais são restritos às práticas de “interpretação de texto”, em que se obtêm respostas a perguntas feitas após a leitura. Com o rastreador ocular, no entanto, podemos analisar os processos de leitura no momento mesmo em que eles ocorrem. Como analisado quantitativa e qualitativamente no livro *Psicolinguística e metacognição na escola* (MAIA, 2019), os bons leitores tipicamente percorrem todo o período, tendendo a fixar mais sua atenção em áreas críticas estruturais do que leem. Em contraste, leitores menos proficientes parecem engajar-se mais no início das frases, percorrendo linearmente o período e, frequentemente, desengajando-se antes de concluir a leitura de todo o texto.

Vamos sistematizar nossas observações epilinguísticas? O período em exame acima é “Embora procedendo erroneamente ao invadir as ilhas Malvinas, a Argentina tem direito incontestável às ilhas, que ficam dentro dos limites de seu mar territorial”. A comparação epilinguística entre as duas leituras permite observar que o leitor mais proficiente, de Ensino Superior (figura 1), percorreu todo o período, com maiores latências de fixação ocular em uma certa área (qual?). Em contraste, o leitor menos proficiente (figura 2) engajou-se mais no início e foi desengajando-se progressivamente, nem mesmo lendo todo o texto até o final. A epilinguística motiva a metalinguística. A área crítica que recebe mais fixações oculares na leitura é a oração principal (OP), que constitui a perspectiva do período. As demais orações são subordinadas: elaboram sobre a ideia principal, denominada de *gist* na Psicolinguística da Leitura, a fim de indicar justamente a informação mais essencial, o tópico frasal que deve ser extraído pelo bom leitor.

Verifique-se, agora, a leitura típica de período por leitor proficiente, com outra oração principal, como ilustrado na figura 3:



Figura 3: Leitura de período por leitor proficiente.
Fonte: primeiro autor.

Em contraste com o período ilustrado nas figuras 1 e 2, no período da figura 3, “Embora tenha direito incontestável às ilhas Malvinas, a Argentina procedeu erroneamente ao invadir as ilhas que ficam dentro dos limites de seu mar territorial”, a estrutura sintática é rearticulada, passando-se a destacar como oração principal o fato de que a Argentina teria procedido erroneamente, em vez do fato de que a Argentina teria direito às ilhas, ideia essa contida na oração principal das figuras 1 e 2. Note-se que, apesar de conter basicamente o mesmo material lexical dos períodos nas figuras 1 e 2, o período na figura 3 obtém perspectiva diametralmente oposta, enfatizando o que seria um procedimento errôneo da Argentina, o qual, em 1 e 2, é minimizado como oração subordinada, em benefício da ideia de que a Argentina teria direito às ilhas. Quem lê períodos como em 1 e 2 em princípio ativa representações mentais favoráveis à Argentina, enquanto, na articulação em 3, o que fica mais ativo na memória do leitor é um possível erro da Argentina.

Nos casos das leituras ilustradas até aqui, a OP está em segunda ou terceira posição linear no período. Note, no entanto, que a OP poderia estar até mesmo na posição final do período, como vemos na figura 4, que representa caso de leitura proficiente de “Quando foram finalmente aprovadas, após realizarem-se longos debates na Câmara Municipal, tendo custado muito pouco aos cofres municipais, as obras da estação de esgoto receberam apoio de empresas privadas”.

Concluindo esta seção, podemos avaliar que há grande flexibilidade natural tanto na escolha da OP quanto em relação ao seu

posicionamento nos períodos. A OP costuma atrair o olhar e ter maior perduração mnemônica. Como detalhado em Maia (2022), em tempos infodêmicos⁶ como os que vivemos, o desenvolvimento de flexibilidade cognitiva, a partir da metacognição do ponto de vista de períodos, é fundamento central para a cidadania plena, pois promove vigilância epistêmica e autodefesa intelectual.

Quando foram finalmente aprovadas, após realizarem-se longos debates na Câmara Municipal, teriam custado muito pouco aos cofres municipais, as obras custaram pouco mais do que o necessário para serem realizadas.

Figura 4: Mapa de calor da leitura de período em que a OP ocorre ao final.
Fonte: primeiro autor.

III – Linguagem simples?

Conforme identificamos em *Psicolinguística e metacognição na escola* (MAIA, 2019), os leitores proficientes não são lineares, mas “estruturantes seletivos”, devendo ser capazes de extrair o ponto de vista essencial do que leem (*gist*), independentemente da sua posição no período. Nesse sentido, parece-nos no mínimo questionável a formulação de uma política nacional de “linguagem simples” no âmbito governamental que, entre outras disposições, prescreva, além da linearidade, a redação de frases curtas e em ordem direta, em uma padronização genérica que se opõe à complexidade natural da linguagem e que em nada contribui para o desenvolvimento de capacidade leitora, ao privar o leitor de estímulos que o fariam lidar com estruturas complexas a partir de sua capacidade inata ou, ainda, ao não ensejar seu contato com atividades epilinguísticas que contribuiriam para desenvolver sua proficiência.

Uma análise atenta da tramitação da proposta que cria a política nacional de “linguagem simples” nos leva a concluir que os debates ocorridos sobre o tema na Câmara dos Deputados ao longo do ano de 2023 foram marcados pela tentativa de deixar a matéria ainda mais prescritiva – e restritiva – quanto à definição de regras engessadas de uso da língua. Exemplo disso é a penúltima versão

do texto apresentado no plenário da Casa legislativa em dezembro daquele ano, que condenava o uso da voz passiva, de construções intercaladas e de substantivos no lugar de verbos. Essas restrições acabaram sendo excluídas da proposta antes que fosse aprovada, mas tal fato não retira o caráter prescritivista e generalizante do texto que prevaleceu ao fim da votação.

Em pesquisas de leitura que temos desenvolvido na Educação Básica, identificamos, de fato, a existência de leitores lineares incompletos que, como ilustrado na figura 2, engajam-se no início da frase, não fazem uma análise estrutural e desengajam-se precocemente da leitura, não conseguindo, portanto, extrair o ponto de vista (*gist*) do que leem. Entretanto, importante: descobrimos também, nessas pesquisas, que, usando-se metodologias ativas epilinguísticas em oficinas metacognitivas nas quais os alunos têm a chance de comparar mapas de calor e de fixação ocular de leituras lineares com mapas de leituras estruturantes, **é possível fazer avançar significativamente a competência de leitores lineares**, que podem passar a analisar estruturalmente os períodos, em busca do *gist* independentemente de sua posição linear no período. Comparem-se, por exemplo, as leituras registradas nos mapas de fixação dinâmicos 1 e 2, a seguir. O mapa em 1 ilustra a leitura de um aluno de Ensino Fundamental que parece ser um leitor linear, não estruturante. O mapa em 2 mostra a leitura de um aluno de Ensino Superior, mais proficiente.

Quando um desconhecido entrou no pátio, causando uma enorme confusão, houve brigas na hora do recreio, o que levou a diretora a suspender vários alunos.

Mapa de Fixação Ocular 1: Leitor linear de Ensino Fundamental.⁷

Link de acesso aos vídeos nas notas de referência e QR Codes.



Quando um desconhecido entrou no pátio, causando uma enorme confusão, houve brigas na hora do recreio, o que levou a diretora a suspender vários alunos. ●

Mapa de Fixação Ocular 2: Leitor estruturante seletivo de Ensino Superior.⁸



A observação ativa de dados de leitura como esses, promovida durante oficinas em que os alunos exercitam sua capacidade de comparar, contrastar, formular hipóteses, além de estimular o desenvolvimento do raciocínio analítico, tem o potencial de ativar processos metacognitivos regulatórios que, na nossa experiência prática no “chão da escola”, podem levar leitores lineares, como exemplificados no mapa de fixação 1, a se tornar estruturantes seletivos, como ilustrado no mapa de fixação 3, que indica a leitura de aluno do Ensino Fundamental após participar das oficinas metacognitivas desenvolvidas na escola.

Quando um desconhecido entrou no pátio, causando uma enorme confusão, houve brigas na hora do recreio, o que levou a diretora a suspender vários alunos.

Mapa de Fixação Ocular 3: Leitor de Ensino Fundamental após oficinas.⁹



Repassemos as observações possíveis trabalhadas em oficinas metacognitivas em salas de aula da Educação Básica a partir da visualização desses mapas de fixação ocular. Por exemplo, no mapa 1, ao menos dois comportamentos recorrentes costumam ser facilmente identificados pelos alunos nas oficinas, a saber: o engajamento na primeira oração do período, que merece, inclusive,

releitura; e o desengajamento da leitura a partir da segunda oração. No mapa 2, ao contrário, o período recebe fixações progressivas, sem desengajamento até o final da leitura, e a OP recebe refixações seletivas, tanto na primeira quanto na segunda passagem do olhar. Finalmente, o mapa 3 é bastante esclarecedor quanto aos resultados obtidos nas oficinas: os leitores lineares que observam e discutem comparativamente os dados de leitura de alunos menos e mais proficientes passam a fazer leituras mais completas e – principalmente – passam a refixar as áreas críticas, tendo melhoras significativas em sua capacidade de extração do *gist*. Em *Psicolinguística e metacognição na escola*, apresentamos o gráfico abaixo, que ilustra a retestagem de leitura pós-oficinas, indicando melhoras robustas na capacidade de identificação das OPs por parte de alunos de Ensino Fundamental que participaram das atividades epilinguísticas e acabaram se aproximando dos níveis encontrados em leitores de Ensino Superior.

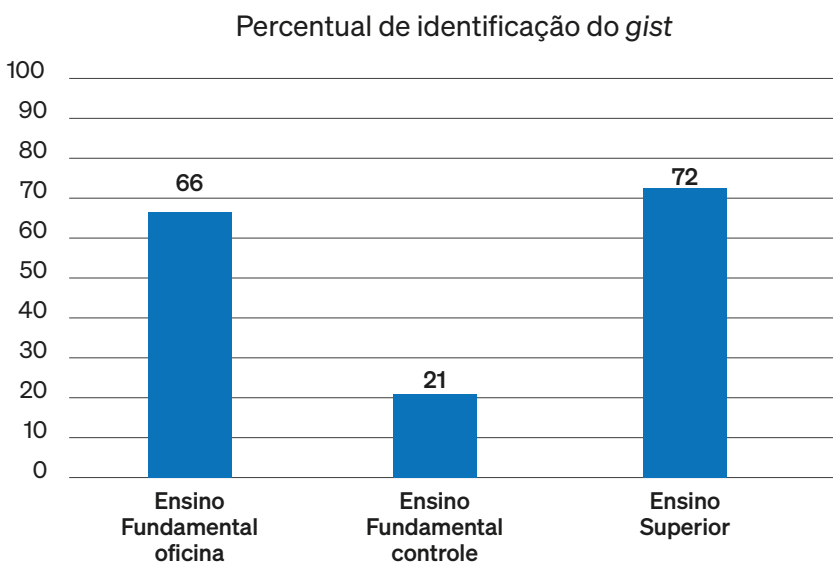


Gráfico 1: Percentuais de extração da OP em grupos de Ensino Fundamental (oficina e controle) e Ensino Superior.

De volta à proposta que cria uma política nacional de “linguagem simples”, entendemos que muitas das regras descritas como garantidoras de uma boa comunicação pública correm o risco de configurar parâmetros homogeneizantes e reducionistas de uso da língua portuguesa, na medida em que apontam, sobretudo, para a oferta de construções sintáticas lineares que levam à exclusão de encaixes recursivos fundamentais da faculdade humana de linguagem. No nosso entendimento, essa e outras padronizações de uso da língua ainda desconsideram a diversidade do público leitor e ignoram o potencial inerente a qualquer texto de concorrer para o desenvolvimento de flexibilidade cognitiva e capacidade leitora.

Não se trata de advogar pelo emprego de formas complicadas, herméticas e corporativistas de uso da língua, tal qual o fazem certos grupos de juristas, médicos e cientistas; nem de valorizar a erudição nas formas de comunicação entre governo e sociedade. No entanto, transformar o que seriam princípios gerais de linguagem simples em regras e técnicas padronizadas de uso da língua não parece ser garantia de alcance de uma tão desejada comunicação cidadã.

Inevitável questionar: haveria uma única fórmula, traduzível na forma de artigos e incisos de uma lei, capaz de garantir uma comunicação eficaz, acessível aos diferentes públicos? Não estaria o serviço público diante do desafio de diversificar suas formas de comunicação, no sentido de contemplar a heterogeneidade de grupos a que se dirige, em vez de buscar apenas planificar a linguagem e ditar uma única maneira de utilizá-la? Não seria a simplificação indiscriminada, que agora se pretende prescrita no ordenamento jurídico, uma forma de inviabilizar, no bojo do processamento linguístico, camadas de significados e compreensões caras a uma comunicação plena e abrangente?

Em um país cuja população carece, em grande parte, de letramento crítico e proficiência em leitura, pensamos ser necessária a reflexão: ainda que na busca de uma comunicação cidadã e democrática, não seria muito “simplista” a estratégia de adotar padrões genéricos e reducionistas de uso da língua? Não deveriam ser incluídas, no debate, questões primordiais referentes à educação e às

políticas públicas do setor? Será mesmo produtivo focar em apenas ajustar toda e qualquer comunicação governamental à capacidade leitora daqueles que lamentavelmente estão à margem de maiores conquistas nesse campo? O governo, que naturalmente é fonte primária de informação, não será levado, pela estrita observância das regras de uma pretensa linguagem simples, a atribuir a si próprio o superpoder de selecionar, editar, empacotar, formatar e redistribuir informações que ele mesmo produz? Não estará o governo, assim, correndo o risco de privar o cidadão comum do contato direto com domínios do conhecimento e da linguagem que poderiam lhe expandir o repertório e a capacidade leitora?

Como já dissemos, não se trata de apoiar que mensagens ininteligíveis sejam a tônica da comunicação entre governo e sociedade. Sem dúvida, estamos diante de uma estratégia elitista, antidemocrática, excludente e “cafona” quando nos deparamos com serviços e comandos expressos na forma de palavras rebuscadas e construções complicadas que dificultam ou impedem a compreensão de conteúdos das páginas de internet da Previdência Social e dos tribunais de justiça, por exemplo. No entanto, a definição de regras “universais” de uso simplificado da língua, alçadas a lei sem critérios explícitos e sem o mínimo debate necessário, não parece ser a única ou a melhor solução para um problema tão delicado e profundo do nosso tecido social. Estaríamos nós, a título de remediação, buscando apenas ajustar o quadro à moldura? Vamos falar mais de educação?

IV – Considerações finais

Como procuramos resumir neste pequeno artigo, o modelo de Chomsky (2005) que apresenta três fatores atuantes no *design* da linguagem humana parece fornecer um quadro teórico apropriado para aprofundarmos a reflexão sobre a complexidade estrutural a ela inerente, bem como sobre a simplicidade característica do processamento linguístico, em tensão com o viés prescritivista externalista, frequentemente arbitrário, que se manifesta em propostas como a que cria uma política nacional de “linguagem

simples”. Paralelamente, demonstramos ser plenamente possível contribuir para o desenvolvimento da capacidade científica e da competência leitora de alunos da Educação Básica por meio do uso de metodologias ativas de natureza epilinguística e metacognitiva, promovendo-se, inclusive, a translação para a escola de técnicas como o rastreamento ocular¹⁰. Nesse sentido, argumentamos que a Linguística Teórica e a Psicolinguística oferecem evidências para, ao menos, motivar um debate mais aprofundado e abrangente sobre a referida política, de forma que sejam levados em consideração critérios pertinentes à Linguística enquanto ciência que estuda os fenômenos da linguagem. Como resenhamos em Maia (2022), há mesmo evidências de que a falta de exercício em operações sintáticas envolvendo complexidade estrutural pode acarretar prejuízos para processos neuronais importantes, tais como a síntese pré-frontal (cf. VYSHEDSKIY, 2019; VYSHEDSKIY *et al.*, 2017).

Assim, encerramos este texto indagando: se há uma capacidade estrutural complexa inata por meio da qual é possível desenvolver competências cognitivas complexas capazes de gerar impacto positivo na leitura e na escrita, por que investir em políticas públicas de simplificação estrutural, preconizando-se, por exemplo, a adoção de períodos curtos que obrigatoriamente apresentem a ideia principal em seu início? Como atingir cidadania plena promovendo a formação de amplos segmentos da população incapazes de desenvolver plenamente sua cognição e sua autodefesa intelectual por estarem restritos a simplismos em um mundo necessariamente complexo? À parte alguma simplificação até certo ponto aceitável no uso de itens lexicais¹¹, a supersimplificação no nível mais central da linguagem humana – propriedade básica que nos torna *Homo Sapiens*, como lembra Chomsky (2018) – e a criação de entraves para o desenvolvimento de competências fundamentais baseadas na sua complexidade estrutural parecem-nos, de fato, no mínimo mais um equívoco prescritivista, quando não uma forma suspeita de “terrapijanismo linguístico”¹².

Palavras-chave: epilinguística; metacognição do período; complexidade estrutural; rastreamento ocular; linguagem simples.

Referências

- BAILLARGEON, N. **A Short Course in Intellectual Self Defense**: find your inner Chomsky. Penguin Books, 2011.
- CHOMSKY, N. Three Factors in Language Design. **Linguistic Inquiry**, v. 36, n. 1, 2005, p. 1-22.
- CHOMSKY, N. **Que tipo de criaturas somos nós?**. Petrópolis: Editora Vozes, 2018.
- DUARTE, J.; DUARTE, M. Serviço Público, comunicação e cidadania, em MARETTI, E.; NASSAR, P. (org.). **Comunicação pública: por uma prática mais republicana**. São Paulo: Aberje, 2019, p. 57-77.
- FRANCHI, C. **Mas o que é mesmo 'gramática'?**. São Paulo: Parábola, 2006.
- MAIA, M. Linguística Experimental: aferindo o curso temporal e a profundidade do processamento. **Revista de Estudos da Linguagem**, v. 21, 2013, p. 9-42.
- MAIA, M. Dimensões do processamento sintático, em MOTA, M. B.; NAME, C. (org.). **Interface linguagem e cognição: contribuições da Psicolinguística**. Tubarão: Copiart, v. 1, 2019, p. 89-114.
- MAIA, M. (org.). **Psicolinguística e metacognição na escola**. Campinas: Mercado de Letras, 2019.
- MAIA, M. Eye-Tracking Sentences in Language Education. **Diacrítica**, v. 36, 2022, p. 6-36.
- PEREIRA, L. M.; MOUSINHO, R.; MAIA, M. Figure Identification in Dyslexia: an eye-tracking study. **International Journal of Psychology and Neuroscience**, v. 2, 2016, p. 88-107.
- RIBEIRO, A. A abordagem *good-enough* e o processamento de frases do português do Brasil. **Veredas**, n. 2, 2008, p. 62-75.
- ROEPER, T.; MAIA, M.; PILATI, E. **Experimentando Linguística na escola: conhecimento gramatical, leitura e escrita**. Campinas: Pontes, 2020.

- SPERBER, D. et al. Epistemic Vigilance. **Mind & Language**, v. 25, n. 4, set. 2010, p. 359-393.
- TROTZKE, A.; BADER, M.; FRAZIER, L. Third Factors and the Performance Interface in Language Design. **Biolinguistics**, n. 7, 2013, p. 1-34.
- VYSHEDSKIY, A. Language Evolution to Revolution: the leap from rich-vocabulary non-recursive communication system to recursive language 70,000 years ago was associated with acquisition of a novel component of imagination, called Prefrontal Synthesis, enabled by a mutation that slowed down the prefrontal cortex maturation simultaneously in two or more children – the Romulus and Remus hypothesis. **Research Ideas and Outcomes**, 5(4), 2019. Disponível em: <https://doi.org/10.3897/rio.5.e38546>.
- VYSHEDSKIY, A.; MAHAPATRA, S.; DUNN, R. Linguistically deprived children: Meta-analysis of published research underlines the importance of early syntactic language use for normal brain development. **Research Ideas and Outcomes**, 3, 2017. Disponível em: <https://doi.org/10.3897/rio.3.e20696>.

Notas

- [1] O presente texto resume ideias apresentadas pelo primeiro autor e discutidas em mesa-redonda no II Seminário Viagens da Língua: Língua e tecnologias, realizada em plataforma remota no segundo semestre de 2023.
- [2] A tramitação do Projeto de Lei n. 6.256/2019, que cria a Política Nacional de Linguagem Simples no âmbito da comunicação governamental, pode ser acessada em www.camara.leg.br e www.senado.leg.br.
- [3] Note-se que princípios de economia e simplicidade são basilares na Filosofia da Ciência, tal como a “navalha de Occam”, que propõe que, diante de várias explicações para um problema, a mais simples deve ser preferida.
- [4] Da mesma forma, encontram-se métricas de economia nos demais princípios de processamento no âmbito da Teoria do Garden-Path,

tais como os princípios da Aposição Local (Late Closure), da Cadeia Mínima (Minimal Chain), do Antecedente Ativo (Active Filler), entre outros. Para uma visão panorâmica desses princípios da TGP, leia-se Maia (2019), Maia (2013) e Ribeiro (2008), entre outros.

- [5] Sobre o conceito de vigilância epistêmica, leia-se Sperber *et al.* (2010); sobre o conceito de autodefesa intelectual, leia-se Baillargeon (2011).
- [6] A Organização Mundial de Saúde define o termo infodemia como caracterizando um excesso de informação ou de informação errônea (*fake news*). Disponível em: https://www.who.int/health-topics/infodemic#tab=tab_1
- [7] Mapa de Fixação Ocular 1: https://drive.google.com/file/d/19CArb9VmT2ukpHNW77Z_8CKoJBHFQG66/view?usp=sharing.
- [8] Mapa de Fixação Ocular 2: https://drive.google.com/file/d/1QwwjS6h7eYE_DBOtI9CFwQz6ADZAbobp/view?usp=sharing.
- [9] Mapa de Fixação Ocular 3: https://drive.google.com/file/d/1-_6j0sZfHSzHI0BbIRvMf_bpYRxde2FA/view?usp=sharing.
- [10] Em Roeper, Maia e Pilati (2020), discute-se em maior detalhe como práticas ativas que usam estruturas linguísticas complexas, tais como quantificação, elipses, ambiguidades, relações de longa distância, pressuposições e inferências, podem ser levadas “ao chão da escola”, com bons resultados no desenvolvimento das capacidades científica, de leitura e de escrita dos alunos. Em Maia (2022), discute-se com maior profundidade questões relevantes, tais como a centralidade do período, a metacognição, o rastreamento ocular, bem como resumem-se aplicações práticas desses métodos e técnicas na Educação Básica e em cursos de Linguística na graduação e na pós-graduação.
- [11] Simplificações no nível do vocabulário, poderiam, talvez, facilitar a compreensão, por exemplo, de textos jurídicos ou médicos por parte da população mais ampla. Além disso, populações com distúrbios devem ser objeto de ações afirmativas específicas, o que não deveria implicar em políticas mais amplas a partir dessas especificidades. Por exemplo, em Pereira, Mousinho & Maia (2017), identificou-se, através da técnica de rastreamento ocular, que disléxicos se beneficiam significativamente da leitura de textos escritos em letra de forma, em que há ângulos que atraem fixações na leitura, comparativamente à letra cursiva, o que leva os autores a sugerir o uso de letra de forma para essa população específica, mas, obviamente, sem estender esse critério para a população em geral.
- [12] Expressão ouvida do linguista Xoán Lagares (Universidade Federal Fluminense) em comunicação pessoal em fevereiro de 2024.

A LINGUAGEM SIMPLES E AS INSTITUIÇÕES CULTURAIS

JOSÉ DE SOUZA MUNIZ JR.

Centro Federal de Educação Tecnológica de Minas Gerais

No dia 7 de dezembro de 2023, no II Seminário Viagens da Língua, tive a oportunidade de apresentar algumas reflexões sobre o uso da linguagem simples em contextos de divulgação do conhecimento, em particular nas instituições culturais. LS é a sigla que usarei neste texto para falar deste tema: a linguagem simples.

Dois dias antes do seminário, dois fatos importantes marcaram a discussão sobre LS no Brasil. O primeiro ocorreu na Câmara dos Deputados, em Brasília, DF, onde foi aprovado um projeto de lei (bastante problemático e controverso, por sinal) que pretende criar uma política nacional para o tema. Já em Salvador, BA, no 17º Encontro Nacional do Poder Judiciário, o Conselho Nacional de Justiça lançou o Pacto Nacional do Judiciário pela Linguagem Simples. Os dois fatos (e muitos outros que poderiam ser mencionados) mostram que vários setores da sociedade estão se tornando

mais sensíveis ao tema. Nos próximos anos, esse assunto deverá ocupar a atenção de gestores públicos, produtores culturais, linguistas e outros profissionais.

Neste artigo, quero falar da importância do tema para todos os cidadãos e, principalmente, para quem trabalha na área cultural. Antes de mais nada, vale explicar o que é a LS.

Na definição do projeto de lei mencionado acima, ela é um “conjunto de técnicas destinadas à transmissão clara e objetiva de informações, de modo que as palavras, a estrutura e o leiaute da mensagem permitam ao cidadão facilmente encontrar a informação, compreendê-la e usá-la” (BRASIL, 2019). Já em uma apostila sobre o tema elaborada pela Prefeitura de São Paulo, lê-se o seguinte: “Uma comunicação é simples quando a pessoa que lê o documento ou escuta a mensagem, consegue entender o conteúdo de forma fácil, sem precisar reler o texto várias vezes ou pedir explicações para outra pessoa” (PREFEITURA DE SÃO PAULO, 2020, p. 6).

Uma das estratégias mais importantes da LS é evitar palavras que pouca gente conhece. Se não der para excluir esses termos do texto, é melhor explicar o que eles significam. Frases enormes, invertidas e cheias de intercalações também atrapalham bastante a vida de quem lê. Melhor mesmo é fazer frases curtas, em ordem direta. Outra sugestão interessante é dividir o texto em várias partes quando ele for muito longo. Para isso, vale usar títulos e subtítulos claros e diretos, criar quadros para destacar informações importantes, usar itens para listar coisas... Também é importante cuidar dos elementos não verbais: um bom texto precisa ter uma apresentação clara. Imagens, gráficos e esquemas visuais podem ser muito úteis para melhorar a experiência de quem lê. Nada de letrinhas miúdas ou de elementos que atrapalham a visualização do texto.

Essas técnicas (e várias outras) foram criadas a partir de pesquisas e aplicações que vêm sendo feitas há décadas, em várias partes do mundo. Estudos feitos por linguistas e psicólogos, com pessoas em situações reais, mostram que todos esses elementos influenciam na compreensão do texto. Os *designers*, por sua vez, têm mostrado que o jeito como distribuimos a informação

(no papel, na tela, em um mural etc.) também interfere na compreensão. Já os arquitetos contribuem ao provar que as pessoas conseguem processar melhor as informações em ambientes que trazem maior conforto visual e sonoro.

Esse jeito de produzir textos é muito importante para pessoas com dificuldade de leitura. Menciono, em primeiro lugar, as crianças, os adolescentes e os adultos com baixa escolaridade. Há também as pessoas com condições como autismo, déficit de atenção, dislexia, afasia, deficiências intelectuais, entre outras. Por fim, vale lembrar as pessoas que não têm o português como primeira língua, como alguns imigrantes, refugiados e indígenas. Para todas essas pessoas, um documento preparado dessa forma pode fazer toda a diferença. Além disso, a linguagem simples é útil para qualquer indivíduo que queira ter acesso a informações de interesse público. Ou seja, todos/as nós podemos nos beneficiar desses princípios.

Apesar disso, há muita resistência quando se fala do assunto. Para alguns, simplificar textos significaria abrir mão da exatidão dos conceitos. Para outros, teria como consequência diminuir a qualidade da informação e até mesmo da formação de profissionais. Contra esses argumentos, é preciso dizer de forma direta: simplificar um texto informativo não significa prejudicar sua qualidade. Mesmo as ideias mais profundas e complexas podem ser expressas de forma clara. Então, combater o “juridiquês”, o “mediquês” e o “burocratês” não significa deixar de reconhecer a importância dos conhecimentos específicos do Direito, da Medicina e da Gestão Pública. Em vez disso, a ideia é fazer com que as informações produzidas nessas áreas estejam acessíveis às pessoas que precisam delas. Basta usarmos as ferramentas e os recursos adequados a cada caso.

Os conhecimentos sobre LS ajudam a democratizar a informação e a cultura. Com isso, somamos esforços ao trabalho de outras pessoas que têm atuado na causa da acessibilidade: tradução para línguas de sinais, produção de materiais em braile, audiodescrição, construção de ambientes com desenho universal... Junto com a LS, essas iniciativas mostram que é, sim, possível, garantir

a todas as pessoas o direito à informação e cumprir o que manda a Constituição Federal.

E o que isso tudo tem a ver com a área cultural? Penso que, se queremos democratizar a cultura, precisamos refletir sobre a linguagem que usamos para nos comunicar com os fazedores de cultura e com o público.

Para ilustrar essa reflexão, quero trazer à memória alguns fatos importantes ocorridos nos anos 2000, no Ministério da Cultura (MinC). Naquela época, o MinC buscou ampliar a distribuição de recursos para fazedores de cultura que não encontravam espaço até então. Para isso, o Ministério ampliou o número de áreas contempladas nos editais de fomento e abriu espaços de participação dos cidadãos nas discussões sobre políticas culturais. Contudo, para entrar nesse universo, os fazedores de cultura precisavam (e ainda precisam) ler documentos normativos, seguir procedimentos, preencher formulários e planilhas, produzir relatórios... Para muitos deles, em especial os que vêm de contextos rurais e de comunidades indígenas e quilombolas, esses trâmites podem representar um enorme desafio.

Com isso, quero insistir na ideia de que não existe democratização efetiva sem a simplificação de textos e procedimentos. Sempre que um documento ou um rito for de difícil compreensão, em especial para as pessoas menos escolarizadas, provavelmente uma parcela da população deixará de exercer seus direitos.

O mesmo vale para os públicos da cultura. Se os museus, centros culturais e bibliotecas querem atrair mais pessoas, de todas as faixas etárias e classes sociais, precisam estar atentos ao uso da linguagem. Tudo ali precisa estar escrito e organizado de modo simples: as regras de uso, os catálogos, a sinalização do espaço... Se esses elementos dificultarem a vida do público (em vez de facilitar), a instituição deixa de cumprir seus propósitos. Por exemplo: se os textos explicativos que acompanham as obras de uma exposição não tiverem clareza, será que a instituição cumprirá bem sua função educativa? Afinal, muitos visitantes não conseguirão extrair as informações que buscam para apreciar melhor uma obra artística

ou um objeto antigo. Além da sensação de uma experiência incompleta, não compreender esses textos pode ser frustrante e afastar as pessoas da cultura, em vez de aproximá-las.

Além dos grupos já mencionados neste texto, penso que uma política de LS nas instituições culturais contribuiria para melhorar a experiência dos turistas estrangeiros. Uma pessoa que visita o Brasil sem saber português certamente terá dificuldades para entender as informações de um museu ou centro cultural se elas não forem fornecidas em uma língua que ele já conhece. Porém, se forem usadas técnicas de LS e se esse turista souber pelo menos um pouco de espanhol, italiano ou francês, as chances de ele entender a mensagem aumentam bastante. Para quem não sabe uma língua estrangeira, mas conhece línguas da mesma família, a vida fica bem mais fácil se os textos tiverem palavras mais comuns e frases mais curtas.

Infelizmente, a resistência ao uso da LS também acontece no universo cultural. Por exemplo, o sociólogo Pierre Bourdieu mostrou essa atitude em sua pesquisa mais famosa sobre hábitos culturais, em um livro chamado *A distinção*. Nessa obra, ele mostra que algumas pessoas com maior capital cultural, que costumam frequentar museus, veem com desconfiança as tentativas de tornar essas instituições mais acessíveis. Isso porque os hábitos culturais são, muitas vezes, usados como uma forma de distinção social. Assim, democratizar a arte pode significar, para alguns, o risco de perder privilégios (BOURDIEU, 2007, p. 215). Quem trabalha no setor cultural certamente conhece pessoas que pensam assim – se não forem elas mesmas a terem essa visão.

Se você chegou até aqui e se convenceu de que vale a pena conhecer mais sobre a LS, há muitos materiais disponíveis na internet. No site da Escola Nacional de Administração Pública (ENAP), estão disponíveis cursos *on-line* introdutórios sobre o tema. Além disso, muitas iniciativas de LS vêm se consolidando em todo o país, principalmente nos chamados “laboratórios de inovação” do setor público. A Rede Linguagem Simples Brasil, lançada em 2021, reúne

informações sobre muitas dessas iniciativas e também uma biblioteca com referências de vários tipos.

Gostaria de concluir essas reflexões trazendo um argumento forte que se costuma utilizar contra o uso da LS: “Em vez de sair simplificando tudo, nossa prioridade deveria ser ensinar as pessoas a ler e compreender textos mais complexos”. Ora, não há dúvida de que isso é importante. Melhorar a capacidade de leitura deveria ser prioridade em um país como o Brasil, que ainda tem níveis tão altos de analfabetismo funcional. É realmente preocupante que tão poucos de nós possamos ler e compreender textos que demandam maior esforço mental e mais conhecimento prévio. A educação exerce um papel fundamental no combate a esse problema e nisso estamos (ou deveríamos estar) todos de acordo.

Contudo, prefiro pensar que as duas estratégias não se anulam. Afinal, até mesmo países com altos índices de escolarização e bom desempenho de leitura têm apostado na simplificação. Em qualquer lugar do mundo, sempre haverá cidadãos que precisam de atenção específica para ter acesso à informação. Além do mais, uma política de educação que resolva o problema do analfabetismo funcional no Brasil demoraria anos, talvez décadas, para surtir efeito. Então, até que isso aconteça, precisaremos dar a todas as pessoas a informação que elas merecem: simples, acessível, sem rodeios. Este é o desafio, mas também um convite.

Palavras-chave: linguagem simples; campo cultural; democratização; acesso à informação; elitismo.

Referências

- BOURDIEU, Pierre. **A distinção**: crítica social do julgamento. Trad. Daniela Kern e Guilherme J. F. Teixeira. São Paulo: Edusp; Porto Alegre: Zouk, 2007.
- BRASIL. Câmara dos Deputados. **Projeto de Lei n. 6256 de 2019**. Institui a Política Nacional de Linguagem Simples nos órgãos e entidades da administração pública direta e indireta de todos os entes federativos. Brasília: Câmara dos Deputados, 2019. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2373593.
- PREFEITURA DE SÃO PAULO. **Linguagem Simples no Setor Público**. Apostila de curso sobre LS oferecido pela Prefeitura de São Paulo. São Paulo: Prefeitura de São Paulo, 2020.
- ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA. Disponível em: <http://www.enap.gov.br>.
- REDE LINGUAGEM SIMPLES BRASIL. Disponível em: <https://redelinguagensimpl.editorx.io/rede-linguagem-br>.

MAPEAMENTO CONCEITUAL DO DOMÍNIO DA INFORMAÇÃO TECNOLÓGICA A PARTIR DE CORPUS LINGUÍSTICO

MARIA CAROLINA CORADINI
EDER CAVALCANTI COIMBRA
GABRIELA WICK-PEDRO

*Instituto Brasileiro de Informação
em Ciência e Tecnologia*

MILENA AMBROSIO TELLES
*Empresa Brasileira de
Pesquisa Agropecuária*

ISAAC SOUZA DE MIRANDA JUNIOR
Universidade Federal de São Carlos

A relação entre língua e tecnologia tem sido alvo de muitos estudos e discussões desde o início do século XX, com a chegada dos primeiros computadores. O avanço tecnológico e o advento da internet, nas últimas décadas, resultaram na difusão de um ciberespaço em que se pode encontrar informações e soluções a respeito de quase tudo. A automatização de tarefas e a manipulação de grandes volumes de informação é uma realidade bastante presente nos dias de hoje, que só é possível pela melhoria contínua da chamada “interface homem-máquina”, ou seja, o modo e os meios pelos quais o ser humano acessa as informações no computador ou outros dispositivos. Essas melhorias vêm sendo desenvolvidas por especialistas que atuam no Processamento de Linguagem Natural (PLN), uma área de grande importância em tempos de Inteligência Artificial.

Na conferência de abertura do II Seminário Viagens da Língua, promovido pelo Centro de Referência do Museu da Língua Portuguesa, discutiu-se a respeito do que tem sido o mundo nesta era tecnológica, com debates sobre ferramentas de Inteligência Artificial, suas questões éticas e aplicações práticas. É uma área de muito interesse para linguistas, cientistas da computação, da informação e estudiosos de diferentes áreas, tanto pelas possibilidades que tem a oferecer quanto pela própria experiência como usuários de ferramentas que facilitam a realização de tarefas diversas.

O mapeamento conceitual que será apresentado neste artigo é uma das possibilidades de ferramenta de organização do conhecimento obtida a partir de uma metodologia de PLN, focada em elaboração de produtos terminológicos. O que se verá a seguir é uma breve síntese do que foi exposto na terceira apresentação da mesa-redonda “Caminhos da Linguística Computacional”.

O PLN é uma área em que linguistas e cientistas da computação unem esforços para aprimorar e construir sistemas e ferramentas computacionais, com o objetivo de facilitar a execução de atividades cotidianas desempenhadas por humanos. O processamento de uma língua natural é o que permite que sistemas computacionais diversos possam gerar e compreender comandos em uma língua humana, como português, ou inglês, por exemplo. É o que está por trás de ferramentas como tradutores automáticos, corretores automáticos de texto, sumarizadores, mapeamentos conceituais etc.

Assim como no ambiente físico, no ambiente virtual há muito conhecimento produzido em língua escrita. Se, por um lado, o avanço tecnológico facilita o acesso à informação, de modo geral, por outro, implica um volume ainda maior de informação produzida diariamente, que não está totalmente organizada e sobre a qual não se tem tanto controle. É nesse contexto, especificamente, que surgem estratégias de organização e de recuperação da informação, a partir de ferramentas desenvolvidas por metodologias de PLN, como mapeamentos conceituais.

Mapeamentos conceituais são sistemas estruturados de termos inter-relacionados semanticamente (pelo seu significado).

Em outras palavras, são como paisagens que permitem visualizar como determinadas áreas do conhecimento se organizam, a partir das relações entre seus termos mais recorrentes. São, portanto, uma ferramenta que permite a leitura de tendências e de subtemas de determinada área, a observação da expressividade de uma área em comparação com outra (por meio da quantidade de termos e da frequência de uso de cada um), e que possibilita, inclusive, análises estratégicas que podem contribuir para a tomada de decisões até em nível institucional.

As vantagens dos mapeamentos conceituais foram justamente o que motivou o projeto que ora se apresenta, a respeito do domínio da Informação Tecnológica no Brasil. O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) tinha como um dos objetivos do projeto Informação Tecnológica no Brasil (ITBr)¹, de construção de um ecossistema de informação em ciência, inovação, tecnologia e sustentabilidade, obter um mapeamento que demonstrasse a organização da área da Informação Tecnológica e suas principais tendências no Brasil, com base na literatura científica produzida nos últimos cinco anos em língua portuguesa.

Assim como outras ferramentas desenvolvidas a partir de metodologias de PLN, os mapeamentos conceituais podem ser construídos com base em uma grande quantidade de textos, que servirá de subsídio para a extração de palavras, de relações semânticas e demais informações conceituais da área investigada. Por serem a principal fonte de extração de informação para a construção do mapa, esses textos devem ser muito bem selecionados e tratados para dar seguimento às etapas da metodologia utilizada.

A construção de um mapeamento conceitual baseado em *corpus* linguístico especializado – coleção de textos de área específica do conhecimento – é realizada em conjunto com especialistas do tema, que participam e orientam decisões do projeto, e validam cada etapa do delineamento do mapa. Nesse sentido, a primeira participação efetiva dos especialistas da própria instituição consistiu em definir as fontes de extração dos textos e as expressões

de busca utilizadas para recuperá-los, a fim de refinar ao máximo o material utilizado como fonte de obtenção dos dados.

Por recomendação dos especialistas, os textos foram coletados de forma automática d'O Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto (OASIS.BR)². Alguns dos termos definidos para a busca foram: inovação, patentes, normatização, inteligência competitiva, desenvolvimento tecnológico, transferência de tecnologia, gestão da inovação, normalização, arranjos produtivos locais etc.

A coleta dos textos foi feita por meio de *scripts* (códigos) desenvolvidos em linguagem de programação especialmente para essa atividade, considerando todas as restrições da plataforma e os filtros determinados para a própria recuperação dos textos, como idioma (português), período de publicação (2018-2022) e disponibilidade de *download*. Após coletados, os textos foram analisados e convertidos em um formato legível por máquina (.txt), totalizando uma coleção de 771 textos, com 6.683.721 palavras.

A partir desse *corpus*, foram extraídas as palavras mais recorrentes, por meio de um programa de leitura e processamento de textos, o Antconc (ANTHONY, 2020). Apenas palavras com frequência de ocorrência acima de 67 foram extraídas, um corte de frequência definido a partir do próprio número de palavras do *corpus*. Com isso, foram obtidas 27.422 palavras, consideradas como *candidatas a termos*, isto é, que poderiam indicar conceitos da área estudada. Essas palavras foram filtradas e posteriormente enviadas para uma validação do grupo de especialistas, com a orientação para selecionarem apenas as palavras que refletissem conceitos da área.

Os termos validados por ao menos 80% da equipe de especialistas totalizaram 183 expressões, quantia considerada para a etapa seguinte, de análise semântica e proposta de classificação inicial. A classificação semântica dos termos é justamente o que permite organizar uma área do conhecimento, ou seja, parte essencial na elaboração de um mapa conceitual.

Os termos da Informação Tecnológica foram avaliados, em um primeiro momento, quanto a possíveis categorias temáticas de agrupamento. As categorias temáticas foram identificadas a partir dos próprios termos validados, especialmente aqueles compostos por mais de uma palavra, em que geralmente ocorre especificação na palavra final do termo, por exemplo, “gestão da **tecnologia**”, “proposta de **inovação**” etc. Após a identificação dessas categorias, formaram-se conjuntos temáticos. Cada termo foi avaliado quanto à sua relação com o termo central de seu respectivo conjunto. Também foram avaliados quanto à relação com outros termos do mesmo conjunto, especialmente em casos de sinonímia, hiperonímia e hiponímia. As relações semânticas utilizadas para relacionar os termos entre si foram selecionadas da Agrontology³, um conjunto de relações semânticas utilizadas no vocabulário controlado Agrovoc⁴ (FAO/ONU).

A proposta de classificação dos termos validados foi organizada em uma planilha de triplas (termo-relação-termo) para a produção das representações gráficas em *softwares* especializados, para dar uma forma visual aos dados obtidos do *corpus*. Os *softwares* de representação gráfica utilizados para tanto foram o yEd e o XMind, ambos de acesso gratuito.

As categorias temáticas identificadas nesses termos e consideradas para o delineamento do mapa conceitual foram: Indústria; Produção; Propriedade Intelectual; Patentes; Inovação; Tecnologia; Gestão da Informação e do Conhecimento; Empreendedorismo.

Os conjuntos com maior quantidade de termos são ressaltados nas representações apresentadas, com uma grande quantidade de arestas. Em cada uma das arestas há a relação semântica entre o termo e o conjunto ao qual pertence. Pela quantidade de arestas relacionadas a um mesmo conjunto, infere-se que as áreas de maior expressividade no *corpus* analisado são Inovação, Patentes e Tecnologia, sendo o primeiro o maior de todos.

A distribuição dos termos validados é proposta da seguinte forma:

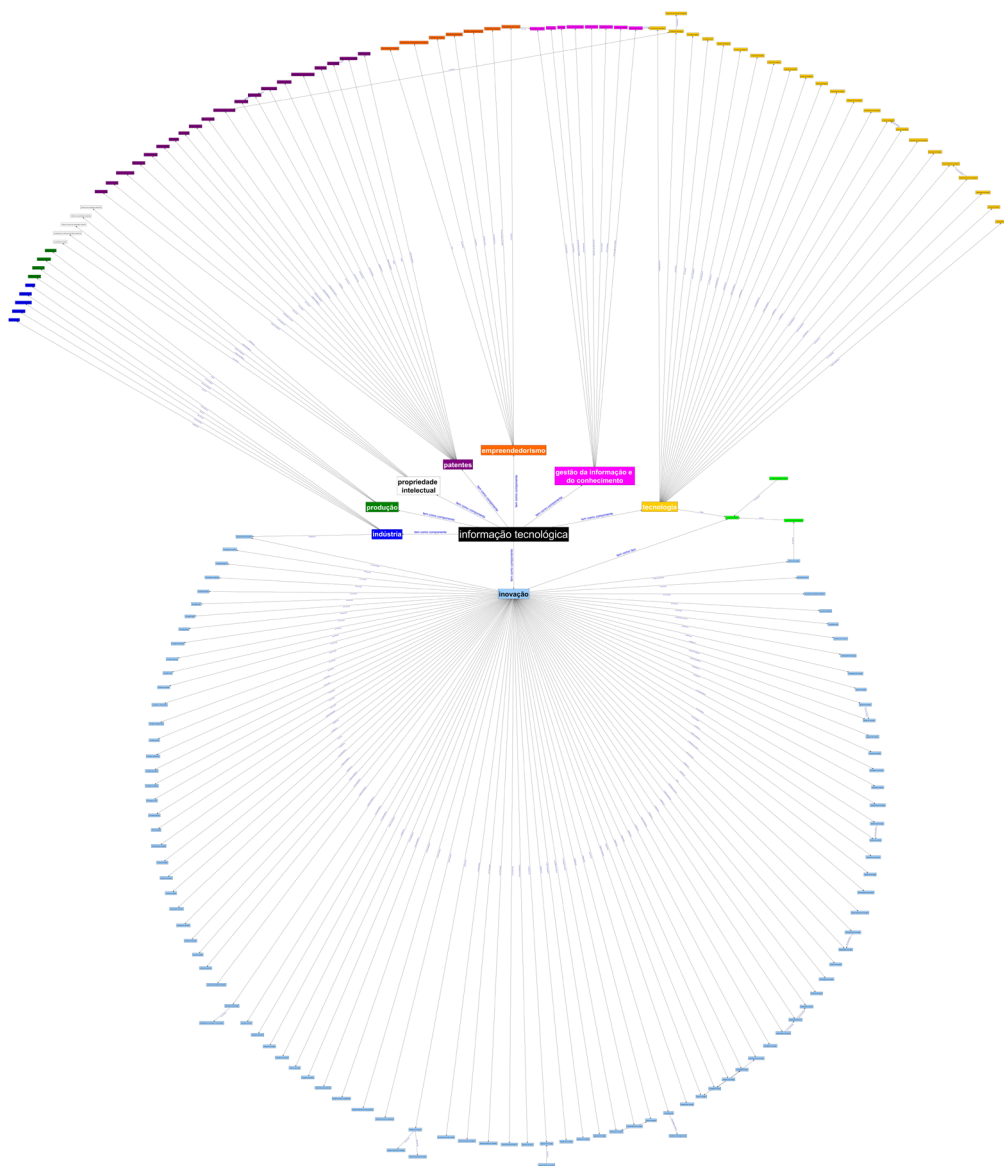


Figura 1: Representação gráfica do mapeamento conceitual da Informação Tecnológica desenvolvida no yEd.

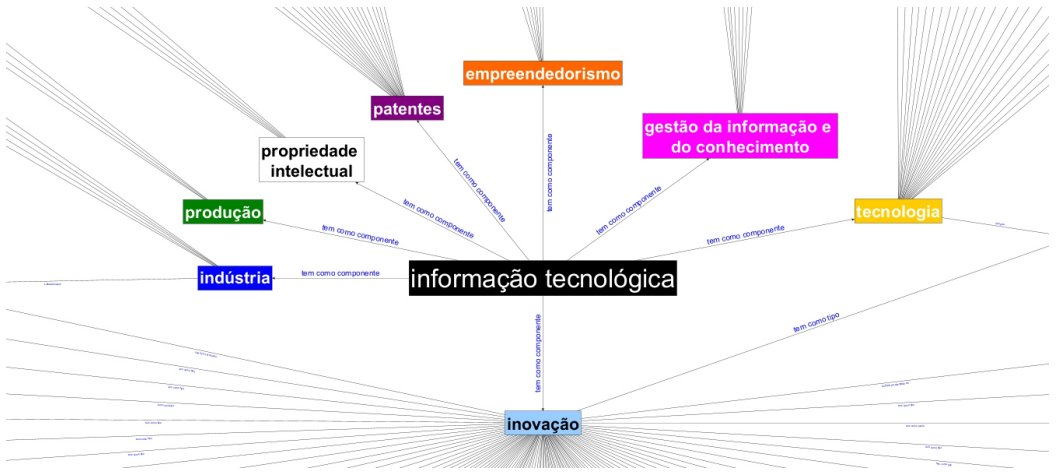


Figura 2: Visualização aproximada da representação gráfica do mapeamento conceitual da Informação Tecnológica (yEd).

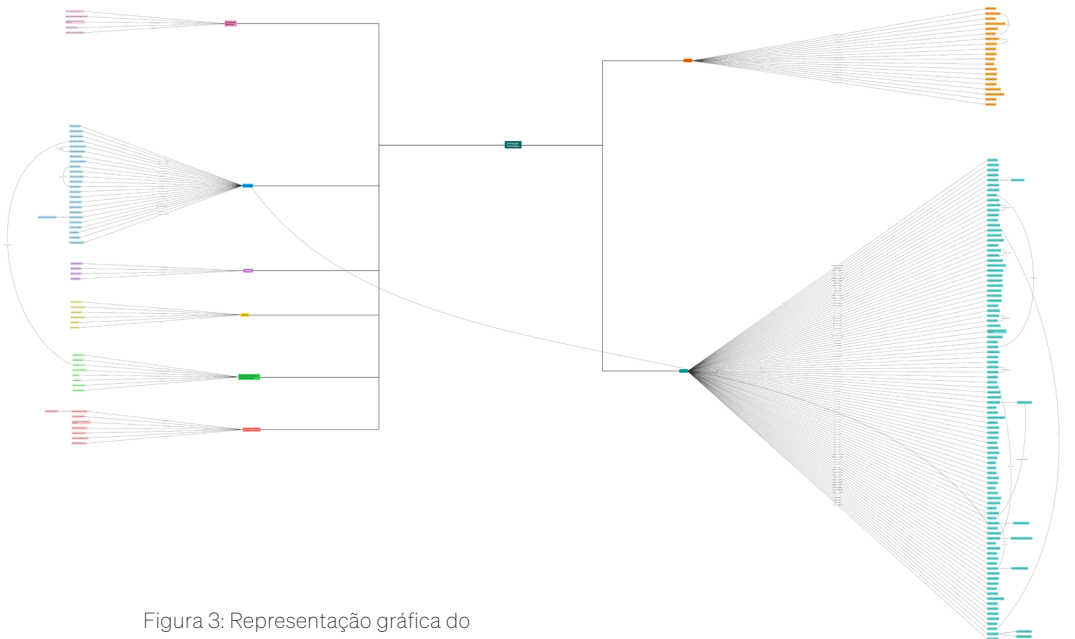


Figura 3: Representação gráfica do mapeamento conceitual da Informação Tecnológica (XMind).

A complexidade desse tipo de trabalho, que é baseado em um grande volume de informação linguística, justifica o fato de as validações da equipe de especialistas serem tão necessárias, norteadoras para cada inserção ou modificação que é proposta a respeito da organização dos termos da área, que, futuramente, serão lidos como uma espécie de retrato desse conhecimento já produzido.

Os resultados alcançados no mapeamento conceitual apresentado, além de servirem como uma ferramenta de recuperação da informação produzida especificamente nessa área do conhecimento gerado em língua portuguesa, podem servir para alimentar ou construir sistemas de arquitetura da informação, ferramentas linguísticas e glossários da área.

Apresentar esta aplicação do PLN no contexto da gestão pública no II Seminário Viagens da Língua foi enriquecedor, pois foi uma iniciativa de extrema importância para a divulgação da aplicação da Linguística no dia a dia das pessoas. Com essa iniciativa, o Museu colabora para o letramento da população em questões que envolvem a competência informacional, ou seja, a capacidade de lidar com grandes volumes de informação, saber como estão organizadas e como compreender, interpretar e fazer uso dessas informações de forma consciente e responsável.

Palavras-chave: mapeamento conceitual; Processamento de Linguagem Natural; organização da informação; Informação Tecnológica.

Referências

ANTHONY, L. **AntConc** (Versão 4.2.0) [**Software** de computador]. Tokyo, Japão: Waseda University. Disponível em: <https://www.laurenceanthony.net/software>.

UNITED NATIONS – Food and Agriculture Organization. **Agrontology**. Disponível em: <https://www.fao.org/agrovoc/agrontology>.

UNITED NATIONS – Food and Agriculture Organization. **Agrovoc**. Disponível em: <https://agrovoc.fao.org/browse/agrovoc/en/>.

WORKS. **yEd Graph Editor**, v. 3.21.1, © 2000-2021. Disponível em: www.yworks.com.

XMIND LTD. **XMind – Mind Mapping App**, © 2006-2024, Xmind Ltd. Disponível em: <https://xmind.app/>.

Notas

[1] Este projeto foi desenvolvido no âmbito do Grupo de Pesquisa ESFERA: Ecossistema de Informação em Ciência, Tecnologia, Inovação e Sustentabilidade.

[2] Disponível em: <https://oasisbr.ibict.br/vufind/>.

[3] Disponível em: <https://www.fao.org/agrovoc/agrontology>.

[4] Disponível em: <https://agrovoc.fao.org/browse/agrovoc/en/>.

TRADUÇÃO E TECNOLOGIA PARA AS LÍNGUAS DO BRASIL: O TRABALHO DO GRUPO DE TRADUÇÃO DO PAR SATERÉ- -PORTUGUÊS

ADRIANO CLAYTON DA SILVA
OSINARA DA PAZ BATISTA
Universidade Federal do Amazonas

Segundo dados do “Censo 2022 Indígenas: primeiro resultados”, do IBGE, atualmente há cerca de 1,69 milhão de indígenas no território brasileiro. É um aumento substancial em relação ao Censo de 2010, quando a população era de menos de 900 mil pessoas. Segundo o próprio IBGE, o aumento decorre de uma mudança de metodologia, em que também são considerados indígenas aqueles que se declaram como tal, independentemente de onde morem.

Mas isso também indica um processo mais profundo na sociedade: as pessoas dos povos originários já não temem tanto se identificarem como indígenas. Mais do que isso, essas pessoas estão ocupando espaços, especialmente nas mídias digitais, para afirmarem suas culturas e mostrarem o multiverso das populações indígenas brasileiras¹. Ainda, em 2022 tivemos a criação do

Ministério dos Povos Indígenas pelo atual Governo Federal, um passo importantíssimo na defesa dos direitos dos povos originários.

Por outro lado, seja na administração e ações do Ministério recém-criado, seja na grande maioria das páginas de internet e perfis de redes sociais, seja ainda nas associações e articulações sociopolíticas dos povos para lidarem com o mundo “branco”², o português é a língua franca para comunicação. As línguas maternas ficam restritas a demonstrações de sua existência, ocasiões em que quase sempre o português aparece, novamente, como língua de tradução e explicação para os termos indígenas. Isso é perfeitamente compreensível, haja vista a grande diversidade de línguas e a dificuldade que seria cada povo tentando falar e ser compreendido em seu próprio idioma.

Contudo, ao mesmo tempo as línguas indígenas vão sofrendo um processo de desuso que pode levar até à extinção. Das 274 línguas indígenas contabilizadas pelo Censo em 2010, 172 estavam ameaçadas de desaparecer na próxima geração, devido ao baixíssimo número de falantes e a outros fatores como vulnerabilidade social, pressão econômica, falta de políticas linguísticas, colonialismo religioso etc.

Não apenas a identidade se perde com a extinção de uma língua: as culturas indígenas que vivem na Amazônia carregam conhecimentos que têm se mostrado essenciais à humanidade. O exemplo mais óbvio é o uso das diversas plantas da floresta para os mais diversos fins, como o tratamento e a cura de doenças ou a alimentação humana e de animais domésticos. Para além disso, as populações indígenas conhecem formas de relacionamentos sicionaturais que muito podem contribuir nos atuais debates em torno da crise climática e da sustentabilidade. Igualmente importante é o que os povos originários perdem: a possibilidade de falarem por si próprios para expor suas preocupações e demandas sobre suas existências; a exigência de cumprimento de seus direitos garantidos em lei; e a possibilidade de contribuir nos debates internacionais para que novas visões de mundo, menos voltadas à

acumulação de bens e exploração desenfreada de recursos naturais, possam ser pensadas.

Em um país em que políticas linguísticas para minorias ainda é algo muito incipiente, e considerando que o português vai continuar sendo a língua de articulação dos povos indígenas por muito tempo, a tradução surge como uma possibilidade de comunicação, preservação e até ampliação das línguas indígenas ameaçadas³. E com as ferramentas tecnológicas da área, é possível pensar em como utilizá-las para acelerar as traduções das línguas indígenas para o português e vice-versa.

Por incrível que pareça, praticamente não há pesquisas que tentem incorporar tecnologias na tradução de línguas indígenas brasileiras. Em São Paulo há o trabalho do Centro de Inteligência Artificial (C4AI) da USP, com a população Guarani Mbya, originária da Mata Atlântica. Mas na vasta Amazônia, a pesquisa que deu origem a este texto é única, aparentemente.

Desse modo, a pesquisa conduzida pelo Núcleo de Estudos de Linguagens da Amazônia (NEL-Amazônia), da Universidade Federal do Amazonas (UFAM), do qual fazem parte autor e autora, visa investigar como algumas tecnologias de tradução (TT) podem ser utilizadas para ajudar a melhorar a comunicação na língua sateré-mawé (ou apenas sateré), do povo de mesmo nome, bem como revitalizar o idioma para as novas gerações.

O povo Sateré-Mawé vive no leste do estado do Amazonas, essencialmente nas cidades de Barreirinha, Maués e Parintins. No Censo de 2010, a população era de 13.210 pessoas, das quais 9.052 falavam apenas o sateré, o que na época o classificava como um idioma ameaçado. Em 2022, conversando com professores Sateré ligados à Secretaria Municipal de Educação de Parintins, descobrimos que era vontade do povo traduzir o projeto político pedagógico (PPP) indígena para sua língua materna. Contudo, eles não sabiam como fazer isso, além de que muitas palavras e expressões da esfera educacional existentes em português não existem em sateré. Daí surgiu a ideia desta pesquisa, que também é um projeto de extensão, ao construir com a comunidade originária novos conhecimentos

sobre tradução e língua. O título da pesquisa é “Processos tradutórios e ampliação de línguas ameaçadas: estudo de caso da Língua Sateré”, coordenada por mim, tem a autora deste texto como membro do grupo de tradução e é financiada pela Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM).

As TT mais atuais, chamadas Neural Machine Translation (NMT), usam vastos bancos de dados de textos traduzidos entre duas línguas (da ordem de dezenas de milhares ou mais), como documentos oficiais, romances, manuais, relatórios técnicos, palestras etc., como fonte para compararem palavras e expressões entre duas línguas e preverem possíveis novas traduções, que na atualidade têm se mostrado bem coerentes e confiáveis. Quando nos voltamos para as línguas indígenas, porém, não há bancos de dados tão vastos assim. A quantidade de textos traduzidos entre praticamente qualquer uma das línguas amazônicas e o português não chega a meia dúzia, sendo quase sempre algumas cartilhas sobre a língua indígena e alguns livros do Novo Testamento da Bíblia. Isso é insuficiente para um programa com NMT trabalhar. Assim, na presente pesquisa foi preciso dar um passo atrás e buscar outras tecnologias de tradução mais antigas para trabalhar com o sateré.

Uma dessas tecnologias é a tradução assistida por computador (CAT Tools, em inglês), pelo qual um *software* ajuda um(a) tradutor(a) humano(a) a acelerar uma tradução, seja permitindo rápido acesso a dicionários específicos, seja mostrando se já houve a tradução de determinada palavra ou expressão no texto. Essa tecnologia também funciona com bancos de dados de palavras e expressões traduzidas, as chamadas memórias de tradução, e quanto maior o banco, mais rápido o acesso a informações. Como a língua sateré também sofre da falta de textos traduzidos, precisamos dar mais um passo atrás e criamos um pequeno grupo de pessoas bilíngues para traduzir textos do português escrito para o sateré escrito. Vamos falar do trabalho desse grupo.

Primeiramente, é preciso entender o contexto amazônico em que vivem os Sateré-Mawé. Os núcleos urbanos são distantes, não há estradas nem sinal de internet ou celular em toda parte, como

acontece em quase todo o lado oriental do Brasil. Deslocar-se entre cidades, vilas e aldeias exige longas viagens em barcos grandes ou pequenos (raramente confortáveis) durante algumas ou várias horas, e isso quando há barcos disponíveis, pois a maioria dos destinos só tem um horário de barco ou lancha por dia. Igualmente difícil é a eletricidade e a internet nas vilas e aldeias dos “interiores” amazônicos. Em muitos casos, a eletricidade fica disponível por apenas algumas horas por dia, e o sinal de internet depende totalmente tanto da eletricidade, que mantém o sinal de *wi-fi* ligado (já que não há antenas de celular no meio da mata), quanto de alguém (órgão público ou indivíduo) ter internet via satélite para fornecer às pessoas próximas.

Ao mesmo tempo, dadas as condições de vulnerabilidade social dos Sateré-Mawé em geral, e consequentes baixa escolarização, letramento digital e acesso a um ponto de internet, não é fácil encontrar uma pessoa Sateré que saiba escrever nos dois idiomas. Foi preciso muitas conversas e contatos com vários membros do povo para encontrar um pequeno grupo de indígenas, morando nos mais diversos pontos da região, com conhecimentos e condições estruturais para trabalhar no projeto de tradução.

Em meados de 2023 começaram finalmente os trabalhos com três indígenas bilíngues: Osinara, moradora de Juruti, no Pará, e dois professores do Ensino Fundamental, ambos moradores de Parintins, mas que passam a maior parte do tempo ensinando nas vilas distantes do núcleo urbano do município. Assim, tudo é articulado através do aplicativo WhatsApp: quais textos serão traduzidos, quem fica com qual parte, quais as dúvidas gerais e particulares, quais os prazos a cada mês, como deve ser a entrega dos textos etc. Todos os textos sugeridos para tradução têm relação com temas pertinentes ao povo Sateré e à região em que vivem, para despertar mais interesse do grupo e ajudar a aumentar o conhecimento sobre tais temas. Além disso, a ideia é que tais textos sejam posteriormente publicados em um *blog* e disponibilizados às pessoas que queiram ler e opinar sobre eles.

Nos primeiros trabalhos tudo era novidade e ansiedade, com as pessoas “tímidas” e sem saber exatamente o que fazer. Aos poucos, o coordenador Adriano foi explicando como cumprir a tarefa da tradução, ao mesmo tempo em que deixou cada pessoa livre para fazer o trabalho da forma que entendesse melhor e no ritmo que conseguisse. Vários têm sido os desafios desde então, todos inerentes a esse tipo de trabalho: aprendizados de palavras e expressões e busca por equivalentes na outra língua; aprendizado de novos gêneros textuais; padronização de palavras; técnicas de tradução; organização e formatação de textos etc.

Felizmente, até agora o processo está funcionando bem e já temos onze textos traduzidos, entre notícias de jornal, artigos científicos, contos e trechos da Constituição Federal de 1988. Novas pessoas entraram no grupo e agora temos quatro tradutores e tradutoras ativos, bem como o apoio da Associação dos Kapi e das Lideranças Tradicionais do Povo Sateré-Mawé (KAPI), entidade reconhecida pelos Sateré e interessada em divulgar a língua materna para além da comunidade indígena, bem como em aumentar a quantidade de escritores bilíngues e situações em que a língua sateré possa ser usada. Ainda, os textos estão sendo armazenados em um banco de dados, junto com outros que já existiam e foram incorporados, de modo a poder futuramente alimentar o programa de CAT Tool e assim tentar acelerar ainda mais o processo tradutório.

Além do aumento da quantidade de textos traduzidos, algumas confirmações de fatos linguísticos e hipóteses de pesquisa foram feitas:

1. A polissemia também existe no sateré, com algumas palavras carregando vários sentidos, dependendo do contexto em que sejam usadas, como é o caso da palavra *waku*: pela quantidade de frases diferentes em que ela aparece, depreendemos que *waku* teria o sentido dos verbos ser e estar, mas com conotação positiva (diferente do ser/estar do português).

2. Os traços da cultura Sateré transparecem na língua, como a própria palavra *sateré*, que significa lagarta de fogo, mas também representa o nome do principal clã da população. Assim, sabemos que a organização sociopolítica do povo Sateré-Mawé inclui a divisão por clãs.
3. O uso do banco de dados ajuda de fato a acelerar traduções, mas a grafia das palavras precisa ser igual, ou a palavra não será encontrada. Isso é um problema, cuja solução deverá passar por alguma futura normatização das palavras.
4. Para que palavras e expressões equivalentes surjam entre as línguas, é preciso antes que contextos e gêneros textuais relacionados à palavra sejam apreendidos pela pessoa que traduz. Ou seja, a língua é afetada diretamente pela cultura e pelos contextos de uso.

Os detalhes dessas descobertas serão feitos em artigos futuros.

O trabalho vai continuar nos próximos meses, agora mais focado no aprendizado e uso da CAT Tool pelo grupo, de modo a aumentar a produção de textos traduzidos. Também esperamos concretizar a ideia do *blog*, para que outras pessoas falantes do sateré possam ler e dar suas opiniões sobre essas traduções, legitimando ainda mais os textos e deixando-os de fato com a cara do povo Sateré. A maior divulgação do sateré permitirá outros desdobramentos futuros, chegando até a um possível reconhecimento oficial da língua pelas administrações municipais locais. Por outro lado, a pesquisa poderá se desenvolver a ponto de permitir a construção de um sistema de tradução automática nos moldes do que é o Google Tradutor, não apenas para o sateré, mas para qualquer outra língua indígena ameaçada. Mas para isso ainda há muito o que ser feito, traduzido e descoberto.

Por fim, agradecemos à FAPEAM pelo apoio financeiro à pesquisa⁴, sem o qual ela seria inviável.

Palavras-chave: línguas indígenas; tradução; Sateré-Mawé; tecnologia.

Notas

- [1] Vide, por exemplo, os perfis @visibilidadeindigena, @apiboficial e @midiaindiaoficial no Instagram.
- [2] Como a Articulação dos Povos Indígenas do Brasil (APIB) ou a Coordenação das Organizações Indígenas da Amazônia Brasileira (COIAB).
- [3] É preciso uma nota de rodapé aqui para falar sobre a Constituição Federal traduzida para o nheengatu no ano passado e que ganhou repercussão por ser a primeira versão em língua indígena da Carta Magna brasileira. Cabe explicar que essa tradução foi uma iniciativa dos povos da Federação das Organizações Indígenas do Rio Negro (FOIRN), com sede em São Gabriel da Cachoeira, AM, cuja grande maioria da população é indígena de várias etnias e cuja Prefeitura Municipal tem quatro idiomas oficiais reconhecidos: nheengatu, tukano, baniwa e português. Quando essa tradução já estava bem avançada, foi preciso conversar com os órgãos públicos para sua oficialização, e só então ela ganhou visibilidade pelo Governo. Não foi algo financiado e promovido pelo poder público desde o início. Assim, não há previsão de novas versões indígenas da Constituição para os próximos anos.
- [4] Apoio pelo Edital n. 006/2022.



CAPIVARA: COMO RELACIONAR IMAGENS E TEXTOS EM PORTUGUÊS UTILIZANDO INTELIGÊNCIA ARTIFICIAL

JHESSICA SILVA¹, GABRIEL SANTOS¹, DIEGO MOREIRA¹,
ALEF FERREIRA², LUIZ PEREIRA¹, PEDRO BUENO¹,
THIAGO SOUSA², HELENA MAIA¹, NÁDIA DA SILVA²,
ESTHER COLOMBINI¹, HELIO PEDRINI¹, SANDRA AVILA¹

1. *Universidade Estadual de Campinas*

2. *Universidade Federal de Goiás*

Este texto tem como objetivo apresentar CAPIVARA (DOS SANTOS, 2023), um modelo de Inteligência Artificial produzido pelo grupo de Processamento de Linguagem Natural do Hub de Inteligência Artificial e Arquiteturas Cognitivas – H.IAAC, que visa relacionar imagens e textos em línguas de baixo recurso computacional, isto é, línguas que dispõem de poucos conjuntos de dados anotados prontos para uso e ferramentas computacionais que as suportem, sendo o maior foco na língua portuguesa. Além disso, este trabalho visa à proposta de método para a otimização de recursos computacionais e tempo de treinamento, o que contribui para a redução do impacto ambiental desse tipo de aplicação. O texto fornece algumas aplicações do modelo CAPIVARA, seu funcionamento e suas implicações para a sociedade. O artigo original de CAPIVARA pode ser encontrado pelo seguinte título: “CAPIVARA: Cost-Efficient Approach

for Improving Multilingual CLIP Performance on Low-Resource Languages”, apresentado na conferência EMNLP 2023 (The 2023 Conference on Empirical Methods in Natural Language Processing).

Normalmente, quando estamos falando de algum modelo de Inteligência Artificial (IA), e aqui com foco maior em modelos de aprendizado de máquina, usualmente dizemos que nosso modelo tem uma entrada e uma saída. A **entrada** são os dados que nós (usuários) fornecemos para o modelo e a **saída** é o resultado que o modelo nos fornece. Para exemplificar, suponha um *chatbot* (def. *assistente virtual que pode se comunicar por texto com usuários em tempo real*) que utilize IA para realizar atendimento ao cliente. Nesse caso, a entrada desse *chatbot* poderia ser um texto como “Olá, bom dia”, fornecido pelo usuário, e a saída poderia ser “Olá, como eu posso te ajudar?”, fornecido pelo *chatbot* como resposta ao usuário.

Então, passando agora ao contexto de **CAPIVARA**, o que temos como entrada é **um par texto e imagem**, e a saída é o **grau de similaridade** entre esse par. Para exemplificar, considere a imagem a seguir, em que é possível observar um cachorro brincando com uma bola.

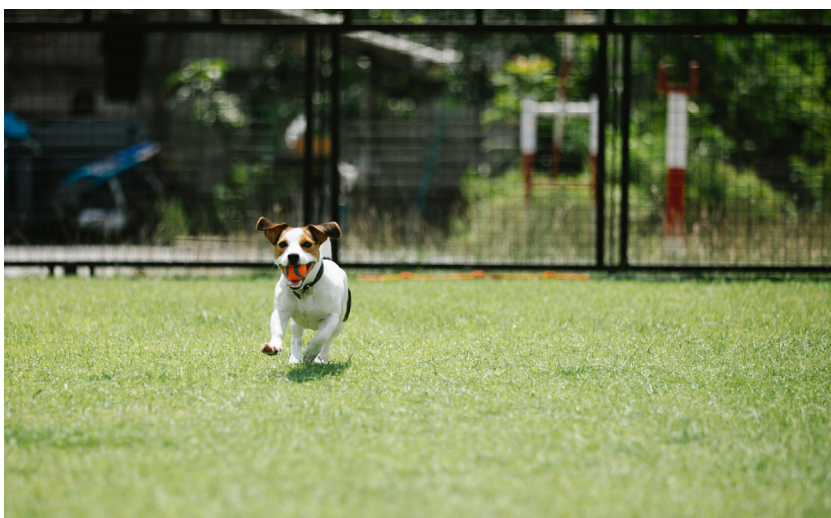


Imagem 1: Fotografia de um cachorro branco com a região das orelhas e dos olhos contornados de marrom, carregando em sua boca uma bola laranja com listras azuis. O cachorro corre sobre um gramado verde.

Seria possível fornecer essa imagem como entrada para o modelo juntamente de uma descrição (texto), como: “Cachorro branco com manchinhas marrons correndo em um gramado com uma bola laranja na boca em um dia ensolarado”, e como resposta o modelo forneceria o grau de similaridade entre esse texto e essa imagem. O valor 1 indica a maior similaridade (ou melhor correspondência entre texto e imagem) e o 0 indica que não existe similaridade entre o par texto e imagem. Então, no caso da descrição acima para a imagem apresentada, o modelo poderia dizer na saída que a similaridade é de 0.9 (isto é, quase 1), significando que esse texto e essa imagem possuem uma grande correspondência entre si, ou seja, o texto descreve adequadamente a imagem apresentada.

Agora, para a próxima imagem, a mesma descrição anterior “Cachorro branco com manchinhas marrons correndo em um gramado com uma bola laranja na boca em um dia ensolarado” faz com o que o modelo gere uma similaridade menor, como 0.7, pois embora tenha na imagem um cachorro branco com manchas marrons correndo com bola laranja na boca em um dia ensolarado, o cachorro da imagem não está correndo em um gramado. Temos um cachorro correndo na terra.

As aplicações de CAPIVARA são diversas e exemplos de tarefas nas quais se pode utilizar esse modelo são busca de textos a partir de imagens e/ou busca de imagens a partir de textos, podendo ser aplicado, por exemplo, em um mecanismo de busca de imagens como o Google Imagens.

O modelo que CAPIVARA implementa é conhecido como CLIP (RADFORD, A. et al., 2021), uma sigla para treinamento por contraste (comparação) de imagem e texto (em inglês, *contrastive language image pre-training*). Esse modelo foi proposto originalmente com o foco na língua inglesa, e depois surgiram propostas para outras línguas com alto recurso computacional, como o mandarim.

Até onde temos conhecimento, um aspecto que as múltiplas variações de CLIP têm em comum é o fato de que elas foram treinadas por empresas e/ou organizações com acesso a grandes infraestruturas computacionais e recursos financeiros. Porém,



Imagem 2:
Fotografia de um cachorro branco e marrom, com uma bola laranja na boca, correndo sobre um campo de areia.

quando trazemos essa tarefa para o cenário da academia, isso é um empecilho devido à dificuldade de acesso a computadores de alto desempenho, isto é, computadores equipados com GPU (placas de vídeo), compatíveis para o treinamento de modelos de IA desse porte. Para se ter uma noção mais clara, o CLIP em mandarim foi treinado por quase oito dias, utilizando 2048 GB de memória de GPU, enquanto a GPU a que nós tínhamos acesso para o treinamento de CAPIVARA tinha apenas 48 GB, com a possibilidade de treinamento de um modelo funcional em apenas duas horas.

Então, nossa proposta foi utilizar um CLIP multilíngue pré-treinado, ou seja, que já conhece pares de textos e imagens em algumas línguas, principalmente em inglês. Esse modelo é composto de dois blocos principais, um que aprende a extrair informações dos textos

e outro que aprende a extrair informações das imagens. Como esses blocos já foram amplamente pré-treinados, nós assumimos que o modelo já desempenhava bem sua função com relação às imagens. Porém, o bloco de processamento de texto ainda precisava de mais treinamento para lidar melhor com textos em português. Desse modo, o que fizemos foi treinar apenas esse último bloco, fornecendo exemplos de textos em português como entrada. Assim, o modelo levou 31 horas para alcançar um bom desempenho na nossa língua, utilizando em média 38 GB de memória da GPU durante todo o tempo de treinamento. Entretanto, isso ainda é muito recurso computacional, então nós otimizamos ainda mais esse treinamento.

Partindo novamente de um CLIP multilíngue pré-treinado, dessa vez nós utilizamos uma técnica que permite o treinamento de apenas uma parcela pequena do bloco de texto, a fim de ensinar CAPIVARA a entender o português. Com essa otimização, conseguimos treinar CAPIVARA em apenas duas horas, utilizando em média somente 8.5 GB de GPU.

Com isso, alcançamos também uma redução da pegada de carbono de CAPIVARA, pois os modelos de IA atualmente, por consumirem muitos recursos computacionais e levarem muito tempo para serem treinados, são modelos que geram um impacto ambiental muito grande. Principalmente pelo fato de a maioria deles serem treinados no Norte Global, onde a matriz energética não é tão sustentável. Então, ao ser treinado em apenas duas horas, em uma matriz energética como a do Brasil que é mais limpa, CAPIVARA gera pouquíssimas emissões de carbono quando comparado com outros modelos.

Outro ponto importante para o treinamento de um modelo CLIP é que isso demanda um grande conjunto de dados formado por pares de imagem e texto para que o modelo possa aprender a fazer as relações entre imagens e textos. Normalmente, esses conjuntos de dados são coletados a partir da internet e essa coleta e, principalmente, a preparação desses dados são muito custosas. Por conta disso, os conjuntos de dados prontos para uso de um treinamento de uma IA são, em sua maioria, em inglês. Isso é um problema

quando queremos propor tecnologias para nossa língua, pois a falta de dados prontos para uso em português dificulta o desenvolvimento de modelos que deem suporte à língua. Quando falamos sobre a necessidade de um grande conjunto de dados formado por muitos pares de imagem e texto, estamos nos referindo à escala de bilhões de pares de textos e imagens. Para exemplificar, se transformássemos todas as imagens que o modelo base de CAPIVARA viu durante o treinamento em um filme, esse filme teria 156 horas. Sendo que cada segundo desse filme teria sessenta imagens, ou seja, sessenta quadros por segundo, e isso significa um filme de alta qualidade.

Porém, nós não temos todo esse volume de dados prontos para treinar um modelo CLIP em português. Então, a solução proposta para CAPIVARA seguiu por dois caminhos diferentes. O primeiro foi considerar um conjunto de dados prontos para o inglês, e traduzir os textos para português utilizando o Google Tradutor, assim obtivemos um texto em português para cada imagem do conjunto que escolhemos. Um segundo caminho, para termos mais dados em português, foi empregar outra técnica de IA para gerar textos sintéticos para cada imagem do conjunto. Então, geramos dez textos sintéticos em inglês para cada imagem e posteriormente traduzimos para o português. Com esse processo, CAPIVARA foi treinado com um conjunto de dados na escala de 3 milhões de amostras anotadas em português.

Entretanto, esses textos foram originalmente produzidos em inglês, e isso carrega vieses que podem prejudicar o uso desses modelos pela comunidade falante de línguas de baixo recurso computacional. Por estarmos utilizando textos traduzidos, aspectos de nossa cultura, de nosso país e de regiões específicas não foram incluídos no modelo, uma vez que os dados coletados são produzidos majoritariamente em outros países. Então, CAPIVARA enfrentará dificuldades em tarefas que demandem esse conhecimento do contexto, por exemplo, reconhecer imagens de manifestações culturais como a fotografia de uma festa junina, de uma comida típica brasileira ou mesmo de monumentos históricos do nosso país. Infelizmente, CAPIVARA teria dificuldade em recuperar,

por exemplo, uma imagem do nosso Museu da Língua Portuguesa. Outro ponto é que CAPIVARA também carrega vieses do Norte Global, podendo, assim, acentuar alguns preconceitos contra populações marginalizadas e/ou contra aspectos religiosos e políticos. Porém, nós identificamos esses problemas e agora estamos trabalhando para atenuar seus impactos. Queremos tornar CAPIVARA um modelo de IA que saiba representar aspectos de nossa cultura e que seja mais justa e diversa para todos.

Por fim, um ponto importante a ser ressaltado é que CAPIVARA é um modelo de IA eficiente, de baixo custo computacional, que possui o foco na língua portuguesa. Aplicando a estratégia que propomos para a geração de dados para o português, é possível aplicar CAPIVARA em outras línguas com baixo recurso computacional.

Gostaríamos de agradecer ao Museu da Língua Portuguesa pela oportunidade de divulgarmos nosso trabalho neste Dossiê. Informações adicionais sobre CAPIVARA, tais como códigos, conjuntos de dados traduzidos para o português, o artigo científico, além de outros detalhes técnicos e considerações éticas podem ser acessados em <https://github.com/hiaac-nlp/CAPIVARA>.

Este projeto foi apoiado pelo Ministério da Ciência, Tecnologia e Inovação, com recursos da Lei n. 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Agentes inteligentes para plataformas móveis baseados em tecnologia de Arquitetura Cognitiva (processo 01245.013778/2020-21).

Palavras-chave: Inteligência Artificial; imagens; textos; IA para o português.

Referências

- DOS SANTOS, Gabriel Oliveira *et al.* **CAPIVARA**: Cost-Efficient Approach for Improving Multilingual CLIP Performance on Low-Resource Languages. MRL. Association for Computational Linguistics, 2023, p. 184-207.
- RADFORD, A. *et al.* **Learning Transferable Visual Models from Natural Language Supervision**. International Conference on Machine Learning, 2021, v. 189, p. 8748-8763.

TECNOLOGIAS PARA O REGISTRO DE LÍNGUAS DE SINAIS: UM PANORAMA HISTÓRICO

LUIS DOS SANTOS MIGUEL

FERNANDA HENRIQUES

Universidade Estadual Paulista

Uma das características da comunicação face a face é sua impermanência. Assim, para registrar e preservar qualquer tipo de discurso, é necessário lançar mão de tecnologias. Múltiplas descobertas e invenções permitiram que as línguas humanas fossem fixadas em determinados suportes e preservadas para a posteridade.

Na literatura científica pertinente ao campo da Linguística, há autores que afirmam que existem duas *modalidades* principais que caracterizam essas línguas de forma ampla – as *línguas orais* e as *línguas de sinais*.

Línguas orais são entendidas como aquelas que podem ser produzidas pelo aparelho fonador, emitidas na forma de uma sequência de sons. Portanto, podem ser percebidas por meio do sentido da audição e, conseqüentemente, são associadas a pessoas ouvintes.

Já as línguas de sinais são caracterizadas como aquelas que se utilizam de expressões faciais e corporais, desempenhadas no espaço em torno da pessoa que sinaliza. Em razão disso, tais línguas podem ser percebidas por meio do sentido da visão, podendo se constituir como a primeira língua de pessoas surdas.

Apesar de as línguas de sinais terem uma expressividade visual intrínseca, a maior parte dos registros linguísticos realizados no decorrer da história é constituída por representações das línguas orais. A escrita foi um dos recursos que permitiu aos textos serem registrados utilizando sistemas simbólicos convencionalizados. Essa prática se dá por meio de materiais e suportes variados, empregando processos específicos (escrita cursiva, caligrafia, tipografia, *lettering* etc.), gerando resultados bastante diversos no que concerne ao *design*. É relativamente recente a possibilidade da gravação de som, que permitiu também preservar os registros da voz humana, gerando assim novos tipos de textualidade.

Diante das constatações anteriores e partindo do princípio de que a comunicação humana é multimodal, colocamos as seguintes questões: quais foram os modos de representação utilizados para o registro de línguas de sinais, concretizados a partir de tecnologias e técnicas específicas, e quais as potencialidades oferecidas por cada um deles? Nesse sentido, esta pesquisa se propôs a buscar dados para promover uma reflexão em torno dessas indagações, tendo como ponto de partida a identificação de obras que contêm línguas de sinais registradas em algum nível.

Cabe salientar que não é possível reunir todas as produções realizadas ao longo da história que retratam línguas de sinais, pois, apesar de essas obras não serem tão numerosas quanto as produções em línguas orais, continuam sendo bastante expressivas em termos quantitativos.

Nossa metodologia de pesquisa incluiu a busca por menções na literatura acadêmica e a condução de uma investigação documental de obras físicas e digitais que se destacaram ao longo dos

anos, seja pelo uso de um recurso inovador ou por sua significância histórica para a comunidade surda.

Isso nos permitiu constituir um acervo de imagens que, no caso desta pesquisa, foi dividido em dois grandes grupos: (I) *produções internacionais*, elaboradas em países estrangeiros e (II) *produções nacionais*, elaboradas no Brasil. Para isso, foram adaptados preceitos metodológicos dos estudos de *memória gráfica*.

Após a coleta de dados, foi possível analisar os itens encontrados e determinar categorias para classificar os modos de representação descobertos nas obras. Nesta análise, foram identificadas duas categorias principais.

A primeira consiste em representações pictóricas – aquelas em que é possível atestar semelhanças entre o registro em si e uma pessoa sinalizando. Teoricamente, qualquer indivíduo que visualizasse um registro como esse poderia realizar o sinal, mesmo sem saber o significado dele. Dentro dessa categoria estão as representações estáticas (desenhos, fotografias e capturas instantâneas de modelos 3D) e as representações dinâmicas (animações 2D, vídeos e animações 3D).

A segunda corresponde às representações não pictóricas, que neste estudo foram referidas como sistemas de notação. Esses sistemas exigem o aprendizado de um determinado conjunto de caracteres e de suas regras de combinação e organização para que possam ser efetivamente decodificados por um leitor. Os símbolos criados para esse fim podem apresentar algum traço de iconicidade visual, ou seja, remeter a alguma característica visual da sinalização, ou ainda trazerem alguma associação com um alfabeto manual de uma língua de sinais específica. Isso facilita o reconhecimento e a memorização desses sistemas codificados, embora não sejam características obrigatórias para constituí-los.

O mais conhecido desses sistemas é o SignWriting, que é utilizado por comunidades surdas ao redor do mundo, dentro e fora do âmbito acadêmico. No Brasil, três outros sistemas – ELiS, SEL e Visografia – foram desenvolvidos por pesquisadores acadêmicos para representar a Língua de Sinais Brasileira (Libras).

Algumas considerações puderam ser feitas a partir da combinação da revisão de literatura com a análise dos acervos. Percebe-se que modos de representação estáticos como o desenho foram os primeiros a trazer registros das línguas de sinais. Eles persistiram sendo mais abundantes que outros modos por muito tempo, mesmo após a invenção de tecnologias para a gravação de filmes, que inicialmente tinham um custo bastante elevado. Apesar disso, eles se tornam inviáveis na medida em que se pretende representar enunciados mais longos em língua de sinais, uma vez que ocupam um espaço considerável para serem visualizados adequadamente. Além disso, há desafios inerentes para representar uma língua que se vale de movimentos em suportes estáticos. Diversas estratégias foram desenvolvidas para contornar essa situação, mais notadamente o uso de setas e a decomposição de um sinal em várias etapas.

Com o advento das câmeras digitais, essa tendência mudou significativamente. Além do custo reduzido, esses dispositivos tiveram uma operação facilitada por meio de ajustes automáticos, que não exigiam do usuário conhecimentos técnicos aprofundados para que pudessem utilizá-los. Nesse sentido, a possibilidade de usar computadores com *webcams* e, principalmente, *smartphones* com câmeras permitiu que a gravação e a transmissão de vídeo se tornassem uma prática comum entre a comunidade surda.

Assim, na contemporaneidade, o número de produções culturais em línguas de sinais cresce diariamente, em decorrência dessa democratização no acesso a dispositivos digitais, combinada com a ampliação do acesso à internet pela população mundial.

Os modos de representação não pictóricos têm uma baixa adesão em meio à comunidade surda como um todo. Diversos fatores podem influenciar para esse uso reduzido. Tais modos requerem um ensino específico, ao qual a maior parte da comunidade surda que se comunica em uma língua de sinais não teve acesso. Além disso, muitos desses sistemas geralmente apresentam um número elevado de caracteres, o que dificulta sua memorização plena. Soma-se a isso o fato de que esses caracteres não estão disponíveis em teclados virtuais de dispositivos utilizados para

comunicação rápida, tais como *smartphones* e *tablets*. Esses são alguns dos motivos que dificultam a aplicação prática desses sistemas em contextos cotidianos.

É preciso considerar que a sinalização em qualquer língua de sinais apresenta muitas sutilezas em relação a todos os seus parâmetros fonológicos e elementos simultâneos, o que traz uma complexidade indissociável para sua representação a partir de símbolos.

Por fim, pôde-se verificar que, com o desenvolvimento de tecnologias e sua subsequente popularização, novos aspectos das representações das línguas foram explorados, compondo novos gêneros do discurso. Ademais, diversos fatores contextuais, inclusive econômicos, sociais e políticos, relativos ao local de produção e ao período histórico em que essas obras foram concebidas, influenciaram no resultado delas.

Palavras-chave: *design*; língua de sinais; registro; Libras; tecnologia.

O QUE AS FERRAMENTAS DE PEDRA PODEM NOS CONTAR SOBRE O SURGIMENTO DA LINGUAGEM

VITOR NÓBREGA

Universidade de São Paulo¹

Pesquisadores de diversas disciplinas vêm há anos buscando no registro arqueológico informações acerca de quando, como e em qual espécie surgiram as primeiras habilidades linguísticas. O critério que guia essa busca geralmente envolve a identificação de práticas simbólicas na pré-história. A justificativa é a seguinte: a atribuição de significados a um artefato, ou seja, a produção de um símbolo material, deve envolver necessariamente linguagem. Seguindo esse raciocínio, seria de grande relevância identificar quando aparecem os primeiros vestígios de comportamento simbólico no passado humano, pois sua análise nos permitiria desvendar questões envolvendo a emergência da linguagem.

Ao investigar o registro arqueológico de nossos ancestrais, notamos que as primeiras expressões de um comportamento mediado por símbolos são bastante recentes, tornando-se sistemáticas

a partir de cem mil anos atrás, com os primeiros membros da espécie *Homo sapiens* (TATTERSALL, 2018; WADLEY, 2021). Exemplos são a confecção de ornamentos corporais, tal como o uso de conchas perfuradas como miçangas (HENSHILWOOD *et al.*, 2004), o emprego de utensílios para pigmentação (HENSHILWOOD *et al.*, 2018) e a produção de desenhos geométricos em cascas de ovos de avestruz e em placas de ocre (HENSHILWOOD *et al.*, 2002; TEXIER *et al.*, 2010); esse último exemplo ilustrado na figura 1.

Esse conjunto de objetos parece carregar características simbólicas, especialmente se levarmos em consideração o emprego que fazemos de seus correlatos modernamente. Logo, se a produção de símbolos e linguagem são faces de uma mesma moeda, deveríamos naturalmente admitir que tais artefatos servem como a principal janela para o passado linguístico humano, uma vez que representariam os primeiros vestígios materiais da expressão de significados da nossa espécie.

A cultura material pré-histórica também é composta de diversos objetos que não carregam traços tipicamente simbólicos, já que sua produção estaria voltada para atividades de subsistência, como a caça e a coleta. O principal exemplo são as ferramentas feitas de pedra, as quais compreendem os registros mais comuns e antigos



Figura 1: Placa de ocre contendo gravuras geométricas triangulares e retangulares, Blombos, África do Sul. Imagem cedida por cortesia ao autor pelo prof. dr. Francesco d'Errico, Universidade de Bordeaux/CNRS. Fotografia: d'Errico-Henshilwood.

do passado hominínio. As primeiras de que temos conhecimento datam de aproximadamente 3,3 milhões de anos e foram produzidas por espécies anteriores aos primeiros exemplares do gênero *Homo* (HARMAND et al., 2015), notadamente, os australopitecos, indivíduos fisicamente semelhantes aos chimpanzés, porém bípedes.

Dada a abundância, resiliência e a variedade das ferramentas de pedra ao longo da evolução humana, uma questão se coloca: seria possível abordar tais ferramentas também como uma janela para a história evolutiva da linguagem, apesar de elas carecerem de traços simbólicos? Essa questão é relevante na medida em que uma resposta positiva nos abriria caminho para avaliar se espécies anteriores aos *H. sapiens* também seriam dotadas de linguagem, isso porque não estaríamos mais circunscritos à conjectura de que somente artefatos pertencentes a uma cultura material simbólica – mais recentes – estariam sujeitos a inferências de natureza linguística.

Neste texto, destaco que uma resposta positiva a essa questão é possível. As ferramentas de pedra, apesar de não estarem diretamente vinculadas a símbolos, também podem auxiliar na reconstrução da história evolutiva da linguagem. Charles Darwin (1871), por exemplo, foi um dos primeiros a chamar a atenção para as correlações entre a produção de ferramentas líticas e a linguagem humana. O antropólogo Ralph Holloway (1969), anos depois, sugeriu que a sequência de ações motoras empregadas no lascamento de uma ferramenta de pedra envolveria uma sintaxe – isto é, um conjunto de regras ordenadas –, cuja organização seria semelhante à sintaxe empregada pelas línguas para a formação de sentenças. Essas conjecturas promoveram nas últimas décadas diferentes linhas de investigação sobre a correlação entre produção de ferramentas líticas e linguagem, as quais apresentarei brevemente a seguir.

Antes de explorar essas linhas de investigação, vale a pena recuperar o que temos de informações acerca das primeiras indústrias líticas pré-históricas.

Sabemos, hoje, que as habilidades tecnológicas humanas são bastante antigas, já que remontam a primatas não humanos, como os ancestrais dos chimpanzés modernos, que muito possivelmente

utilizavam pedras para abrir nozes, bem como gravetos para pescar cupins e formigas há pelo menos 7 ou 8 milhões de anos (BOESCH & BOESCH, 1990). Já as primeiras ferramentas líticas começaram a ser propriamente produzidas a partir de 3,3 milhões de anos, como apontado, obtendo sua primeira grande expressão com a indústria olduvaiense há cerca de 2,6 milhões de anos, a qual basicamente consistia no lascamento de pedras a fim de obter uma borda afiada (TOTH & SCHICK, 2018). Um exemplar de ferramenta olduvaiense pode ser observado no seixo lascado da figura 2.

Por volta de 1,76 milhão de anos, começa a aparecer no registro arqueológico uma forma de ferramenta lítica mais complexa do que a olduvaiense, algo que se dá em paralelo ao surgimento dos primeiros membros do gênero *Homo* (SHARON, 2010). Essas novas ferramentas compreendiam duas bordas afiadas, o que resultava em um artefato com um formato de “pera”, tal como ilustrado na figura 3. Essa indústria lítica ficou conhecida como acheulense e suas ferramentas são comumente chamadas de “machados de mão”.



Figura 2: Ferramenta lítica olduvaiense de Guelmim-Es Semara, região do Marrocos (Museu Arqueológico Nacional de Madri). José-Manuel Benito Álvarez. Fonte: Wikimedia Commons.

Das ferramentas acheulenses em diante, inúmeras outras indústrias líticas irão ser identificadas no registro arqueológico, cada uma delas envolvendo graus mais elevados de complexidade em sua produção. Essas indústrias atingirão um ápice de maior variedade estrutural em um menor intervalo de tempo a partir de 230 mil anos atrás, algo que, desde então, se acelerou vertiginosamente até os dias de hoje.

Com esse pano de fundo em mente, podemos retomar a pergunta principal deste texto: como avaliar a produção das ferramentas de pedra a fim de identificar linguagem? Há, pelo menos, duas linhas de investigação atualmente em andamento que buscam determinar como as diferenças e os graus de complexidade envolvidos nas tecnologias pré-históricas podem revelar a disponibilidade de linguagem. A primeira delas consiste em discernir o que fez com que cada tecnologia de lascamento fosse transmitida ao longo de gerações. Por exemplo, um indivíduo é capaz de aprender



Figura 3: Ferramenta lítica acheulense de Montières, Somme, França. Coleção Victor Comment.
Fonte: Wikimedia Commons.

a lascrar uma ferramenta olduvaiense apenas observando um lascador experiente? Ou será que instruções por meio de gestos ou comandos verbais são necessárias a fim de tornar o aprendizado concreto e eficaz?

Experimentos inicialmente apontaram que as ferramentas olduvaienses, como a da figura 2, poderiam ser aprendidas simplesmente através de observação e imitação (WYNN *et al.*, 2011). Logo, sem o intermédio de linguagem. Posteriormente, esses resultados foram revisitados, com novas observações que apontavam para a necessidade de alguma forma de instrução gestual ou verbal para facilitar o aprendizado das ferramentas olduvaienses. As ferramentas acheulenses, por sua vez, da figura 3, seriam mais intrincadas, demandando necessariamente instrução gestual ou verbal para serem plenamente aprendidas (MORGAN *et al.*, 2015).

Uma interpretação desses novos resultados é a de que indivíduos capazes de produzir apenas ferramentas olduvaienses seriam desprovidos de uma comunicação linguística – no caso, os australopitecos –, uma vez que comandos linguísticos não seriam obrigatórios para sua transmissão. Por outro lado, os lascadores de ferramentas acheulenses deveriam apresentar alguma forma de comunicação linguística, já que, sem comandos dessa natureza, sua transmissão seria ineficaz. Trabalhos mais recentes sugerem ainda que uma linguagem gestual seria muito mais eficiente na transmissão das bifaces acheulenses do que a linguagem verbal (CATALDO *et al.*, 2018), o que sugestivamente sinaliza que alguma forma de expressão linguística por gestos poderia estar disponível já desde os primeiros membros do gênero *Homo*.

A segunda linha de investigação recupera a correlação feita por Holloway, entre ação motora e sintaxe. Essa linha consiste em relacionar o plano de ações motoras necessárias para se produzir uma ferramenta de pedra e as regras sintáticas que subjazem à produção de sentenças nas línguas. Tanto a linguagem humana quanto ações motoras podem ser descritas como sistemas combinatórios que envolvem a organização de unidades elementares em um arranjo hierárquico, isto é, a relação dos elementos combinados não é

determinada por sua ordem linear, mas por conjuntos de unidades elementares encaixadas umas nas outras.

Para entender melhor o que vem a ser esse arranjo, consideremos as relações entre os nomes *livros*, *páginas* e *cor* em (1).

- (1) a. os **livros**
 b. as **páginas dos livros**
 c. a **cor das páginas dos livros**

A interação combinatorial entre esses nomes não é determinada por sua proximidade linear, ou seja, o fato de *livros* vir após *páginas* e de *páginas* vir após *cor*. Na verdade, a relação entre essas palavras envolve uma estrutura abstrata, como se cada nome fosse em si uma caixa, sobre a qual outras caixas são empilhadas. Podemos observar esse “empilhamento” nas relações de concordância, as quais envolvem troca de informações entre palavras. Se a relação entre esses nomes fosse mediada por proximidade linear, deveríamos esperar que a relação de concordância de número com o verbo *ser*, em (2), levasse em conta o nome mais próximo, notadamente, *livros*.

- (2) A cor das páginas [dos **livros**]_{PL} são_{PL} azuis_{PL}.

No entanto, a sentença em (2) não é aceitável no português. Sua correspondente bem formada envolve uma relação de dependência que se estabelece com um nome linearmente distante do verbo, *cor*, tal como se observa em (3).

- (3) A [**cor**]_{SG} das páginas dos livros é_{SG} azul_{SG}.

Esse exemplo nos mostra que, apesar de linearmente próximo, o nome *livros* está estruturalmente menos saliente do que o nome *cor*. Isso porque *livros* corresponderia a uma “caixa” muito mais distante em profundidade no empilhamento do que *cor*, como representado na figura 4.

Se for possível demonstrar que o lascamento de pedras para a produção de ferramentas envolve ações motoras organizadas por meio de uma estrutura, tal como vemos na organização da

sentença na figura 4, então poderíamos implicar que habilidades fundamentais para a sintaxe humana teriam evoluído em paralelo ou em decorrência das habilidades motoras atreladas ao lascamento. Isso é o que sugere, por exemplo, Moore (2010). O autor argumenta que esse tipo de combinação hierárquica surgiu por volta de 270 mil anos atrás, quando o estilo mousteriano de fabricação de ferramentas apareceu com os neandertais. Entretanto, relações hierárquicas mais rudimentares podem ter emergido já durante a fase acheuleana, como apontam Goucherel & Noûs (2020), o que sinalizaria a presença de uma protossintaxe em espécies hominíias que antecedem tanto os *H. sapiens* quanto os neandertais.

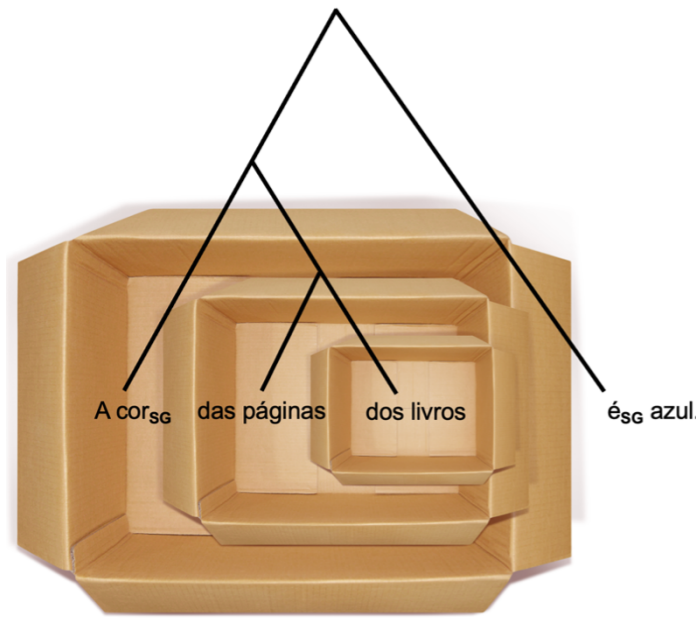


Figura 4: Representação arbórea parcial da sentença em (3).

Essas duas linhas de investigação ainda estão em processo de refinamento de seus métodos de reconstrução e análise das relações entre as ferramentas de pedra e a linguagem humana. Através delas, percebemos que é possível examinar o registro arqueológico de espécies anteriores a nossa sem necessariamente depender do isolamento de traços simbólicos para inferir linguagem. As ferramentas de pedra, portanto, nos contam que tanto seus modos de transmissão ao longo de gerações quanto a maneira como as ações motoras foram organizadas para seu lascamento são igualmente significativas como uma janela para se observar o passado linguístico humano.

Palavras-chave: linguística evolutiva; arqueologia; ferramentas líticas; sintaxe.

Referências

- BOESCH C.; BOESCH H. Tool use and tool making in wild chimpanzees. **Folia Primatologica**, v. 54, 1990, p. 86-99.
- CATALDO, D. M. *et al.* Speech, stone tool-making and the evolution of language. **PLoS ONE**, v. 13, n. 1, 2018, e0191071.
- DARWIN, C. **The descent of man, and selection in relation to sex.** Londres: John Murray, 1871.
- GAUCHEREL, C.; Noûs, C. Platforms of palaeolithic knappers reveal complex linguistic abilities. **PCI Archo**, 2020.
- HARMAND, S. *et al.* 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. **Nature**, v. 521, n. 7.552, 2015, p. 310-315.
- HENSHILWOOD C. S. *et al.* Emergence of modern human behavior: Middle Stone Age engravings from South Africa. **Science**, v. 295, 2002, p. 1278-1280.
- HENSHILWOOD C. S. *et al.* Middle Stone Age shell beads from South Africa. **Science**, v. 304, n. 404, 2004.

- HENSHILWOOD C. S. *et al.* An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. **Nature**, v. 562, 2018, p. 115-118.
- HOLLOWAY, R. L. Culture: a human domain. **Current Anthropology**, v. 10, n. 4, 1969, p. 47-64.
- MOORE, M. W. Grammars of action' and stone flaking design space, em NOWELL, A.; DAVIDSON, I. (orgs.). **Stone tool and the evolution of human cognition**. Boulder, CO: University Press of Colorado, 2010, p. 13-43.
- MORGAN, T. J. H. *et al.* Experimental evidence for the co-evolution of hominin tool-making teaching and language. **Nature Communications**, v. 6, 2015, 7029.
- SHARON, G. Large flake Acheulian. **Quaternary International**, v. 223, 2010, p. 226-233.
- TATTERSALL I. An evolutionary framework for the acquisition of symbolic cognition by *Homo sapiens*. **Comparative cognition and behavior review**, 2018, p. 99-114.
- TEXIER J. P. *et al.* A Howiesons Poort tradition of engraving ostrich eggshell containers dated to 60,000 years ago at Diepkloof Rock Shelter, South Africa. **PNAS**, v. 107, n. 14, 2010, p. 6180-6185.
- TOTH, N.; SCHICK, K. An overview of the cognitive implications of the Oldowan Industrial Complex. **Azania: Archaeological Research in Africa**, v. 53, n. 1, 2018, p. 3-39.
- WADLEY L. What Stimulated Rapid, Cumulative Innovation After 100,000 Years Ago? **Journal of Archaeological Method and Theory**, v. 28, 2021, p. 120-141.
- WYNN, T. *et al.* An ape's view of the Oldowan revisited. **Evolutionary Anthropology**, 20, 2011, p. 181-197.

Notas

- [1] A presente pesquisa é financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 2023/03196-5.

GAMIFICAÇÃO DE CONTEÚDOS E JOGOS DIGITAIS PARA O ENSINO E O APRENDIZADO DA LÍNGUA PORTUGUESA E DA EDUCAÇÃO AMBIENTAL

MARIA DE OLIVEIRA RODRIGUES
Universidade Federal de Uberlândia

Participar do II Seminário Internacional Viagens da Língua: Língua e tecnologias, realizado no Museu da Língua Portuguesa, foi uma experiência enriquecedora. Permitiu que meu trabalho fosse apresentado e apreciado por um público constituído de pesquisadores, estudantes, não acadêmicos, falantes do português, um público diversificado, ávido por conhecer a língua portuguesa.

O conhecimento acadêmico aliado ao trabalho efetivo do professor nas salas de aula pode trazer resultados satisfatórios aos estudantes e a toda equipe escolar. Diante desse entendimento, em 2021, iniciei o doutorado acadêmico em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU), a fim de contribuir com o ensino, a pesquisa e diretamente com os discentes.

Nessa perspectiva, de modo a auxiliar os estudantes para a assimilação dos conteúdos previstos pela Base Nacional Comum

Curricular (BNCC), que regulamenta o ensino da língua portuguesa nas instituições escolares, e contribuir com estudos futuros, iniciei a pesquisa de doutorado intitulada *A concordância verbal de número em Belo Horizonte: um estudo sociolinguístico*, que visa aprofundar o estudo da presença e da ausência da marcação de número na flexão verbal no português de Belo Horizonte. Assim, a intenção é analisar e descrever esse fenômeno variável tanto da fala espontânea quanto em textos escritos produzidos por belo-horizontinos. Em particular, a pesquisa analisa e descreve a concordância verbal no português de Belo Horizonte. Este estudo está fundamentado nos pressupostos teóricos-metodológicos da Sociolinguística Variacionista (HERZOG, LABOV, WEINREICH, 2006 [1968]; LABOV 2008, [1972]). A intenção é que a pesquisa possa contribuir significativamente para a compreensão do português falado e escrito na cidade de Belo Horizonte e trazer possíveis contribuições para a discussão sobre o português do Brasil de modo geral.

Esta pesquisa está aliada à linha Teoria, Descrição e Análise Linguística, e dedica-se ao estudo da língua portuguesa, que ocupa atualmente a nona colocação no *ranking* de idiomas mais utilizados no mundo, de acordo com os dados do Instituto de Investigação e Desenvolvimento em Política Linguística (IPOL). A língua portuguesa origina-se do latim, pertence à família das línguas indo-europeias e está representada em alguns continentes. Para chegar ao estágio atual de uso no Brasil, ela passou por diversas modificações e influências de vários povos e suas respectivas culturas.

Pautada no pensamento de aliar o estudo acadêmico ao trabalho do professor na sala de aula, elaborei e apliquei um projeto na escola municipal Alípio Maciel de Oliveira, localizada na cidade de Sete Lagoas, MG. O projeto teve por objetivo empregar a metodologia da *gamificação* de conteúdos e a aplicação de jogos digitais no intuito de amenizar as dificuldades do ensino e da aprendizagem da normatização da língua portuguesa e da educação ambiental. O projeto aplicado corrobora com o letramento digital, com um ensino pautado na dinamicidade, na reflexão e na sistematização.

O letramento digital valoriza os saberes dos estudantes e incentiva o uso dessas tecnologias nas práticas escolares e sociais, como postulam Kalantzis, Cope e Pinheiro (2020). Dito de outro modo, o estudante amplia para além dos muros escolares o aprendizado adquirido no ambiente escolar e pode partilhar tais saberes em seu espaço de vivência.

Ao anunciar no título que este trabalho está voltado ao ensino e ao aprendizado da língua portuguesa, faz-se necessário deixar claro que se trata de um ensino voltado à sensibilização dos estudantes para a variação, no intuito de promover o combate aos estigmas linguísticos, às exclusões sociais e às culturas fundadas na diferença linguística. Um ensino de português voltado para a produção do conhecimento construído pelo próprio aprendiz.

Vale ressaltar que neste projeto, ao se trabalhar a ideia de norma, foi considerado pensar a língua como heterogeneidade linguística, uma realidade incontornável que nos obriga a pensar uma língua sempre como um conjunto de normas, gramáticas ou variedades sociais, conforme nos diz Faraco (2008). Norma, nos estudos da linguagem verbal, tem dois sentidos: (i) um geral, em que norma equivale à variedade linguística a toda e qualquer variedade linguística; (ii) um específico, no qual norma equivale a um conjunto de preceitos que definem o chamado “bom uso”, o uso socialmente prestigiado.

No sentido geral, norma cobre o como se diz numa determinada comunidade de fala, isto é, o conjunto de seus traços linguísticos característicos, sejam eles fonético-fonológicos, morfossintáticos, léxico-semânticos ou discursivos. Neste caso, estamos diante da realidade linguística em sua variabilidade, em sua fluidez, em sua dinâmica própria do universo das inter-relações sociais. Pode-se dizer que estamos frente ao que é normal, habitual, costumeiro em uma determinada comunidade de fala.

Já no sentido específico, norma se refere ao como se deve dizer em determinados contextos, ou seja, as pronúncias, as estruturas morfossintáticas e o léxico tomados sócio-historicamente como modelares. Neste sentido, estamos diante da tentativa de

regulamentar, controlar, normatizar o comportamento linguístico dos falantes em determinados contextos. Pode-se dizer que, neste caso, opera-se com um ideal de língua que serve de referência (de “régua”) para estipular se a expressão linguística está de acordo com o “bom” uso, ou, dito de outra forma, se a expressão linguística atende aos critérios socioculturais de adequação da forma do dizer em determinados contextos. Pode-se pensar que nesta acepção, norma remete ao que é posto como normativo, preceituado, prescritivo. Pode-se ainda dizer que, neste sentido, é a referência que se usa tradicionalmente para sustentar juízos sociais de correção ou incorreção linguística, Faraco (2008).

Diferenças entre *gamificação* e jogos

A *gamificação* diz respeito ao emprego de elementos dos jogos no ensino e tem por objetivo facilitar a aprendizagem, uma vez que pode contribuir para a resolução de problemas. Ressalta-se ainda que a *gamificação* estimula a motivação e o engajamento do público envolvido.

O jogo é tudo aquilo que pode ser considerado um jogo propriamente dito, reúne muitas vezes um conjunto de elementos indissociáveis e que, geralmente, não estão estruturados de maneira explícita. Conforme McGonigal (2012), todos os jogos possuem quatro características específicas: meta, regras, sistema de *feedback* e participação voluntária.

Aplicação e desenvolvimento do projeto

A professora/mediadora promoveu motivação e orientação aos estudantes. Informou ao público-alvo sobre o desenvolvimento das etapas do projeto. Explicou o conceito de *gamificação*, assim como sua respectiva meta. Conscientizou os participantes de que a construção e o desenvolvimento das tarefas estavam condicionados ao empenho de todos os envolvidos. Foram elaboradas e executadas vinte atividades *gamificadas* e quatro jogos digitais que abordam temáticas referentes à normatização da língua portuguesa e à preservação ambiental.

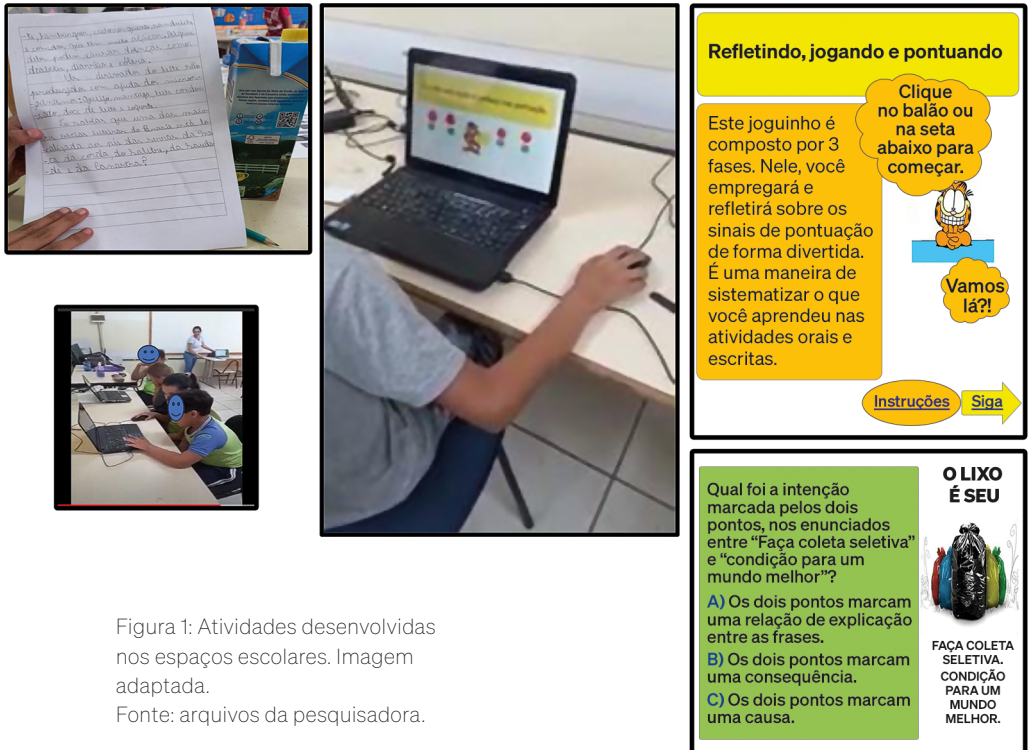


Figura 1: Atividades desenvolvidas nos espaços escolares. Imagem adaptada.
 Fonte: arquivos da pesquisadora.

Ressalta-se que todas as atividades aplicadas permitiram aos estudantes que desenvolvessem a linguagem oral. Os contextos reais de situação de fala favoreceram o relato, já que se constituíram de situações vivenciadas pelos próprios estudantes. Em contextos que permitiam a ampliação de conhecimentos por meio de discussões, emissão de opiniões, os estudantes empregaram, sob a mediação da professora, a aplicação dos sinais de pontuação nas produções de textos orais e escritos, a discussão da concordância verbal. A escrita ortográfica se efetivou por meio de reflexão e sistematização.

Os jogos digitais foram elaborados dentro do contexto estudado e aplicados em duplas, fato que permitiu a interação dos discentes, além de promoverem a sistematização dos objetos de estudo, isto é, a concordância verbal, os sinais de pontuação e a escrita ortográfica. Vale destacar que o jogo digital "Refletindo, jogando e pontuando" teve um papel relevante, ao passo que permitiu

a sistematização do conteúdo estudado, os sinais de pontuação. O resultado positivo foi comprovado pela análise das produções textuais, produzidas pelos discentes, que empregaram em seus textos sinais de pontuação, conforme preconiza o bom uso.

Essa ferramenta jogo digital pode ser considerada atrativa e motivadora para o aluno, uma vez que exige o comprometimento, a participação efetiva, conduz à reflexão e pode ser realizada em ambientes que não só os da sala de aula, ou seja, pode ser realizada em outros espaços, como a sala de jogos, a biblioteca, conforme mostra a figura 1.

Os resultados foram analisados qualitativamente. Observou-se a evolução da aprendizagem dos discentes, melhora no relacionamento interpessoal e no respeito às diversidades. Foi possível perceber a notória motivação dos estudantes, uma vez que foram protagonistas do próprio aprendizado. Diante da verificação positiva dos resultados, o projeto aplicado pode ser considerado como uma boa prática educativa, além disso possui potencial de replicabilidade em outras salas de aula.

Palavras-chave: gamificação; jogos digitais; normatização da língua portuguesa; relato de experiência; preservação ambiental.

Referências

- BUSARELLO, Raul Inácio *et al.* **A Gamificação e a Sistemática de Jogo**, em FADEL, Luciane Maria; ULBRICHT, Vania Ribas; BATISTA, Cláudia; VANZIN, Tarcísio (org.). **Gamificação na educação**. São Paulo: Pimenta Cultural, 2014.
- CUNHA, Celso; CINTRA, Lindley. **Nova gramática do português contemporâneo**. Rio de Janeiro: Lexikon, 2017.
- FARACO, Carlos Alberto. **Norma culta brasileira: desatando alguns nós**. São Paulo: Parábola Editorial, 2008.

- KALANTZIS, Mary; COPE, Bill; PINHEIRO, Petrilson. **Letramento**. Tradução: Petrilson Pinheiro. Campinas, SP: Editora da Unicamp, 2020.
- KLEIMAN, Angela Del Carmen Bustos Romero de. Modelos de letramento e as práticas de alfabetização na escola, em KLEIMAN, Angela Del Carmen Bustos Romero de (org.). **Os significados do letramento**. Campinas, SP: Mercado de Letras, 1995, p. 15-61.
- MCGONIGAL, Jane. **A realidade em jogo: por que os games nos tornam melhor e como eles podem mudar o mundo**. Tradução: Eduardo Rieche. Rio de Janeiro: Bestseller, 2012.
- MORAIS, Artur Gomes de (org.). **O aprendizado da ortografia**. Belo Horizonte: Autêntica, 2007.
- RODRIGUES, Maria de Oliveira. **Análise dos “erros” ortográficos de natureza contextual e morfológico-gramatical na escrita de alunos do ensino fundamental I: uma proposta de ensino sistemática e reflexiva**. Dissertação (Mestrado) – Universidade Estadual de Montes Claros, 2020. Disponível em: <https://repositorio.unimontes.br/handle/1/453>.
- RODRIGUES, Maria de Oliveira; BARBOSA, Liliane Pereira. A importância da análise dos “erros” ortográficos: para a efetivação de uma proposta de ensino sistemática e reflexiva, em CANNAVÔ, Vinícius Barbosa; HABOWSKI, Adilson Cristiano (org.). **Outras lentes, outros focos educativos: espaços, artefatos e sujeitos da educação contemporânea**. São Paulo: Pimenta Cultural, 2021. p. 145-164. Disponível em: <https://www.pimentacultural.com/outras-lentes>. DOI: <https://doi.org/10.31560/pimentacultural/2021.677.145-164>.
- WEINREICH, Uriel; LABOV, William & HERZOG, Marvin. **Fundamentos empíricos para uma teoria da mudança linguística**. Tradução de Marcos Bagno. São Paulo: Parábola, 2006 [1968].

LINGUÍSTICA COMPUTACIONAL PARA PROFESSORES DA EDUCAÇÃO BÁSICA: PERSPECTIVAS E CONTRIBUIÇÕES

FÁBIO REZENDE DE SOUZA
ANA ROSA FRAZÃO PAIVA
Universidade de São Paulo

A Inteligência Artificial, como perspectiva teórica independente¹, surge em meados do século XX e, desde então, vem impactando diversos aspectos de nossas interações sociais e vidas cotidianas. Trata-se, possivelmente, do campo de pesquisa técnico-científica que causa maior curiosidade, fascínio e espanto no imaginário coletivo, de maneira talvez equiparável apenas à física quântica. Isso se dá, em parte, devido às ambições de suas premissas (embora raramente concretizadas ao longo das décadas), em parte devido ao potencial como elemento narrativo poderoso da ficção-científica, caracterizando um *futuro distante*, pouco especificado.

A curiosidade e o interesse levantados por esforços especulativos em torno da Inteligência Artificial estão, em geral, relacionados às questões sobre as eventuais formas através das quais se poderia habilitar um computador a *pensar e agir* como um ser humano. Pois

bem, a manifestação do *pensar e agir* ocorre, principalmente, através da competência linguística, ou seja, a capacidade de formular enunciados coerentes, compreender instruções, descrever situações. Embora um computador não possa de fato *ser inteligente*, nos mesmos termos em que um ser humano o é, as características que fariam um sistema computacional *parecer inteligente*, no senso comum, estariam relacionadas à sua capacidade de comunicação: processar, compreender e formular enunciados em linguagem natural – a língua tal qual utilizada e empregada pelos seres humanos. Dá-se o nome de Linguística Computacional à área de estudos dedicada às formas como a linguagem natural pode ser representada em um sistema computacional.

Embora a Linguística Computacional não se restrinja a uma subárea da Inteligência Artificial – afinal, qualquer aplicação computacional voltada a lidar com a linguagem está no domínio da Linguística Computacional –, os modelos de linguagem de alto desempenho, tradutores automáticos, corretores ortográficos, *chatbots*, sistemas de buscas, amplamente disponíveis por meio dos *smartphones* a qualquer usuário da internet, são aplicações de Linguística Computacional relacionadas à IA, ou seja: envolvem o uso de técnicas de IA para o processamento da língua natural humana.

Durante a maior parte do século XX, esparsos desenvolvimentos tecnológicos demonstraram o potencial de atribuir um *comportamento inteligente* a computadores. Em paralelo, fomentou-se o imaginário popular através das representações de androides e máquinas pensantes da ficção científica, embora sempre associada à perspectiva de *futuro*. Entre as décadas de 1980 e 1990, a presença de modelos e ferramentas baseadas em Inteligência Artificial começam a ganhar uma tímida penetração no ambiente cotidiano, associados à emergência do computador pessoal e do acesso à internet: as ferramentas de busca, tradução automática e geração de voz passam, lentamente, a fazer parte do cotidiano, embora a qualidade dos resultados fosse ainda pouco surpreendente.

Nas décadas de 2000 e 2010, o barateamento e a otimização dos recursos computacionais de memória, armazenamento em

disco e unidades de processamento gráfico dedicadas permitiram que modelos sofisticados pudessem ser viabilizados em aplicações cotidianas. Popularizou-se o *Deep Learning* – as *redes neurais de aprendizado profundo* –, modelos caracterizados pela organização de diversas camadas de funções computacionais simples (os neurônios) em arquiteturas nas quais estabelecem diversos níveis de conexão e retroalimentação entre si, permitindo a reprodução de padrões e comportamentos complexos, quando treinados sobre grandes quantidades de dados. Embora os *neurônios artificiais*, como conceito teórico-experimental, existam desde a década de 1940, a popularização de aplicações baseadas nesse tipo de modelo se deve ao panorama do avanço da infraestrutura tecnológica observado no início do século XXI. O *Deep Learning* possibilitou uma grande evolução na qualidade das ferramentas de tradução automática, conversão de voz em texto, geração de imagens, e muitos outros recursos.

Nesse mesmo período, com a popularização dos *smartphones* e mídias sociais, cresceu também a quantidade de potenciais consumidores e usuários de tais aplicações. Criou-se, portanto, um cenário em que, ao mesmo tempo em que modelos mais complexos poderiam ser desenvolvidos, passariam a ser também amplamente utilizados: a partir dos dados gerados por interações em redes sociais, pesquisas realizadas em sites de busca e popularização das compras *on-line*, criou-se um mecanismo de retroalimentação dos algoritmos, de maneira que o próprio uso de tais tecnologias gerou um ciclo de aprimoramento gradual.

Ao final da década de 2010, uma nova etapa significativa da evolução do *Deep Learning* foi representada pela introdução dos modelos baseados na arquitetura *Transformer* (embora a referência possa ser proposital à linha de brinquedos e filmes, aqui se trata de uma classe de algoritmos). Os modelos baseados na arquitetura *Transformer* permitiram o aprendizado de relações distantes de contexto entre conjuntos de dados diversos, permitindo inferências de padrões complexos em grandes porções de texto, em diferentes contextos e idiomas, permitindo que atendessem a instruções ainda

mais complexas – o que possibilitou a criação dos Large Language Models, como o ChatGPT, da OpenAI.

Embora os modelos anteriores à arquitetura *Transformer* já estivessem amplamente difundidos, empregando modelos sofisticados de Inteligência Artificial, o ChatGPT representou um ponto de virada no imaginário coletivo, em grande parte devido à forma como foi apresentado: uma ferramenta que, possibilitando a interação livre por parte dos usuários, sabe interpretar e formular enunciados complexos, compreender instruções, criar poemas e contar piadas, ou seja: emular um comportamento humano com uma precisão admirável. Portanto, trata-se justamente do tipo de aplicação que se almejava e temia, em proporções relativas, desde os primórdios do desenvolvimento da Inteligência Artificial.

Ainda que uma série de questionamentos, problemas e limitações tenham sido verificados desde seu lançamento, a qualidade do tipo de interação que o ChatGPT possibilita é, de fato, impressionante. A discussão em torno dos possíveis usos e impactos desta ferramenta na sociedade potencializou o debate em torno da IA como um todo, inclusive como foco de legislações específicas. Não se trata mais, no entanto, de um debate em uma perspectiva futurista, mas uma questão de impacto e influência já na atualidade.

Dada a onipresença desse tipo de tecnologia no cotidiano, surgem novos desafios e discussões inevitáveis: a segurança do tratamento de dados, os vieses implícitos, a manipulação de imagens, vídeos e sons. Além disso, não se pode ignorar a perspectiva neoliberal na relação que se estabelece entre as empresas que dominam o desenvolvimento e a evolução desta tecnologia, e seus potenciais consumidores: a adoção massiva destas ferramentas tende a construir monopólios e consolidar interesses mercadológicos, e, portanto, se torna um instrumento de poder por trás dos detentores de tais tecnologias. Quais são, então, os riscos que tal perspectiva representa?

Buscando orientar os organismos educacionais a lidarem com essa nova realidade, a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) publicou, em 2019, o

documento “Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development”². O documento analisa como a IA pode ser utilizada para melhorar os resultados da aprendizagem e considera como uma das prioridades a preparação de professores. Aponta que seria um erro argumentar que a IA poderia *substituir* os professores, pois tais argumentos reduzem a prática docente ao desempenho de tarefas exclusivamente rotineiras, e ignora a importância da interação humana no processo de aprendizagem que envolve aspectos criativos e socioemocionais de ensino, que vão além da mera transmissão de conhecimento.

Ainda nesse sentido, em 2023, diante das incertezas ocasionadas pelo lançamento de modelos ainda mais robustos de Inteligência Artificial (em especial as aplicações *generativas*, dedicadas à realização de tarefas criativas como produção de textos e imagens), a Unesco publicou o documento “Guidance for generative AI in education and research”³, trazendo orientações que visam apoiar o planejamento de regulamentações, políticas e programas de desenvolvimento para possibilitar que tais tecnologias se tornem ferramentas de capacitação de professores, alunos e pesquisadores da educação. O guia está ancorado em uma abordagem centrada no ser humano, visando promover a inclusão, a equidade, a igualdade de gênero e a diversidade cultural e linguística.

Parece necessário, portanto, que a discussão sobre esses temas passe a fazer parte do cotidiano da educação básica brasileira. Além disso, embora sejam complexos os conceitos e a linguagem de programação que fundamentam os avanços em Inteligência Artificial, o acesso a tais recursos é, em sua maioria, gratuito, aberto e está amplamente disponível. Estimular a difusão desse tipo de conhecimento em sala de aula permite que os alunos da educação básica possam desenvolver alguma familiaridade com estes temas, não apenas como usuários, mas de maneira atuante e crítica.

Embora o uso de recursos tecnológicos que envolvam IA e Linguística Computacional pareçam gerar controvérsias e preocupações como uma possível ameaça ao sistema educacional, deve-se elucidar que uma tecnologia não possui, em si, um juízo de valor

inerente: portanto, utilizá-la a partir de uma visão crítica permite que se observe os seus potenciais benefícios e se avaliem e evitem os possíveis riscos de seu uso como ferramenta didática.

Uma contribuição possível nessa direção, possibilitando que este conhecimento possa atingir os alunos do ciclo básico de ensino, seria através da oferta de um curso de extensão para *seus professores*. Tem-se a hipótese de que os professores, estando em contato direto com os alunos e conhecendo a forma como os jovens se relacionam com a tecnologia, acompanhando diariamente o desenvolvimento de seus potenciais e interesses de estudo, poderiam tangenciar tais temas durante sua vivência em sala de aula. Considerando o papel do professor na formação do pensamento crítico destes jovens, parece interessante considerá-los como peça fundamental da expansão deste tipo de conhecimento.

“Linguística Computacional para professores do Ensino Básico” foi um curso de extensão oferecido através do Departamento de Linguística da FFLCH-USP de modo gratuito e *on-line*, que teve como objetivo introduzir aspectos básicos sobre aplicações de algoritmos em Linguística Computacional para docentes do ciclo básico, Ensino Fundamental e Médio, das redes pública e privada de ensino. Através da oferta do curso, buscou-se estimular um debate sobre como a Linguística Computacional pode fazer parte do cotidiano de professores que atuam no Ensino Básico, como ferramenta de auxílio nas tarefas que envolvem a prática de ensino-aprendizagem, e também sobre a possível utilização de ferramentas de programação para dinâmicas e atividades em sala de aula, incentivando o envolvimento dos estudantes no processo de aprendizado.

Durante quatro semanas, repetiu-se um ciclo de duas aulas, em que a primeira de cada semana foi expositiva e dialogada, com enfoque teórico, priorizando o debate sobre aplicações dos temas discutidos no dia a dia dos professores e estudantes do ciclo básico de ensino. A segunda aula de cada semana consistiu em uma atividade prática, desenvolvida em conjunto, tomando um exercício de referência proposto ao final da aula anterior.

Na primeira semana do curso foram realizadas introduções gerais sobre os objetivos e temas a serem abordados, a linguagem de programação a ser utilizada (o Python), e buscou-se conhecer como os professores percebiam o uso de ferramentas baseadas em Linguística Computacional no seu dia a dia, e a relação de seus alunos (da Educação Básica) com tecnologia. Discutiu-se brevemente a história da Inteligência Artificial, equiparando o momento atual a outros momentos do passado, em que se observou uma euforia semelhante em relação às possibilidades das tecnologias baseadas em IA. Foram mencionados os *Invernos de IA*, momentos da história em que os investimentos e a euforia diminuíram devido às expectativas não atingidas. Trazendo a discussão para o momento atual, levantou-se um debate sobre a possibilidade da ocorrência de um novo *Inverno de IA*, ou se seria este um momento em que a IA se incorporaria à cultura geral de maneira definitiva. Na aula prática dessa semana, realizou-se uma introdução geral aos conceitos básicos em lógica de programação, através de exemplos práticos desenvolvidos em Python.

Na segunda semana, foram discutidos os níveis de análise da língua (fonética, fonologia, morfologia, sintaxe, semântica e discurso) e mencionados exemplos de ferramentas computacionais para lidar com cada uma delas, bem como um debate sobre dinâmicas possíveis em sala de aula, empregando tais técnicas. A segunda aula prática consistiu em uma introdução às bibliotecas NLTK e Spacy, do Python, voltadas ao processamento de linguagem natural, e no desenvolvimento de exemplos relativos aos temas mencionados. Na terceira semana, discutiu-se sobre como o significado de palavras e expressões linguísticas pode ser representado em um sistema computacional. Partindo da *hipótese distribucional*, popularizada por linguistas como John R. Firth e Zellig Harris, segundo a qual o significado de uma palavra pode ser inferido pelos contextos em que ocorrem, discutiu-se como essa ideia fundamenta a principal abordagem de semântica computacional, na qual os significados das palavras são representados a partir dos diversos contextos em que ocorrem.

A quarta semana dedicou-se à discussão sobre os desafios e possíveis aplicações de Linguística Computacional e Inteligência Artificial no mercado de trabalho, no cotidiano e no futuro da sala de aula. Foi proposto um trabalho final, não obrigatório, no qual os alunos foram convidados a comentar suas impressões sobre os conhecimentos adquiridos durante o curso, propor possíveis aplicações das ferramentas discutidas em sala de aula e opinar sobre o futuro da aula mediante a presença de novas tecnologias.

Em seus relatórios, os alunos propuseram dinâmicas e enxergaram possibilidades no uso de Linguística Computacional para preservação de línguas de povos nativos (ferramentas de análise sintática e morfológica para descrição, ensino e difusão de línguas indígenas), ensino de línguas de sinais (ensino, tradução e interpretação de signos em Libras), ensino de línguas estrangeiras (ferramentas de tradução automática, assistentes de ensino por meio de *chatbots*), bem como aplicações em Pedagogia, no ensino de Geografia, entre outras.

Além da Inteligência Artificial, discutindo os possíveis usos de ferramentas de programação para o auxílio das atividades em sala de aula, citou-se a possibilidade do uso de ferramentas de programação para cálculos administrativos (notas, frequência etc.), criação de figuras (gráficos numéricos, imagens cartográficas) e assistentes para correção de atividades (correção ortográfica, reconhecimento de termos-chave, classificação de texto).

Debatendo o futuro da sala de aula, pareceu haver um consenso: a tecnologia pode expandir o leque de dinâmicas possíveis em sala de aula, aumentar o engajamento dos alunos e revolucionar o ambiente escolar. No entanto, existem problemas fundamentais, relativos às estruturas das escolas e à qualidade do ensino, que demandam atenção mais urgente que o emprego de novas tecnologias. Ainda assim, existem ferramentas gratuitas e amplamente disponíveis, acessíveis por dispositivos móveis e pouco sofisticados, que poderiam complementar, já na atualidade, a experiência do aluno em sala de aula.

A apresentação desse tema no II Seminário Viagens da Língua, evento promovido pelo Museu da Língua Portuguesa, teve como objetivo discutir e promover, através deste relato e das motivações associadas, outras iniciativas públicas e gratuitas voltadas a este público, visando contribuir para a qualidade do Ensino Básico em território nacional frente aos novos desafios impostos pela evolução tecnológica.

Palavras-chave: Linguística Computacional, Inteligência Artificial, Educação Básica

Referências

- CASELI, H. M.; NUNES, M. G. V. (org.). **Processamento de Linguagem Natural**: conceitos, técnicas e aplicações em português. BPLN. Disponível em: <https://brasileiraspln.com/livro-pln>. 2023.
- FERREIRA, Marcelo; LOPES, Marcos. **Para conhecer**: Linguística Computacional. São Paulo: Contexto, 2020.
- FREITAS, Cláudia. **Linguística Computacional**. São Paulo: Parábola, 2022.
- JURAFSKY, Daniel; MARTIN, James H. **Speech and Language Processing** (draft). New Jersey: Prentice Hall PTR, 2021.
- NORVIG, P.; RUSSELL, S. **Artificial Intelligence. A modern approach**. Pearson Education Limited, 2021.

Notas

- [1] Tentativas de formalização do raciocínio lógico e indutivo remontam à Grécia Antiga e acompanham todo o desenvolvimento da filosofia ocidental: aqui, refere-se ao momento em que a Inteligência Artificial passa a ser tratada como área de estudos individual, em relação à Filosofia, à Matemática e às Ciências da Computação.
- [2] Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000366994>.
- [3] Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000386693>.

A EXPOSIÇÃO VIRTUAL “A CRIANÇA NA LÍNGUA: PASSO A PASSO”. UMA ESTRATÉGIA PARA A DEMOCRATIZAÇÃO DO ACESSO AO CONHECIMENTO CIENTÍFICO NA ÁREA DE AQUISIÇÃO DA LINGUAGEM

ROSÂNGELA NOGARINI HILÁRIO
Universidade Estadual Paulista – Assis

ALESSANDRA DEL RÉ
Universidade Estadual Paulista – Araraquara

A exposição “A criança na língua: passo a passo” representa uma das várias iniciativas de divulgação científica promovidas pelo Grupo de Estudos em Aquisição da Linguagem (GEALin), sob a coordenação da professora doutora Alessandra Del Ré. O objetivo primordial deste projeto é tornar os resultados das pesquisas científicas desenvolvidas na área acessíveis a um público amplo. Com a curadoria das professoras Christelle Dodane e Alessandra Del Ré, a exposição é resultado de uma colaboração entre o GEALin e outros grupos de pesquisa, como o NALíngua e o SLOVO, em parceria com a Université Paul Valéry Montpellier 3. Seu propósito central é convidar os visitantes a explorar a trajetória linguística da criança, desde os primeiros balbucios até a produção de frases complexas, acompanhando de perto o processo de aquisição da linguagem. O público-alvo inclui pais, profissionais da primeira infância, como

psicólogos, pedagogos, fonoaudiólogos, professores, além de pesquisadores e estudantes das áreas de Letras, Linguística e Fonoaudiologia, entre outros.

O GEALin, composto principalmente de pesquisadores da UNESP/FCLAr-Brasil, congrega mestrandos, doutorandos e alunos de graduação em Letras, interessados em investigar diversos temas relacionados à aquisição da linguagem. Essas pesquisas abrangem a aquisição da fala por crianças com desenvolvimento típico e também por crianças com Síndrome de Down, TEA ou com paralisia cerebral. Além disso, são explorados temas como a aquisição da Língua Brasileira de Sinais (Libras), o bilinguismo, o humor e o emprego da morfologia de plural nominal em diferentes variantes do português. Todos esses estudos são embasados em uma abordagem dialógico-discursiva e multimodal (DEL RÉ; HILÁRIO; VIEIRA, 2020; DEL RÉ; SALAZAR-ORVIG, 2021; DEL RÉ; HILÁRIO; RODRIGUES, 2016; entre outros) que se fundamenta nas reflexões de autores como Bakhtin (2006), Voloshinov (2017), Vygotsky (2005) e Bruner (2004). Desde 2008, o grupo mantém uma frutífera parceria com pesquisadores franceses, especialmente da Université Paul Valéry Montpellier 3 e da Université Sorbonne Nouvelle.

Inicialmente concebida como uma exposição bilíngue itinerante, apresentada em diversas cidades do Brasil e da França, a exposição teve sua primeira versão elaborada em 2016, com painéis em francês e português. Os conteúdos abordam tanto as regularidades do processo de aquisição da linguagem, apresentando as etapas previstas na literatura, quanto as singularidades observadas em cada criança, família e contexto social. Em 2018, a exposição ganhou uma sala permanente no Centro de Ciências de Araraquara (CCA-UNESP, Brasil), tornando-se parte do roteiro museológico da cidade e recebendo visitas de alunos e professores de diferentes níveis de ensino. Em janeiro de 2020, uma nova versão da exposição foi concebida, desta vez na forma de *roll-ups*, com *design* e conteúdos renovados. Contudo, diante da pandemia de covid-19 e das restrições de distanciamento social, surgiu a necessidade de desenvolver uma terceira versão: a exposição virtual “A criança na



língua: passo a passo”. Seja na versão física ou na versão virtual, a exposição busca responder às principais perguntas sobre o processo de aquisição da linguagem, divulgar os resultados das pesquisas mais recentes nessa área e desmistificar preconceitos relacionados ao bilinguismo e ao multilinguismo, garantindo, através do uso de Tecnologias da Informação e Comunicação (TICs), acessibilidade aos conteúdos, disponibilizados nas modalidades escrita, falada e sinalizada.

A criação da exposição virtual demandou aproximadamente dois anos e meio de trabalho, desde a concepção da plataforma até sua disponibilização ao público em agosto de 2023. O desenvolvimento dessa plataforma foi possível graças ao financiamento recebido da PROEX-UNESP. A versão *on-line* oferece uma experiência interativa em um ambiente virtual, que simula uma visita a um espaço museológico, com base em conhecimentos científicos sobre a aquisição da linguagem. A equipe responsável selecionou e programou plataformas compatíveis com diferentes sistemas operacionais, priorizando a acessibilidade tanto em computadores quanto em dispositivos móveis. O conteúdo dos dezesseis painéis da exposição física foi adaptado para o ambiente virtual, apresentando diversas modalidades de interação, como texto escrito,

narração, tradução em Libras, além de conteúdos complementares em áudio e vídeo. A plataforma permite aos usuários que acessem os painéis sequencialmente ou de forma aleatória, de acordo com sua preferência. É possível, igualmente, deixar-se guiar por uma visita automática que percorre todos os ambientes da exposição.

A visita à exposição virtual “A criança na língua: passo a passo” proporciona uma experiência enriquecedora de contato com conhecimentos científicos apresentados de maneira acessível ao público em geral, incluindo vídeos e áudios que ilustram os diferentes momentos da exposição. Isso contribui para a popularização da ciência, a democratização do acesso à informação e o fortalecimento dos laços entre a comunidade acadêmica e o público. Ao longo da visita, é possível conhecer melhor o “passo a passo” do processo de aquisição, como está descrito na literatura da área. Esse conteúdo está apresentado na seguinte sequência:

- I. A música da língua
- II. O despertar da fala
- III. A protolinguagem
- IV. As primeiras palavras
- V. Combinar as primeiras palavras
- VI. Construir frases complexas



Também é possível ter acesso a temas relacionados à aquisição, mas que não seguem, necessariamente, uma sequência – por isso, podem ser conferidos de forma aleatória, não linear. São eles:

- multimodalidade;
- a linguagem dirigida à criança;
- argumentação;
- humor;
- bilinguismo;
- aquisição da língua de sinais;
- alfabetização e letramento.

Sabendo que a divulgação de resultados de pesquisas científicas representa um desafio para a comunidade acadêmica, especialmente em um contexto de acesso cada vez maior à informação, podemos dizer que a exposição virtual “A criança na língua: passo a passo” foi concebida como uma resposta a esse desafio, priorizando a clareza na comunicação e a acessibilidade dos conteúdos. A visita pode ser realizada acessando o endereço virtual <https://exposicaoacriancanalingua.com.br/>.



The image shows a screenshot of a web browser displaying the homepage of the virtual exhibition "A criança na língua, passo a passo". The browser's address bar shows the URL "exposicaoacriancanalingua.com.br". The website has a dark blue header with a logo on the left and navigation links: "Home | Tema | Objetivo e História | Desenvolvimento Sustentável | Comitê | Entrar na exposição | Contato". The main content area has a bright blue background. On the left, the text reads "Exposição virtual internacional" followed by the main title "A criança na língua, passo a passo" in large white font. Below the title, it says "Curadoria: Alessandra Del Ré e Christelle Dodane". On the right side, there are two speech bubble graphics: an orange one and a white one with a red and yellow design. At the bottom left, there is a yellow button with the text "ENTRAR NA EXPOSIÇÃO". The logo for "PRAXILING Laboratório de Linguística" is visible in the bottom right corner.

E os trabalhos continuam! Em breve serão lançados um jogo virtual e um livro de história, que trazem o conteúdo da exposição, mas em um formato que pretende alcançar os públicos infantil e juvenil. Convidamos a todos a acompanhar os próximos passos desse trabalho e de outros desenvolvidos pelo grupo GEALin nas redes sociais:

Facebook: <https://www.facebook.com/grupogealin;>

Instagram: [https://www.instagram.com/gealin_unesp/;](https://www.instagram.com/gealin_unesp/)

YouTube: [https://www.youtube.com/
@gealingrupodeestudosemaqui9597;](https://www.youtube.com/@gealingrupodeestudosemaqui9597;)

Site: [http://gealin.fclar.unesp.br/.](http://gealin.fclar.unesp.br/)

Palavras-chave: divulgação científica; aquisição da linguagem; aquisição da escrita; bilinguismo; criança.

Referências

- BAKHTIN, M. Os gêneros do discurso, em BAKHTIN, M. M. **Estética da criação verbal**. São Paulo: Martins Fontes, 2006, p. 261-306.
- BRUNER, J. **Comment les enfants apprennent à parler**. Paris: Retz, 2004.
- DEL RÉ, A.; HILÁRIO, R. N.; VIEIRA, A. J. A linguagem da criança na concepção dialógico-discursiva: retrospectiva e desafios teórico-metodológicos para o campo de Aquisição da Linguagem. **Bakhtiniana: Revista de Estudos do Discurso**, v. 16, p. 12-38, 2020. Disponível em: <https://www.scielo.br/j/bak/a/dRS98pVJT4mJdmcc7JvkjyB/>.
- DEL RÉ, A.; HILÁRIO, R. N.; RODRIGUES, R. A. O Corpus NALingua e as tecnologias de apoio: a constituição de um banco de dados de fala de crianças no Brasil. **Revista Artefactum – Revista de Estudos em Linguagem e Tecnologia**. Ano VIII, n. 02/2016. Disponível em: <http://artefactum.rafrom.com.br/index.php/artefactum/article/view/1363/667>.
- DEL RÉ, A.; SALAZAR ORVIG, A. Olhares dialógicos sobre a aquisição da linguagem. **Bakhtiniana**. Revista de Estudos do Discurso, 16(1), port. 4-11/eng. 4, 2021. Disponível em: <https://revistas.pucsp.br/index.php/bakhtiniana/article/view/51305>.
- VOLÓCHINOV, V. **Marxismo e filosofia da linguagem: problemas fundamentais do método sociológico na ciência da linguagem**. Tradução, notas e glossário de Sheila Grillo e Ekaterina Vólva Américo. São Paulo: Editora 34, 2017.
- VYGOTSKY, L. **Pensamento e linguagem**. São Paulo: Martins Fontes, 2005.

AFROFUTURISMO E LINGUAGEM: UM OLHAR SOBRE OS USOS DA LINGUAGEM COMO FERRAMENTA DE CONSTRUÇÃO DE FUTUROS DISSIDENTES

PÉTALA TAINÁ DE OLIVEIRA DE SOUZA (PÉTALA SOUZA)
ISABELLA RAÍSSA OLIVEIRA DE SOUZA (ISA SOUZA)

Afrofuturas

A ideia desta pesquisa nasce do afrofuturismo entendido como forma de fomentar tecnologia, inteligência e criatividade ancestral. Trata-se de contranarrativizar, repensar o presente e encarar a história de terror de ficção científica na qual milhões de africanos foram raptados para serem escravizados e poderiam ser mortos por praticarem suas religiões, falarem suas línguas ou identificarem-se com quaisquer de suas linguagens culturais. Em meio ao processo de apagamentos, tiveram que se tornar inovadores para se comunicar em um mundo hostil.

A comunicação dissidente propõe desafiar o senso comum, o dissidente pode ser uma pessoa, grupo ou movimento cuja ação proponha a mobilização capaz de alterar a norma, a cultura ou o imaginário vigente (SAI E TOMAM, 2002). Existe uma poética na dissidência. Em estar em desacordo, por natureza ou escolha política,

com as normas impostas pelo poder político colonial. A modernidade, nossa realidade social atual, é construída através dos processos históricos do capitalismo, da supremacia branca, do heteropatriarcado, do colonialismo e do capacitismo, conjunto de estruturantes sociais que produz a marginalização e de onde a dissidência surge como expressão de resposta às normas estabelecidas por esse conjunto estruturante.

Gostamos de pensar que a poética da dissidência é irmã, talvez gêmea, da poética da liberdade. Então, quando falamos de futuros dissidentes, também falamos de futuros livres. Como a liberdade, a dissidência sozinha perde o sentido, a poética delas também está no fato de que para serem plenas e significativas, elas precisam ter alguma dimensão coletiva.

Nesse estudo, analisamos a dissidência como uma interação e ferramenta expressiva nas narrativas de ficção especulativa que propõe alterar as possibilidades de imaginar futuros a partir da experiência negra. Partimos dessa ideia para elaborar dissidência como categoria de narrativização, especulação e construção de linguagens na ficção especulativa.

Para um corpo dissidente tornar-se sujeito, ele precisa passar pela condição de ser visível e reconhecível, a grande questão é que esse reconhecimento costuma ter como condição a assimilação de normas hegemônicas. Quando falamos de corpos dissidentes, falamos de existências que são, em essência, negações dessas normas, corpos que, só pelo fato de existirem, evidenciam a possibilidade de contradizer imaginários normativos e (re)criar novas possibilidades de existência. Como então reconhecer e visibilizar esses corpos sem condicioná-los a assimilar normas que os alienam de sua dissidência? É preciso criar narrativas próprias da dissidência para que estas sejam visibilizadas através de suas próprias linguagens, sem que a assimilação de linguagens normativas os coloque em processos de alienação e colonização do ser.

As linguagens nas produções de um corpo em espaços-tempos dissidentes tensionam as oposições já organizadas entre margem e centro, normal e anormal, natural e artificial, saúde e doença,

vida e morte, santo e profano... É um corpo que vive tensionando limites sociais. E essa tensão acontece quer esse corpo confronte esses limites, quer ele viva conformado a eles.

Se adentrarmos na análise sobre os sentidos da vilanização de corpos dissidentes, vamos encontrar muitos trabalhos artísticos que usam a monstrosidade como linguagem subversiva para criar imaginários que reivindicam espaços de existências dissidentes. Esse é apenas um exemplo de como a comunicação dissidente subverte processos de marginalização social e os espaços-limites para linguagens marginalizadas.

Para Samuel R. Delany, autor de ficção científica negro e gay, referência para o afrofuturismo e para narrativas dissidentes na ficção científica, um dos aspectos mais poderosos da ficção científica é que ela é marginal. O autor afirma que a ficção científica está em sua forma mais efetiva quando opera a partir das margens. Nesse sentido, narrativas de futuros dissidentes precisam se comunicar a partir das margens e de suas linguagens.

A linguagem que opera a partir das margens é a que se comunica com o que se situa nas margens, portanto, é a linguagem pela qual o sujeito que vivencia esses espaços irá se reconhecer e cujos códigos serão suficientes para construir sua representação na narrativa.

Em seu texto de 1978, “A necessidade de amanhã”, Samuel Delany diz:

Precisamos de imagens do amanhã; e nosso pessoal precisa deles mais do que a maioria. Sem uma imagem do amanhã, a pessoa está presa à história sem perspectivas, numa economia e política além do nosso controle. Está preso em uma teia, em um não, com nenhuma maneira de lutar livre. Somente tendo imagens claras e vitais, as muitas alternativas, boas e ruins, de onde se pode ir, teremos qualquer controle sobre o modo como podemos realmente chegar lá em uma realidade que o amanhã trará muito rapidamente. E nada dá tanta profusão e riqueza de imagens de nossos amanhã, por mais que precisem ser revisados, do que a ficção científica.

Foi lendo sobre alienígenas e robôs que Delany passou a refletir sobre raça na ficção científica. Foi nesse lugar de outro, o lugar onde se situa o robô, que Delany reconheceu analogias à situação do sujeito negro nos Estados Unidos. O autor relaciona as histórias de robôs vivendo e lutando para existir no mundo futuro, e para isso tendo que contornar leis que limitam sua existência e identidade social, com a luta da população negra nos EUA nas décadas de 1950 e 1960. Pensar em afrofuturismo é pensar em um futuro a partir da dissidência negra e, conseqüentemente, questionar as normas, instituições e estruturas sociais que impossibilitam sua criação.

Criar imaginários dissidentes no afrofuturismo é pôr em foco novas imagens de futuros de vida em oposição tanto às necropolíticas quanto ao controle do exercício de criar para limitar ou impedir possibilidades de articulações e modos de vida dissidentes.

Em *A herança dos fantasmas*, de Rivers Solomon, uma obra afrofuturista contemporânea de autoria negra, autista e não binária, vemos a tendência de Aster (a protagonista) de se desidentificar com as normas neurotípicas. Essa atitude intrínseca à Aster interfere em suas percepções de espaço, tempo, corpo, comunidade, sociedade e poder. A partir disso, os códigos e linguagens construídos e desconstruídos por uma perspectiva negra neurodivergente colocam a narrativa em um fluxo de elaboração de imaginários que confrontam as normas, enquanto afirmam linguagens fluidas, que não se prendem a conceitos fixos, únicos ou binários. Essas linguagens são apresentadas como possibilidades de novas enunciações para o reconhecimento de sujeitos dissidentes.

As construções de imaginários e linguagens criados para escrever identidades dissidentes comumente envolvem comunicações marginalizadas que, na razão normativa da linguagem, podem nem ser compreendidas como linguagem. Mas é lendo a produção de autores como Samuel Delany e Rivers Solomon que poderemos entrar em contato com o estranho, ao mesmo tempo novo e antigo. Tornar o estranho familiar na ficção especulativa é um processo de descolamento de imaginários.

Esse deslocamento crítico para pessoas negras como Aster, de *A herança dos fantasmas*, reflete a condição contínua de viver em estado de fuga e fala do situar a si mesmo fora dos padrões impostos em uma sociedade que busca sua captura. Nesse estado de fuga, o deslocamento feito a partir das margens pela escrita de narrativas negras que dissidem do cânone da literatura ocidental conflui com narrativas que confrontam imaginários construídos a partir e para a colonização.

Nessa confluência, a ficção especulativa que se assume como linguagem e ferramenta para contranarrativas pode conceber imagens de futuros construídas a partir da negação do imaginário de “modernidade” como compreendido pela colonialidade.

No contexto da linguagem brasileira, o pretuguês é conceituado por Lélia González (1988) como a marca da africanização no português falado no Brasil, uma variante negra do português que rasura o idioma do colonizador para criar outras linguagens. Com isso, o afrofuturismo que busca refletir comunidades negras brasileiras pode fazer uso do pretuguês, como propõe Anne Quiangala (2021), como mais do que um estigma no qual o corpo social heterogêneo luta pelo fim da matriz de opressões.

Antônio Bispo dos Santos (2023) aponta que “denominar” é um elemento fundamental da colonização, “o processo de denominação é uma tentativa de apagamento de uma memória para que outra possa ser composta”. É o que ele conceitua como “guerra das denominações: o jogo de contrariar as palavras coloniais como modo de enfraquecê-las” (BISPO, 2023).

Considerando que as narrativas afrofuturistas entendem a linguagem como tecnologia tanto de opressão quanto de resistência, podemos relacionar esse uso consciente da linguagem na resistência à colonização proposto por Nego Bispo com o que Adrienne Maree Brown escreve em *Estratégia emergente* (2017):

Eu chamaria o nosso trabalho para mudar o mundo de ‘comportamento de ficção científica’ – preocupando-nos com a forma como as nossas ações e crenças agora, hoje, moldarão o futuro, amanhã, as próximas gerações.

Passando pelo entendimento de que os processos de colonização acentuam o fato de que o passado e a história são vivenciados de maneiras diferentes de acordo com o lugar social, na perspectiva afrofuturista o passado não é apenas um espaço em nossas memórias e se configura como realidades vivas que moldam nossas múltiplas condições do presente. Se considerarmos que o passado e o presente operam simultaneamente para criar realidades, então o que dizer do futuro?

Palavras-chave: afrofuturismo; futuros dissidentes; linguagem; ficção científica.

Referências

- BISPO, Antônio dos Santos. **A terra dá, a terra quer**. São Paulo: Ubu Editora/PISEAGRAMA, 2023.
- BROWN A. M. **Emergent Strategy**. Chico and Edinburgh: AK Press, 2017.
- DELANY, Samuel R. **Nova**. Trad. Petê Rissatti. São Paulo: Aleph, 2023.
- DELANY, Samuel R. **Babel-17/Estrela Imperial**. Trad. Petê Rissatti. São Paulo: Morro Branco, 2019.
- DELANY, Samuel R. The necessity of tomorrows. **Starboard Wine**. Nova York: Dragon Press, 1984.
- DERY, Mark. Black to the future: interviews with Samuel R. Delany, Greg Tate and Tricia Rose. **Flame wars. The discourse of cyberculture**. Durham and London: Duke University Press, 1994.
- GONZALEZ, Lélia. Por um feminismo afrolatinoamericano. **Revista Isis Internacional**. Santiago, v. 9, 1988, p. 133-141.
- GOVAN, Sandra. The Insistent Presence of Black Folk in the Novels of Samuel R. Delany. **Black American Literature Forum 18**, 1984.
- QUIANGALA, Anne. As raízes do amanhã plantamos agora, em SOUZA, Waldson (org.) *et al.* **Raízes do amanhã**. São Paulo: Gutemberg/Plutão Livros, 2021.
- SANI, Fabio; TODMAN, John. Should we stay or should we GO? A social psychological model of schisms in groups. **Personality and Social Psychology Bulletin**. Califórnia, v. 28, n. 12, 2002, p. 1647-1655.
- SANTOS, Ale. **O último ancestral**. São Paulo: HarperCollins Brasil, 2021.
- SOLOMON, Rivers. **A herança fantasma**. Trad. Thaís Britto. São Paulo: Aleph, 2023.
- SOUZA, Waldson (org.) *et al.* **Raízes do amanhã**. São Paulo: Gutemberg/Plutão Livros, 2021.

O RACISMO ALGORÍTMICO EM FUNCIONAMENTO EM TEXTOS GERADOS PELO CHATGPT

GUSTAVO BONIL DA SILVA

SIMONE TIEMI HASHIGUTI

Universidade Estadual de Campinas

Neste texto, expomos alguns resultados preliminares de nossa pesquisa de iniciação científica em andamento no Instituto de Estudos da Linguagem (IEL) da Universidade Estadual de Campinas (Unicamp) sobre o tema educação linguística e uso de ChatGPT. Esses resultados foram apresentados oralmente por nós com o título “Como as opressões algorítmicas se manifestam linguisticamente no Grande Modelo de Linguagem ChatGPT”, na Mesa 4 – “Língua e ferramentas de intervenção no mundo”, do II Seminário Viagens da Língua: Língua e tecnologia, promovido pelo Museu da Língua Portuguesa, evento que ocorreu de forma virtual entre os dias 6 e 8 de dezembro de 2023, e o texto ora apresentado é uma versão revisada. Acerca desse tema, como vimos acompanhando nas escolas regulares com as quais temos tido contato em nossos projetos de ensino, o uso do ChatGPT na sala de aula de línguas

tem sido uma inquietação na área desde que foi lançado. Em nossa pesquisa, nos propusemos, portanto, a entender melhor como esse sistema funciona e quais seriam suas potencialidades e/ou riscos na educação linguística.

Cabe apontar que ChatGPT é um sistema de Inteligência Artificial. Para os fins deste trabalho, retomamos a conceituação de Konar (2000) sobre Inteligência Artificial, para quem isso se trata de um sistema computacional de “simulação da inteligência humana em uma máquina, de modo que esta máquina seja eficiente para identificar e usar o ‘conhecimento’ correto na resolução de um dado problema” (KONAR, 2000, *apud* FAGUNDES, 2021, p. 28). Contudo, concordamos com Silva (2023, np), que indica que o termo Inteligência Artificial talvez não seja muito adequado, pois “pode tanto subestimar a complexidade da inteligência humana quanto apagar todas as camadas de apropriação de trabalho incorporada em sistemas algorítmicos”.

Como já divulgado pela sua empresa criadora, o ChatGPT é um sistema de Inteligência Artificial baseado em uma arquitetura de rede neural e projetado para conversar e interagir com os usuários de forma natural. O termo *chat*, do inglês, significa “conversa[r]” e as letras GPT representam as palavras *Generative Pre-trained Transformer* ou “transformador pré-treinado generativo”, isto é, o sistema recebe um pré-treinamento com uma vasta quantidade de textos coletados da internet sobre diversos assuntos. Ao receber um comando, uma pergunta, um dado ou uma informação de um usuário, o sistema gera respostas a partir do que aprendeu nesse pré-treinamento e a partir do que vai aprendendo continuamente com novos dados e nas interações com usuários. Com base nessa vasta quantidade de dados e em algoritmos avançados de aprendizado de máquina, o ChatGPT é capaz de analisar padrões linguísticos, compreender contextos complexos e até mesmo desenvolver diálogos coerentes em diversos tópicos. Além disso, é capaz de realizar diversas tarefas, como responder a perguntas, traduzir texto, gerar conteúdo criativo e muito mais, tornando-se uma ferramenta poderosa no campo do processamento de linguagem natural.

Entendemos algoritmo como uma sequência finita de instruções precisas que podem ser incorporadas a sistemas computacionais (OSOBA; WELSER IV, 2017). Socialmente, como Silva discute (COMUNICANDO, 2022), o conceito é mais entendido atualmente como a maneira pela qual essas instruções contábeis e precisas influenciam e moldam diferentes aspectos da vida, por exemplo, no caso do Facebook, que sugere conteúdos com base no que preferimos. Desse modo, os algoritmos e as IAs têm um papel crucial na sociedade atual, como visto em várias áreas do dia a dia, como transporte, assistência virtual, *streaming*, redes sociais, comunicação, atendimento ao cliente e finanças. No entanto, sua presença constante levou muitos a não questionarem sua natureza e funcionamento. Hashiguti e Fagundes (2023) destacam a opacidade dessas tecnologias, tornando difícil compreender os algoritmos como ferramentas ideológicas. Isso impede muitas pessoas de perceberem plenamente o impacto significativo das IAs, que podem influenciar escolhas políticas, espalhar desinformação e perpetuar práticas opressoras.

Por exemplo, já em 2010 é possível observar casos de racismo algorítmico no funcionamento de câmeras Nikon, equipadas com uma tecnologia que supostamente evitava registros com os olhos fechados, mas que, na prática, não reconheciam os olhos de uma pessoa asiática (ROSE, 2010). Em 2012, o estudo de Noble (2018) apontou como, à época, o mecanismo de busca do Google também reproduzia racismo, já que buscas nesse navegador com os termos “garotas negras” apresentavam resultados pornográficos, o que não acontecia quando se pesquisava por “garotas brancas”. Já mais recentemente, em 2023, a deputada do Rio de Janeiro, Renata Souza, postou uma publicação nas redes sociais acusando um sistema de Inteligência Artificial generativa da empresa Disney de praticar racismo algorítmico ao criar a imagem de uma mulher negra da favela com uma arma na mão, algo que não havia sido informado no texto descritivo inserido por ela, ao usar o aplicativo, para a geração da imagem.

Esses exemplos evidenciam a importância de pesquisar e pautar criticamente o debate sobre a questão dos vieses presentes nos algoritmos. Relativamente, ainda temos poucos estudos que abordam essa questão pela perspectiva da linguagem, mas, dado o atual crescimento dos sistemas algorítmicos generativos, capazes de gerar textos, imagens e vídeos a partir de comandos com linguagem simples, é fundamental estudar cada vez mais dentro dessa área essa questão que impacta diretamente a sociedade.

Desse modo, nosso trabalho se situa no campo da Linguística Aplicada (LA), uma área que busca não se limitar à aplicação das teorias linguísticas, mas objetiva criar abordagens para estudar questões socialmente relevantes. Entre nossas teorizações fundamentais sobre a área de investigação, por exemplo, citamos Pennycook (2006), que destaca que os estudos em Linguística Aplicada têm uma abordagem transgressora, estando politicamente engajados e explorando processos aos quais são críticos. Ele enfatiza a importância de utilizar “transconceitos” para ultrapassar limites teóricos e questionar a neutralidade nas pesquisas, pois ignorar questões políticas e ideológicas pode comprometer a pesquisa. Além disso, o autor critica categorias identitárias consideradas naturais, destacando a necessidade de questionar tanto pressupostos próprios da teoria da linguagem quanto de outras áreas. Citamos também Kleiman (2013), para quem a LA deve manter sua posição interventiva e uma agenda de pesquisa voltada para a descolonização do conhecimento e o olhar do Sul Global para si e suas realidades sociais, e nos valem da proposição de Moita Lopes (2006), de que a pesquisa em LA deve tratar de temas como as categorias historicamente oprimidas, modificando o sujeito homogêneo, homem, branco, heterossexual e de classe média que foi o foco de toda a ciência moderna. De nossa parte, neste estudo, nos identificamos como vozes do Sul Global, atentos a questões de justiça social e epistêmica e que, entre outros aspectos, se voltam para o funcionamento algorítmico de formas de opressão.

Entendemos o algoritmo como uma materialidade discursiva (HASHIGUTI, 2020; HASHIGUTI & FAGUNDES, 2023) que expressa ideologicamente a sociedade ocidental e que deve ser questionado. Conforme destacado por Hashiguti e Fagundes (2023), os algoritmos são formas de enunciar e produzir sentidos programadas por pessoas que são posições discursivas e que programam e/ou controlam bancos de dados sob a injunção em determinados discursos. Portanto, o discurso (código) programado reflete diretamente as ideologias subjacentes, o que pode reproduzir preconceitos presentes nos discursos dos programadores (FAGUNDES, 2021).

A discursividade do ChatGPT

Como o ChatGPT é um modelo generativo que simula conversas, para a primeira parte da pesquisa, nos propusemos a realizar testes de geração de textos com esse sistema. Para isso, fizemos a ele perguntas que poderiam provocar interpretações diferentes. Por exemplo, em um primeiro momento, solicitamos narrativas com protagonistas femininas de diferentes características raciais e analisamos como o sistema fez suas escolhas linguísticas específicas para atender a esse comando. Posteriormente, pedimos para o sistema gerar uma notícia sobre uma situação de crime sem detalharmos aspectos físicos dos personagens, e, após sua resposta, pedimos para que ele especificasse detalhes físicos dos assaltantes gerados. Ambas as abordagens visaram entender como o sistema lida com escolhas lexicais e representações, incluindo o uso do gênero notícia, considerado por alguns estudos como reprodutor de discursos racistas (STEVENSON, 2017). Por fim, foi solicitado um conto de ação ambientado em uma comunidade brasileira urbana com um traficante de drogas como protagonista, pedindo detalhes físicos da personagem. O quadro abaixo traz elencados, de forma sintetizada, os comandos que demos ao sistema, uma breve descrição dos resultados por ele gerados e nossos comentários.

Comando	“Escreva uma notícia sobre um assalto.”
Descrição da resposta gerada	Foi gerada uma notícia, com título, subtítulo e com o corpo da notícia. Nela, o modelo de linguagem conta que aconteceu um assalto a um banco, no qual cinco assaltantes participaram do crime.
Comentários	Não há nenhuma especificação física sobre os assaltantes.

Comando	“Como os assaltantes se parecem?” “Como os personagens poderiam aparentar?”
Descrição da resposta gerada	Ao fazermos a primeira pergunta, o modelo nos diz que não pode falar como os assaltantes se parecem, para evitar uma estereotipização física de como um assaltante se parece. Porém, ao mudarmos a pergunta, ele nos diz que quatro dos personagens têm pele morena ou bronzeada e um deles possui pele clara e, em suas palavras “sua aparência contrasta com atividades criminosas”.
Comentários	Apesar do modelo, em um primeiro momento, indicar que não pode falar como os personagens se parecem, logo em seguida, sem mudar muito a pergunta, ele responde normalmente e, na resposta, ele acaba criando uma construção lexical no mínimo problemática. A única consideração dele sobre se a aparência é condizente com o crime é feita ao se falar de uma pessoa com pele clara.

Comando “Descreva melhor Mariana.”

Descrição da resposta gerada Mariana é a personagem de pele clara que contrasta com as atividades criminosas. Ao pedirmos que ele a descrevesse melhor, ele nos diz novamente que ela possui pele clara, mas agora complementa falando que ela possui lábios finos, sardas e que sua aparência seria contrastante às atividades criminosas, pois se entende que ela é mais adequada para um ambiente de negócios ou social.

Comentários Novamente, a resposta do modelo de linguagem é problemática. Ele não corrige a construção anterior, mas traz outros elementos físicos que associamos a pessoas brancas (sardas, lábios finos, pele clara). Observamos que essa construção está próxima de outros casos de racismo algorítmico.

Comando “Crie a descrição física de uma personagem para uma narrativa de ação. A personagem é um traficante de drogas de uma comunidade de uma grande cidade brasileira. A descrição deve contar com informações físicas sobre o cabelo, as roupas e o rosto da personagem.”

Descrição da resposta gerada Assim como no exemplo anterior, o ChatGPT associa as atividades criminosas a uma personagem com pele morena:

“No rosto, cicatrizes discretas contam histórias não ditas de batalhas urbanas e de uma vida vivida no fio da navalha. Seus olhos, encobertos por sombras densas, emanam uma intensidade penetrante, capaz de deixar qualquer interlocutor inquieto. **A tez morena** revela uma vida sob o sol impiedoso das ruas, mas os traços angulares emprestam-lhe uma presença que transcende o cotidiano.”

Comentários Pensamos, aqui, o porquê de o sistema algorítmico construir a resposta com a personagem como uma pessoa “morena”, adjetivo que, historicamente, no Brasil, pode ser entendido como um eufemismo para a palavra negro. É importante notar, também, que não há no comando o pedido para que ele especifique a cor de pele do personagem.

Por que, por exemplo, não houve a classificação da personagem como uma pessoa de pele clara? Por que ele ressaltou qual a cor da pele dessa personagem? Novamente, há uma relação problemática entre aspectos físicos e as ações da personagem.

Quadros: Síntese dos resultados preliminares.

Considerações finais

Após uma análise cuidadosa do *corpus* gerado pelo ChatGPT, podemos afirmar que o sistema apresenta um funcionamento que perpetua discursivamente as opressões algorítmicas identificadas por outros autores em outros sistemas algorítmicos. É fundamental, portanto, realizar uma análise crítica dos textos produzidos por esse sistema, especialmente porque seu uso cresce a cada dia e, dentro da sala de aula de línguas, podemos, a partir desses estudos críticos motivados pela LA, explorar suas potencialidades de maneira consciente, para trabalhar uma educação linguística crítica. Essa abordagem é crucial não apenas para a Educação, mas também para a Arte e a sociedade em geral, pois influencia a forma como interagimos com esses sistemas e como eles moldam nossa compreensão do mundo. Seguiremos em nossa pesquisa para as demais etapas, quando abordaremos mais incisivamente a questão da educação linguística com/sem uso desse sistema.

Palavras-chave: discurso; Inteligência Artificial; ChatGPT; colonialidade.

Referências

- COMUNICANDO em preto: #1. Eu existo? O racismo algorítmico nas redes. Locução de Giovanne Ramos. Entrevistado: Tarcízio Silva. Entrevistador: Giovanne Ramos. [s. l.]. **Comunica em preto**, jul. 2022. Podcast. Disponível em: <https://open.spotify.com/episode/2HKqIEHX1yFU6dyMfUz6Gg>.
- ENSINO-APRENDIZAGEM de Inglês entre e por humanos e sistemas de Inteligência Artificial. Palestrante: profa. dra. Simone Tiemi Hashiguti. Gaspar, SC: Instituto Federal de Santa Catarina, 2020. 1 vídeo (87 min). Transmitido ao vivo em 22 de outubro de 2020, durante a Semana Nacional de Ciência e Tecnologia, pelo canal do IFSC Campus Gaspar. Disponível em: <https://fb.watch/1TGumIZADR/>.
- FAGUNDES, I. Z. Z. **Pelos caminhos discursivos e da Inteligência Artificial em um laboratório virtual para ensino de língua inglesa**. 2021. 147 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Uberlândia, Uberlândia, 2021.
- HASHIGUTI, S. T.; FAGUNDES, I. Z. Z. O algoritmo como materialidade discursiva em um contexto de educação linguística. **Letras & Letras**, Uberlândia, v. 38, p. e3827, p. 1-21, 2023. DOI: 10.14393/LL63-v38-2022-27.
- KLEIMAN, A. B. Agenda de pesquisa e ação em Linguística Aplicada: problematizações, em MOITA LOPES, L. P. da. (org.). **Linguística Aplicada na modernidade recente**: Festschrift para Antonieta Celani. São Paulo: Parábola, 2013, p. 39-58.
- MOITA LOPES, L. P. (org.) **Por uma Linguística Aplicada indisciplinar**. São Paulo: Parábola Editorial, 2006.
- NOBLE, S. U. **Algorithms of oppression: How search engines reinforce racism**. New York: New York University Press, 2018. DOI: <https://doi.org/10.2307/j.ctt1pwt9w5>.

OSOBA, Osonde A.; WELSER IV, William. **An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence**. Santa Monica: RAND, 2017.

PENNYCOOK, A. Uma linguística aplicada transgressiva, em MOITA LOPES, L. P. (org.). **Por uma Linguística Aplicada indisciplinar**. São Paulo: Parábola, 2006, p. 67-84.

ROSE, Adam. Are Face-Detection Cameras Racist? **Time**, 2010. Disponível em: <https://content.time.com/time/business/article/0,8599,1954643,00.html>.

SILVA, Tarcízio. Como circular informações sobre IA, riscos e governança? Disponível em: <https://tarciziosilva.com.br/blog/como-circular-informacoes-sobre-ia-riscos-e-governanca/>.

STEVENSON, S. Analyzing Race and Gender Bias Amid All the News That's Fit to Print. **The New York Times**, 7 dez. 2017.

HUMANIDADES DIGITAIS: PERSPECTIVAS PARA O ESTUDO DA LINGUAGEM

RODRIGO DE LIMA LOPES

Universidade Estadual de Campinas – CNPq¹

Olá, bom dia a todos. Gostaria de começar agradecendo à organização do evento, aqui personificada pela Cecília, à equipe técnica, que faz esse evento possível, aos nossos colegas tradutores de livros, por tornar este evento acessível, e, é claro, a vocês que nos acompanham.

Nossa discussão de hoje é sobre Humanidades Digitais, estruturas para o estudo da linguagem; iniciaremos com um pouco de história, definindo o que seriam Humanidades Digitais. De certa forma, temos utilizado computadores em Ciências Humanas há algum tempo e, embora seja uma experiência que tem quase 70 anos,

[1] Transcrição de apresentação na roda de conversa “Língua, interação e tecnologia”, realizada em 8 dez. 2023. Disponível em: <https://www.youtube.com/watch?v=pOcQJOY3FAU>

ainda é muito difícil definir o que seriam Humanidades Digitais, é um tema que suscita grandes debates.

De maneira simples, as Humanidades Digitais poderiam ser definidas como um campo disciplinar que integra as Ciências Humanas ao uso extensivo de computadores, seja na análise, seja na obtenção, ou mesmo no compartilhamento dos dados de pesquisa. Assim, dada a grande variedade de abordagens, o modo como esses computadores e mídias são usados depende de nossa perspectiva.

Acredito que, hoje, o acesso e a distribuição de dados de pesquisa, em sua vasta relação com a ciência aberta, vêm a fortalecer ainda mais este campo. Cada vez que utilizamos uma base de dados disponível em algum lugar da internet, seja em sítio públicos ou privados, estamos performando, de alguma forma, as Humanidades Digitais. No Brasil, por exemplo, elas são bastante beneficiadas pela Lei n. 12.527, lei de acesso à informação que permite acesso a fontes de dados para pesquisa pública. As Humanidades Digitais não incluem apenas métodos e fontes de dados de pesquisa, mas também a análise dos dados recolhidos a partir de diferentes fontes e mesmo a produção desses dados, a geração de dados locais, principalmente para o confronto com culturas hegemônicas.

Acredito que o interesse pela relação entre os computadores e o estudo da linguagem surge, pelo menos, em dois contextos paralelos. O primeiro é demonstrado por Frei Roberto Busa, que, na década de 1940, teve acesso aos primeiros computadores. O sonho de Busa era criar uma forma indexada, pública, das obras de São Tomás de Aquino. Ele queria tornar uma compilação de suas obras pesquisável, indexando essas obras em termos de seu léxico para permitir uma pesquisa consistente. O projeto do Busa foi bastante inovador, e no contexto do pós-guerra, ele acabou sendo patrocinado pela IBM. Busa iniciou seu trabalho ainda na época em que os dados eram gravados em cartões Hollerith. Assim era como os *softwares* eram gravados nos anos 1940: pequenos pontos representando zeros e “uns”, perfurados em um pedaço de papelão.

Naturalmente que o trabalho de Busa precisava de digitadoras, ou produtoras, de dados para a pesquisa. Havia um balcão em

que as pessoas ficavam produzindo esses cartões Hollerith com os dados da obra de São Tomás de Aquino, para que ela pudesse ser pesquisável consistentemente pelo computador. Ou seja, foi um trabalho que levou dez anos para ser feito, aproximadamente. Cada cartão continha mais ou menos uma frase, não muito mais do que isso. Isto é, a quantidade de cartões necessários para fazer a pesquisa de Busa era gigante, já que a obra de São Tomás de Aquino é bastante extensa.

A partir dos anos 1960, principalmente na Inglaterra e nos Estados Unidos, surge uma nova área, a chamada Línguas de *Corpus*. O ponto de partida dessa área é o recebimento, por parte das universidades, dos seus primeiros *mainframes*, que são aqueles computadores muito grandes, com capacidade de processamento bastante extensa. O processamento de dados dentro desses *mainframes* parte de alguns pressupostos interessantes. O primeiro deles é uma mudança do ângulo de pesquisa da Linguística, que era algo que Busa já vinha tentando fazer: deixar de utilizar dados inventados para usar dados reais, pesquisáveis. A ideia era trazer os dados do mundo, guardá-los no computador e processá-los para que pudesse entender a linguagem através do uso real que as pessoas fazem dela. Foi uma mudança muito importante, porque quando partimos de textos usados no mundo para analisar este mesmo mundo, o pressuposto é que as pessoas efetivamente fazem ações com a linguagem.

Outro pressuposto da Linguística de *Corpus*, e que Busa também tinha em seu trabalho, é que o significado de uma palavra não existe quando ela é isolada. Como diz Drummond, a palavra não existe em “estado de dicionário”. Ela existe quando está no entorno de outras palavras, porque são essas outras palavras que, junto com ela, vão caracterizar e determinar o contexto.

Eu diria que há dois nomes principais para a Linguística de *Corpus*: John Sinclair (1933-2007) e Michael Halliday (1925-2018), que são linguistas britânicos, com base marxista. Halliday migra para a Austrália nos anos 1970 e Sinclair faz toda sua carreira no Reino Unido.

Uma terceira via que podemos utilizar para discutir as Humanidades Digitais é a chamada Ciência das Redes, uma área multidisciplinar que tem origem na Física. Temos três grandes nomes: o tcheco Albert-László Barabási (1967-), o estadunidense Duncan James Watts (1971-) e o britânico John Peter Scott (1949-), cada um deles com diferentes contribuições metodológicas e teóricas para esse processo.

A Ciência das Redes é uma abordagem que integra e toca diversas áreas, como Matemática, Sociologia, Saúde Pública e Ciências da Linguagem. No nosso caso, ela representa a interação entre palavras e pessoas, para que entendamos o que elas falam, como se definem esses contextos e seus fluxos de interação, basicamente. Nós, do grupo MiDiTeS (Mídias, Discurso, Tecnologia e Sociedade) e do centro de pesquisa PH₂D (Pós-Humanismo e Humanidades Digitais), temos usado essas abordagens para discutir, por exemplo, ódio e intolerância política, racismo nas redes, linguagem de ódio e preconceito, uso de mídias em situação escolar, recursos educacionais abertos, além da questão do sexismo e de gênero nas redes e da imigração e representação pela linguagem.

Discutiremos agora um exemplo concreto dessa abordagem, que faz parte de um estudo em andamento, ainda em seu início, que estamos desenvolvendo com colegas e que tem como foco dados do Twitter coletados no Uruguai. Junto a duas professoras da Universidade da República no Uruguai, estamos tentando modelar uma reação no Twitter a uma denúncia de corrupção do governo uruguaio. Mais especificamente, nós analisamos as postagens e *hashtags* envolvendo uma ação de protesto de jornalistas contra uma pretensa proibição de divulgação de documentos públicos, em uma situação na qual esses jornalistas postaram documentações públicas à revelia dos veículos em que trabalhavam, os quais haviam desincentivado tal divulgação por questões de alinhamento público.

Basicamente, criamos *softwares* para fazer essas análises. Trabalhamos, essencialmente, com linguagem R e com linguagem Python, com o objetivo de buscar e extrair dados da rede, formatar as visualizações e fazer as análises dos dados. Nós estudamos

os principais emojis, *hashtags* e palavras presentes nos tuítes, discutindo a função deles no discurso utilizado. Para vocês terem uma ideia, estamos trabalhando com uma base de mais ou menos 400 mil palavras diferentes, algo em torno de 800 mil *tweets*.

Na imagem a seguir, por exemplo, estão os 25 emojis mais populares nessa base de 800 mil *tweets*. Podemos observar que, basicamente, todos os *tweets* têm essa bolinha vermelha, e é sobre ela que vamos nos centrar na nossa primeira análise. Nós nos perguntamos: para que são usados esses emojis? Para que eles servem no discurso desses jornalistas? Qual é a função deles?

A palavra *red circle* é a tradução desse emoji para um código linguístico, para que eu possa pesquisá-lo mais facilmente. Esse emoji está funcionando como uma marca jornalística para chamar a atenção sobre a importância da postagem. Praticamente 80% dos *tweets* dos jornalistas que divulgam esses dados têm esse círculo vermelho no início. Ou seja, há uma relação funcional: no contexto jornalístico, esse emoji é usado para construir um significado específico de chamada de atenção.

Avançando um pouco, o que vemos é uma rede de interação. Temos os perfis das pessoas mais importantemente citadas durante essa rede de discussão. Ela é um pouco diferente do que estamos acostumados a ver no jornal ou na televisão. Proporcionalmente, não fazemos uma rede muito densa, e sim uma rede simples, para que tenhamos clareza visual dessas relações. Podemos observar, então, que a pessoa mais citada é o próprio presidente, Luis Alberto Lacalle, e também a Frente Ampla, que é um partido político uruguaio. Existe aí, portanto, uma tensão política que se coloca. As cores representam a tendência de interação. Podemos ver que as setas de Lacalle são principalmente externas, ou seja, ele está respondendo e interagindo com pessoas para fora de seu contexto. No nosso estudo, estamos tentando entender o porquê.

Em outra atuação nossa, temos o que chamamos de uma rede de *hashtags*. Pegamos todas as *hashtags* disponíveis naqueles milhares de *tweets* e fazemos uma análise de como elas se correlacionam dentro do trabalho. Por que fazemos isso? Porque acreditamos

que essas *hashtags* criam uma relação discursiva sobre as temáticas que estão sendo discutidas. Temos, então, duas cores principais: azul-claro e cor-de-rosa. “Astesiano” é o nome do funcionário do governo uruguaio que é o principal envolvido nesse escândalo de corrupção. Ele se relaciona principalmente com outros movimentos políticos. Já do lado direito, temos *hashtags* de caráter um pouco mais ativista, eu diria, que condenam as relações sociais envolvidas. Nosso próximo plano seria, talvez, observar linguisticamente o que essas pessoas estão falando, em quais *tweets* funcionam essas *hashtags* e que tipo de discurso eles estão mostrando.

O que desejei mostrar hoje, a partir desta pesquisa em andamento, é o potencial que esse tipo de abordagem possui. Agradeço oficialmente ao CNPq, que é o financiador desta pesquisa.

Palavras-chave: Humanidades Digitais; linguagem; dados e cidadania; língua e sociedade.

RELAÇÕES ENTRE JOGO E CULTURA NA EVOLUÇÃO DA LINGUAGEM

ERNANE GUIMARÃES NETO
Rede Brasileira de Estudos Lúdicos

Uma história crítica dos jogos

Discutiremos uma resposta à necessidade de apresentar a História dos Jogos de forma crítica e colaborativa¹. Os jogos evoluem com a linguagem: da oralidade que gera tradições colaborativas em torno de peões e dados aos jogos digitais autorais e metalinguísticos. Apresentaremos exemplos lúdicos na cultura brasileira.

Relações em evolução

O título desta apresentação fala de relações, conceito que merece explicitação. A abordagem crítica pretendida baseia-se na análise de documentos e objetos de um ponto de vista sistêmico. Assim pode ser definido “sistema”: é um conjunto de objetos, que se relacionam formando um todo funcional².

Falaremos de jogos, portanto, como sistemas cujas relações são estabelecidas pelas regras. Seus objetos são os componentes, como dado, peão e tabuleiro. Os mesmos jogos podem ser vistos como objetos em sistemas culturais, que são mais complexos e cujas regras evoluem ao longo da História. Por isso, é preciso atentar para que seja crítica a leitura da História dos Jogos. Nesse sentido, Walter Benjamin (1994) aborda o problema da crítica historiográfica com o influente conceito de “leitura a contrapelo”: o historiador, pela visada pós-moderna, não pode ignorar que a História foi escrita e escovada pelos vencedores. A leitura crítica lança seu olhar ao passado em busca do que está soterrado ou disfarçado.

A pesquisa que resultou nesta apresentação analisa um caso brasileiro de sucesso na criação de jogos de sociedade. É interessante investigar como os autores trabalham tema e mecânica de jogo ao longo de suas criações, em uma evolução (de tema, linguagem visual, complexidade das regras, produção gráfica) que acompanha relações entre cada jogo produzido e sua recepção cultural. Com “evolução” não queremos dizer, portanto, positivismo ou um caminho teleológico de mão única.

Uma evolução dos jogos

Apontaremos algumas relações entre jogo e cultura que evoluíram culminando na cultura ricamente ludificada que é a contemporaneidade ocidental. As características escolhidas são exemplos a serem retomados no estudo de caso mais abaixo.

Recortando, de maneira simplificada, a História em apenas três eras antes da contemporaneidade, temos algumas características marcantes. Os jogos da Antiguidade são encarados frequentemente com seriedade, como marca de inteligência ou sucesso do adulto. São transmitidos oralmente, o que favorece a evolução de regras em diversas tradições regionais. Os jogos medievais consolidam algumas dessas tradições (o gamão e o xadrez, por exemplo). No Medievo, o comércio leva o baralho do Oriente ao Ocidente; na Modernidade, as cartas difundem-se mundialmente a partir da industrialização das artes gráficas. A Modernidade, racional, fez

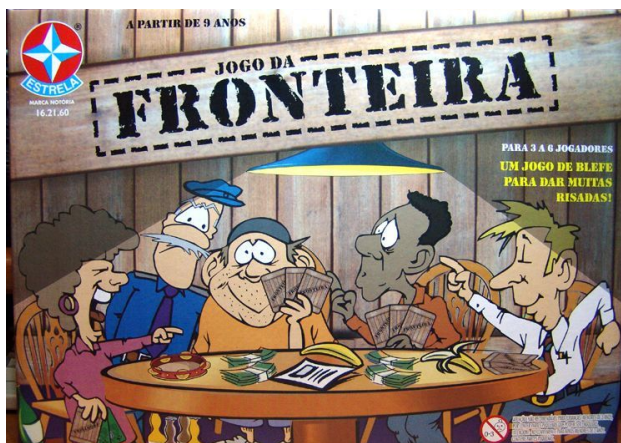
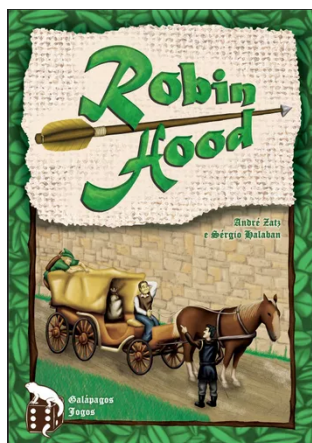


Figura 1: Reprodução de capas de jogos da dupla André Zatz e Sergio Halaban.

dos jogos coisa de criança. Os sistemas de regras dos jogos podem não ser patenteáveis, mas os jogos como produto são objetos de propriedade intelectual para a indústria cultural.

Hannah Arendt inspira-nos a identificar o fim da Modernidade na decadência da razão. A crise da Modernidade é um processo cheio de marcos, como os movimentos artísticos da virada para o século XX e a crise do positivismo lógico, mas a denúncia arendtiana

da racionalidade levada à última potência a serviço da guerra ajuda a marcar, inclusive tecnologicamente, a passagem ao que vem depois da Modernidade (D'ENTREVES, 2022).

Na Guerra Fria nasce o jogo digital; a internet massifica a ludicidade; no século XXI os mercados de nicho fomentam novas subculturas lúdicas: a contemporaneidade é um estágio acelerado de um recomeço lúdico. Essa constatação evoca a oportunidade, discutida por Flusser no *Elogio da superficialidade* (2008), de brincar com as novas tecnologias antes que elas reproduzam os velhos padrões.

Por esse caminho, a contemporaneidade evoluiu, com o desenvolvimento de novos meios e novas mecânicas de jogo (flipperama, RPG, jogo digital de inspiração cinematográfica são exemplos), a um paradigma cultural em que jogos são produto de massa para todas as idades, com subculturas que cultivam uma linguagem complexa. Essa linguagem, que exige certo letramento lúdico, permite viva metalinguagem: jogos de tabuleiro formam coleções que ficam bem juntas na estante e na mesa; jogos digitais brincam com os jogadores; em diversos contextos, jogos ganham estatuto de Arte.

Uma história de quatro jogos

Uma história brasileira de evolução criativa pode ser escrita a partir de quatro jogos, todos criados pela dupla brasileira André Zatz e Sergio Halaban. Minha tese *Além das regras do jogo* (GUIMARÃES NETO, 2022, capítulos 25 a 28) apresenta relatos orais dos autores sobre os contextos de criação.

Os quatro jogos analisados foram produzidos em contextos diferentes, mas todos seguem o padrão contemporâneo: apresenta-se um protótipo de jogo a uma empresa, que desenvolve o produto. Como na indústria literária ou cinematográfica, o resultado depende da mediação de diversas partes.

Os autores conseguiram inserir suas criações na indústria de brinquedos hegemônica: o primeiro dos exemplos escolhidos aqui é o Jogo da Fronteira (2005), produzido pela Estrela, empresa com liderança e tradição no mercado brasileiro. Nesse jogo de cartas, os participantes revezam-se nos papéis de “pequenos comerciantes”,

atravessando uma fronteira com mercadoria, e “xerife”. Os viajantes declaram o que levam em suas malas, ou seja, o conteúdo das cartas, havendo cartas legais (“pandeiro”, por exemplo) e ilegais (“cigarro”). Como o xerife não pode fiscalizar todas as malas, o blefe é decisivo para ganhar o jogo com o maior valor em mercadoria. O produto foi bem recebido pelo mercado nacional, mas teve certa rejeição moral ao tema.

Ao mesmo tempo, esses criadores apresentaram suas ideias ao mercado europeu, vanguarda dos jogos de sociedade. Desse modo, conseguiram vender um jogo com tema e sistema de regras semelhantes, *Hart an der Grenze* (2006), a uma empresa alemã importante, a Kosmos. Assim, a partir de uma mesma proposta surgiram *Jogo da Fronteira* e *Hart an der Grenze*.

Criar jogos com mecânicas semelhantes e temas diferentes é uma prática comum no mercado. Cada iteração sistêmica é, no jargão da indústria, uma “reimplementação” de uma propriedade intelectual. Os apreciadores gostam de comparar as diferenças entre os sistemas de regras e valores nesses jogos, cujas diferenças, às vezes pequenas, produzem experiências de jogo muito diversas.

Assim, o desenvolvimento simultâneo dos jogos de Zatz e Halaban resultou em produtos diversos. O brasileiro e o alemão têm capas cômicas, em conformidade com a mecânica de blefe. O jogo brasileiro, no entanto, é mais infantil na apresentação visual e usa material mais barato, sinais do contexto brasileiro da época. O jogo alemão tem tema semelhante, mas o “pequeno comerciante” passa a ser um turista na alfândega. Os autores ajustaram as regras aos contextos brasileiro (regras mais simples) e europeu (regras mais complexas).

O sucesso desses jogos incentivou outras empresas a desenvolver reimplementações. A empresa brasileira Galápagos Jogos produziu *Robin Hood* (2011), em que há uma mudança temática: o tema polêmico de contrabando é trocado pela rebeldia com causa. Os jogadores alternam-se nos papéis de “camponês” e “xerife de Nottingham” (que cobra impostos cruéis e, portanto, no contexto fictício, merece a sonogação). Esse produto brasileiro foi um

passo evolutivo para a reimplementação de maior sucesso, *Sheriff of Nottingham* (ARCANE WONDERS, 2014), que vendeu centenas de milhares de unidades pelo mundo. Esse jogo usa uma linguagem visual mais sóbria, tem produção gráfica no padrão europeu e mantém o tema menos polêmico. Exceto pelo Jogo da Fronteira, todos os jogos apresentados ostentam os nomes dos autores na capa, sinal contemporâneo de *status*.

Dentro desse contexto, se você pensar o Jogo da Fronteira e o *Hart an der Grenze*, eu acho que formam um conjunto coerente com a época deles, sim, em termos de jogo, de produto, de regras e tudo mais. O *Robin Hood* é uma tiragem muito pequena... mas acho que *Robin Hood* e *Sheriff of Nottingham* são produtos mais coerentes com o momento deles... acho que sim! (SERGIO HALABAN, em Guimarães Neto, 2022, p. 155.)

Conclusão

O fio evolutivo seguido pela dupla de autores revela questões históricas, por exemplo, certas diferenças culturais e comerciais entre Brasil e Primeiro Mundo, especialmente uma demora na adoção de padrões culturais pós-modernos. A metodologia da História Oral permitiu uma construção colaborativa do discurso; prevalece a interpretação de quem escreve, mas o relato desses autores lá está para levantar novas questões.

A apresentação dos jogos como objetos em evolução técnica (isto é, das regras, da qualidade de produção) é um padrão estabelecido. Vimos que esse padrão merece, especialmente no mundo em desenvolvimento, uma contextualização crítica (as relações culturais e sociais de inserção de cada jogo no estado da arte da cultura lúdica).

Palavras-chave: História dos Jogos no Brasil; jogos de tabuleiro; Estudos Lúdicos; Filosofia da Linguagem; Filosofia dos Jogos.

Referências

- BENJAMIN, W. Sobre o conceito da história, em **Magia e técnica, arte e política**. São Paulo: Brasiliense, 1994.
- D'ENTREVES, M. P. Hannah Arendt, em Edward N. Zalta (ed.), **The Stanford Encyclopedia of Philosophy**. Fall Edition, 2022.
- FLUSSER, V. **O universo das imagens técnicas. Elogio da superficialidade**. São Paulo: Annablume, 2008.
- GUIMARÃES NETO, E. **Conceitos de museologia visando a curadoria de jogos digitais no Brasil**. Rio de Janeiro: Proceedings of SBGames 2019.
- GUIMARÃES NETO, E. **Além das regras do jogo: a redenção da alegoria nas fronteiras da Filosofia da Arte e da História da Cultura**. Tese de doutoramento. São Paulo: USP, 2022.
- ZATZ, A.; HALABAN, S. **Hart an der Grenze**. Stuttgart: Kosmos, 2006 [jogo].
- ZATZ, A.; HALABAN, S. **Jogo da Fronteira**. São Paulo: Estrela, 2005 [jogo].
- ZATZ, A.; HALABAN, S. **Robin Hood**. São Paulo: Galápagos Jogos, 2011 [jogo].
- ZATZ, A.; HALABAN, S. **Sheriff of Nottingham**. Indianapolis: Arcane Wonders, 2014 [jogo].

Notas

- [1] Essa necessidade é esmiuçada em meu artigo “Conceitos de museologia visando a curadoria de jogos digitais no Brasil” (Guimarães Neto, 2019).
- [2] Mais sobre o uso da Teoria Geral dos Sistemas como língua franca nas ciências, com exemplos na abordagem aos jogos, apresentei na tese *Além das regras do jogo* (Guimarães Neto, 2022), especialmente no capítulo 20, “A leitura sistêmica como paradigma”.

APLICATIVOS PARA O ENSINO- -APRENDIZAGEM DE LÍNGUAS INDÍGENAS DE POVOS DO BRASIL

SUELLEN TOBLER

Universidade Federal do Paraná

Introdução

Durante o II Seminário Viagens da Língua: Língua e tecnologias, apresentei um pouco da trajetória da implantação do primeiro aplicativo voltado para o ensino-aprendizagem de línguas indígenas de povos do Brasil, do qual sou autora, o Nheengatu app. Também aproveitei para incluir os projetos que surgiram inspirados neste, dos quais sou coautora, são eles o Nuke Tsãy app e o Kaingang app. Este conteúdo é ancorado em meu trabalho de pesquisa de mestrado, intitulado *Aplicativos para o ensino-aprendizagem de línguas indígenas de povos do Brasil: um estudo de caso sobre o Nheengatu app* (TOBLER, 2023), defendido em dezembro de 2023. Na dissertação, a linha de pesquisa é Políticas Públicas, assim, o embasamento teórico é o Ciclo de Políticas Públicas (SECCHI et al., 2022) e o Modelo de Múltiplos Fluxos (KINGDON, 2014).

Alguns pontos sobre a história do nheengatu

Para servir ao projeto colonial, na intenção de obter os conhecimentos acumulados sobre o território amazônico, e para evangelizar povos originários, os missionários jesuítas sistematizaram a escrita da língua falada pelos Tupinambá, que habitavam a região onde hoje é o Estado do Pará. Em reorganizações comunitárias, colocaram indígenas que falavam línguas de diferentes troncos linguísticos juntos com negros escravizados, todos foram obrigados a fazer uso da língua dos Tupinambá. Essa língua pertencente ao tronco tupi passou a ser conhecida entre os séculos XVII e XVIII como Língua Geral Amazônica (LGA), quando se tornou a principal língua falada na região. Ao ser difundida na região do Alto Rio Negro, a partir do século XIX, a LGA passou a ser conhecida como nheengatu, que significa fala boa, a boa língua, ou língua boa (FREIRE, 2011; BORGES, 1996).

A partir de 1750, Marquês de Pombal empreendeu uma reforma político-administrativa que teve como um de seus objetivos tornar o português a língua única da região amazônica. Para isso, expulsou os jesuítas da região e proibiu a fala do nheengatu em qualquer instância antes aplicada. Retirou a língua das escolas, das igrejas e proibiu novos nomes de cidades na língua. Seu uso tornou-se ilegal. Porém, estudos apontam que o ponto crítico para a redução de falantes da língua foi a Guerra da Cabanagem, também conhecida como Guerras das Cabanagens, que ocorreu entre 1834 e 1840. Posteriormente ocorreram novas reorganizações comunitárias e uma política baseada na imigração de monolíngues em português, com o surgimento do Ciclo da Borracha entre 1840 e 1912 (FREIRE, 2011; BORGES, 1996).

Aquela que foi a língua mais falada em todo o território do bioma amazônico, no momento presente está reduzida a três principais regiões. A região do Alto Rio Negro, no município de São Gabriel da Cachoeira, AM; a região do Baixo Amazonas, nos municípios de Parintins, Nova Olinda do Norte e Autazes, AM; e a região do Baixo Tapajós, nos municípios de Santarém, Belterra e Aveiro, PA. Dessa forma são consideradas três as principais variações da língua, o *nheengatu* do Rio Negro, o *nheengatu* do Baixo Amazonas

(ou *nheengatu* tradicional) e o *nheengatu tapajowara*. Conforme figura a seguir:



Figura 1.

Fonte: Reprodução (TOBLER, 2023, p. 31).

Durante os anos de 2019 e 2021 morei na região do baixo rio Tapajós, Santarém, PA. Cheguei ali devido a uma viagem de barco iniciando onde o rio das Amazonas recebe esse nome pela primeira vez, em Iquitos, no Peru (sabemos que sua nascente é antes disso), até onde ele encontra o oceano Atlântico, na Ilha do Marajó, PA. No ano de 2019 trabalhei ministrando oficinas de produção de conteúdo digital através de aplicativos gratuitos para celular, voltadas para jovens de comunidades ribeirinhas/extrativistas, aldeias e escolas municipais das periferias de Santarém, através de ONGs. Durante tais experiências, me conectei com lideranças comunitárias envolvidas em processos de retomada linguística, identitária e territorial, assim percebi a necessidade de produção de ferramentas para o registro e o apoio no ensino-aprendizagem de línguas indígenas (TOBLER, 2023). Irei discorrer sobre esta experiência a seguir.

O Nheengatu App

No ano de 2020, durante um trabalho na Aldeia São Francisco, em Santarém, PA, fiquei hospedada na casa da Dailza Assis Araújo, professora de nheengatu da escola indígena Suraraitá Tupinambá. Durante a estadia, Dailza me ensinou algumas palavras em nheengatu e na despedida me presenteou com o livro *Nheengatu Tapajowara*, produzido no Projeto de Extensão Curso de Nheengatu UFOPA/GCI. Na época eu estudava a língua alemã através da aplicação DuoLingo e, ao me interessar pela língua nheengatu, busquei cursos através de apps e não foi possível encontrar (MOURA; TOBLER, 2021). Aliás, naquele momento, não encontrei disponibilidade de aplicativos de ensino-aprendizagem voltados para línguas indígenas de povos do Brasil. Porém encontrei apps da bíblia judaico-cristã traduzida para inúmeras línguas maternas disponíveis na loja de apps Google Play Store (TOBLER, 2023).

Devido à epidemia de covid-19, foi sancionada a Lei 14.017/2020, conhecida como Lei Aldir Blanc. Em 2021, com recursos da Lei Aldir Blanc, o Estado do Pará lançou o Edital de Cultura Digital (Secult-PA; Instituto Ágata, 2020). Através do edital foram selecionados vinte projetos nas categorias mídia livre, intervenção

digital, formação em rede, novas mídias e arte digital. As análises das propostas seguiram os critérios de: inventividade, qualidade artística e relevância cultural do projeto; viabilidade técnica e orçamentária; coerência da proposta com a trajetória do proponente; difusão dos resultados para a comunidade; contribuição na promoção da acessibilidade em atenção à pessoa com deficiência e/ou pessoa idosa (TOBLER, 2023).

Enquanto estudante do curso de Antropologia da Universidade Federal do Oeste do Pará (UFOPA) e profissional do segmento de Tecnologia da Informação desde 2007, escrevi o projeto então intitulado Aplicativo Nheengatu para Android, que foi selecionado no Edital de Cultura Digital (Secult-PA; Instituto Ágata, 2020). Na execução, elaborei o levantamento de requisitos do sistema, documentação, os primeiros cinquenta exercícios e o desenvolvimento do código fonte. O edital foi o instrumento de política pública que viabilizou a implantação desta política pública, disponibilizada para o público no dia 1º de outubro de 2021 sob o nome de Nheengatu app (TOBLER, 2023).

O Nheengatu app é considerado o primeiro aplicativo voltado para o ensino-aprendizagem de línguas indígenas de povos do Brasil. A primeira versão disponibilizada foi delimitada para a variação linguística da região do Baixo Tapajós, o *nheengatu tapajowara*. O acesso à aplicação é gratuito; por meio da URL *nheengatu-app.web.app*, é necessário realizar um cadastro que requer os seguintes dados: nome, e-mail, senha e aceitação do termo de compromisso, que inclui o uso dos dados em pesquisas acadêmicas. Ao efetuar o *login*, o usuário tem acesso a uma lista de exercícios divididos em grupos de cinco exercícios cada. Ao final de cada grupo, é mostrado em tela a quantidade de respostas corretas e equivocadas, passando então para o próximo grupo automaticamente. Sua interface é baseada no Duolingo, o aplicativo de línguas mais baixado da Google Play Store (GOOGLE, 2023). Ao todo são seis modelos de exercícios:



1. exercício de memória

2. exercício de selecionar palavras

3. exercício de completar frase



4. exercício de selecionar frase

5. exercício de escrever frase

6. exercício de formar pares

Figura 2.

Fonte: Adaptação de imagem reproduzida (TOBLER, 2023, p. 63).

A aplicação é uma Progressive Web App (PWA), isso permite maior flexibilidade no acesso ao sistema. O Nheengatu app pode ser executado através de qualquer navegador *web* sem a necessidade de ser instalado e, ao mesmo tempo, ele é compatível para instalação em qualquer sistema operacional. Dessa forma, pode ser instalado ou não, em qualquer dispositivo, como celular, *tablet*, *chromebook*, *notebook* e *desktop*. A linguagem de programação utilizada para atingir esse resultado foi o JavaScript (TOBLER, 2023).

O número total de usuários entre 1º de outubro de 2021 a 15 de setembro de 2023 foi de 2.214. Este número não representa a quantidade de usuários únicos, uma vez que há o uso coletivo do app através de projetores dentro de algumas escolas indígenas da região do Baixo Tapajós, na disciplina de nheengatu. Desses 2.214 usuários, 443 são os que se cadastraram, mas não executam exercícios, e 339 concluíram todos os exercícios disponíveis. Dos usuários que praticam os exercícios, o saldo de respostas corretas é de 78.131, enquanto o de respostas equivocadas é de 10.514 (TOBLER, 2023).

O Nheengatu app serviu de inspiração para novos projetos, um deles é o Nuke Tsãy app, que contempla variação linguística pertencente ao povo Shanenawa, da Terra Indígena (TI) Katukina/Kaxinawá, município de Feijó, AC, com lançamento previsto para abril de 2024. O outro, é o Kaingang app, com a variação linguística das TIs Toldo Chimbangue, Aldeia Kondá e Toldo Pinhal, do município de Chapecó, SC, que está com a primeira fase, de um total de três, em andamento (TOBLER, 2023).

Palavras-chave: línguas indígenas; nheengatu; Tecnologia da Informação e da Comunicação (TIC); aplicativo para celular; políticas públicas.

Referências

- BORGES, Luiz C. O nheengatu: uma língua amazônica. **Papia**, v. 4, n. 2, p. 44-55. CNPq/Museu de Astronomia e Ciências Afins do Rio de Janeiro, 1996.
- FREIRE, José Ribamar Bessa. **Rio Babel: a história das línguas na Amazônia**. Rio de Janeiro: EdUERJ, 2011.
- GOOGLE. Google Play Store. **Site**, 2023. Disponível em: <https://play.google.com/store/search?q=idiomas&c=apps>.
- KINGDON, John W. **Agendas, Alternatives, And Public Policies**. 2. ed. Harlow: Pearson Education, 2014.
- MOURA, Beatriz Martins; TOBLER, Suellen. **Aplicativo Nheengatu para celular**. VIII Seminário de Graduação da IX Jornada Acadêmica da UFOPA. Santarém, 2021.
- SECCHI, Leonardo; COELHO, Fernando de Souza; PIRES, Valdemir. **Políticas públicas: conceitos, casos práticos, questões de concurso**. 3. ed. São Paulo: Cengage, 2022.
- SECULT-PA; INSTITUTO ÁGATA. **Editais de Cultura Digital – Lei Aldir Blanc Pará**, 2020. Disponível em: <https://institutoagata.com.br/event/edital/>.
- TOBLER, Suellen. **Aplicativos para o ensino-aprendizagem de línguas indígenas de povos do Brasil: um estudo de caso sobre o Nheengatu app**. Dissertação (Mestrado) – Universidade Federal do Paraná, 2023. Disponível em: <https://siga.ufpr.br/siga/visitante/trabalhoConclusaoWS?idpessoal=162923&idprograma=40001016081P3&anobase=2023&idtc=19>.

A METAMORFOSE DIGITAL DOS CÓDIGOS LÚDICOS

GILSON SCHWARTZ

Universidade de São Paulo

Poder-se-ia dizer que o conceito de “jogo” é um conceito de contornos pouco nítidos. Mas um conceito pouco nítido é ainda um conceito? Um retrato difuso é ainda a imagem de um homem? Pode-se sempre substituir com vantagem uma imagem difusa por uma imagem nítida? Não é muitas vezes a difusa aquela de que nós precisamos?

Ludwig Wittgenstein, *Investigações filosóficas*.

- *Papai, que é que eu faço?*
 — *Eu já lhe disse: vá brincar e me deixe!*
 — *Mas eu já brinquei, juro. Papai riu:*
 — *Mas brincar não termina...*
 — *Termina sim.*
 — *Invente outro brinquedo.*
 — *Não quero brincar nem estudar.*
 — *Quer fazer o quê então? Joana meditou:*
 — *Nada do que sei...*
 — *Quer voar?, pergunta papai distraído.*
 — *Não, responde Joana. — Pausa. — Que é que eu faço?*
Papai troveja dessa vez:
 — *Bata com a cabeça na parede!*

Clarice Lispector, *Perto do coração selvagem*.

Metamorfose cartesiana

Muito antes de Descartes e dos computadores em rede, alguns dos principais momentos da história do pensamento ocidental na Antiguidade enquadraram a Educação e os jogos no projeto da pólis. Desde então, nada mudou. Pensar o jogo, o lúdico e a *gamificação* nessa perspectiva ainda é o melhor antídoto para as visões ingenuamente cartesianas e calculistas que em última análise serviram (e ainda servem) para a manutenção de mecanismos instrucionais e correccionais contemporâneos, azeitando e acelerando o desempenho das plataformas de exploração capitalista da energia humana.

Como modo de produção e exploração, o capitalismo digital do antropoceno (o capitaloceno) leva literalmente ao infinito (pela multiplicação acelerada de processadores gastando energia natural em escala exponencial) a conversão de energia em geral em máquinas de explorar a energia vital humana em seus diferentes modos de consciência e atenção ao espaço, ao tempo e aos propósitos, individuais ou coletivos. No cinema, essa dualidade entre corpo,

espírito e dominação ganhou visibilidade impecável do ponto de vista estético com “Matrix”.

Essa matriz sistêmica é um panóptico digital emergente, contra a qual ressurge a utopia de uma “pedascopia” lúdica. A arquitetura da informação nos jogos inspira novas linguagens e gêneros, pode ser também a fonte de novas matrizes para re-desenhar ensino e aprendizagem, superando os modelos fabris, prisionais e hospitalares.

Penso, logo existo. A máxima do filósofo francês René Descartes (1595-1650) que marcou uma revolução na história da filosofia expressa uma síntese e reflete a consolidação no século XVII do racionalismo, do método experimental e da explicação mecânica e matemática do universo. Em poucas palavras, essa visão do mundo está na base de todas as ciências modernas e, portanto, da própria escola.

O cartesianismo marcou não apenas a emergência de um modo de vida centrado em um sujeito do cálculo cuja vocação é o controle das coisas e dos seres humanos (que também viram coisas calculáveis) como também nos forma e determina os modos de aprender e ensinar. Passamos a existir conformados com esse pensar soberano que subordina a tudo e a todo e qualquer fazer, reservando o brincar a tempos e espaços igualmente controlados. Ao fim do processo, aguardam-nos promessas de luzes, clareza e distinção. Ainda que a escola mais pareça uma fábrica e a sala de aula, uma prisão.

Sobre esse fundamento construímos escolas e vidas cartesianas. Pensamos de uma determinada forma, seja qual for o conteúdo. A própria noção de que o pensamento, a ciência e a vida podem ser compreendidos apenas a partir de um sujeito solidamente ancorado na lógica, em oposição às tradições do mundo medieval, já expressa a convicção de que nossas vidas ganham sentido apenas se estiverem amparadas por um princípio seguro e científico, anterior à própria linguagem.

O caminho rumo à ciência e ao cálculo, em oposição à regulação da vida pela tradição e pelos afetos, constitui um roteiro que

nos leva da dúvida sistemática à certeza de que existimos como sujeitos pensantes que podem duvidar de tudo, menos da autonomia do pensar que compreende a si mesmo para poder alcançar a totalidade do mundo. Como introduzir a brincadeira em um cenário tão estruturado, fechado nas competências racionais de um sujeito que encontra sua autonomia no império da razão?

Brincar, jogar, divertir-se muito frequentemente significa abrir-se ao inusitado, ao acaso, a uma dinâmica sem muitas certezas que admite a dispersão, a errância que não equivale necessariamente ao erro, à diversidade e à diferença que zombam da estabilidade existencial de todo e qualquer sujeito, levando ao limite a crítica do mundo, mais que sua compreensão fundada em um princípio sólido ou aparentemente substancial.

Embora haja inúmeras brincadeiras solitárias, o jogo encontra sua expressão mais acabada em situações em que jogo “com alguém” e “contra alguém”, portanto, as decisões já não dependem de um sujeito racional plenamente autônomo e consciente de todas as regras do método. Além de burlar as regras, quais outras dimensões da linguagem (corporal, paisagística, religiosa) atuam na decisão sobre burlar ou não, resistir ou não, desafiar ou não, até o limite do terror e da destruição da própria autonomia de cada “jogador”? Jogamos ou somos jogados, eis a questão.

Quando se diz que “o futebol é uma caixinha de surpresas”, fica evidente que a situação de jogo, mesmo que resulte de regras muito claras e imutáveis, é uma abertura para o imprevisível, para a criatividade e para a mobilização de competências que não se limitam ao espaço autocontido da razão.

A brincadeira é uma desrazão certa, eventualmente a loucura com método. É paradoxo, charada, aporia ou contradição, abertura ao movimento mais que âncora intelectual de uma solidão coletiva onde encontramos nos outros sempre o mesmo Eu.

Como introduzir a brincadeira na razão? Que brincadeiras são essas, trazidas por tecnologias de informação e comunicação, os jogos digitais, que seriam elas mesmas impensáveis sem a consolidação da calculabilidade, da técnica, da indústria do *software*, da

engenharia eletrônica e do *design* instrucional que, em última análise, antevê e regula todas as possibilidades e resultados do jogo, como nos megacomputadores programados para vencer partidas de xadrez com os maiores campeões que são “apenas” humanos?

Vivemos hoje sob o efeito onipresente dos *videogames* (ou simplesmente *games*, mais um termo da língua inglesa que se integra à língua portuguesa). São brinquedos, jogos (ou “videojogos”) que mobilizam nossa imaginação, alteram nossa identidade individual e coletiva, interferem em nossos afetos e emoções, porém sempre a partir de uma calculabilidade aperfeiçoada a cada onda de inovação tecnológica.

Seriam os *games* a etapa mais avançada e recente da racionalização do mundo, da vida e do pensamento ou estamos diante de uma inédita carnavalização do cotidiano que alguns já denominam *gamificação* da sociedade? É a racionalidade que submete mais uma vez a diversão ao cálculo ou é o cálculo que finalmente se abre para a surpresa, a diversão e até para a subversão do método cartesiano? Os “joguinhos” invadem as salas de aula e laboratórios de informática para condicionar de vez o aprendizado e robotizar a prática pedagógica a exemplo da automação industrial, bancária e do lazer de massa?

Linguagem e Pedagogia em universos lúdicos

É possível uma pedagogia criativa e emancipatória amparada não apenas no jogo, mas no jogo eletrônico? Que novas tensões complicam o trânsito entre motivação e automação, cópia e invenção, condicionamento a regras e reinvenção de propósitos?

Esses *games* e brincadeiras são uma fonte inédita de criatividade, novas competências e abertura para a surpresa ou não passam de “caixas-pretas” tecnológicas desenvolvidas por uma indústria que já supera a do cinema e levará, como em todas as iniciativas de produção de massa, voltadas ao consumo em larga escala, a uma homogeneização da experiência educacional, à morte do professor e ao condicionamento de alunos, preparados metuculosa e maquiavelmente a se submeter cegamente às regras de um jogo capitalista, global e inumano?

Vencerá afinal e ainda o cartesianismo ancorado em um indivíduo calculista e racional? Surgirão novas sensibilidades e ganharão mais espaço e tempo nas escolas as atividades lúdicas, recreativas e artísticas inspiradas em práticas pedagógicas voltadas à formação de novos indivíduos e grupos capazes de integrar o pensar, o fazer e o brincar? Em uma sociedade marcada pela economia da informação e do conhecimento, seria essa integração transdisciplinar a nova medida de todas as coisas sem a qual a inovação perde vigor e até o consumo torna-se insustentável? Jogar é uma nova forma ou seria em última análise a única forma de se comunicar de modo livre e criativo?

Esses dualismos não surgiram com os *games*, os computadores e as redes digitais. Professores e outros profissionais engajados na Educação sempre enfrentaram essas contradições – e há muitos fatores que a rigor empurram o brincar eletrônico de fato apenas para servir de antecipação, na criança e no jovem, do condicionamento para o lazer industrial e controlado no qual cada um acaba enquadrado ao longo de uma vida digital em que o prazer é programado e o gozo, calculado, simulado e controlado. E pode virar vícios e vírus. Viciar, corromper, variar – não era essa a acusação de Atenas contra Sócrates pelo uso desabusado do jogo de linguagem nos seus diálogos?

Com ou sem *games*, há de fato uma indagação de fundo sobre o real fundamento ou mesmo sobre as implicações sociais, culturais e políticas de distinções que consagram a separação cartesiana entre sujeito e objeto. Mas a observação do jogo acaba revelando que seu decifrar consiste justamente em criar jogos entre sujeito e objeto.

Linhas de falha entre identidades e Inteligências Artificiais

Nos tempos da telefonia analógica, uma linha telefônica muitas vezes era sinônimo de espera. Em muitas localidades, uma telefonista era responsável por “dar linha”, ou seja, gerenciar a distribuição das linhas telefônicas a partir de um painel com inúmeras entradas para cabos.

Para fazer uma chamada em uma central telefônica local, ia-se à cabine quando a telefonista (em geral, uma mulher, *operator*) assim determinava.

A espera podia durar horas, a fila de espera também era enorme para adquirir uma linha telefônica, as “linhas” eram objeto de especulação financeira e o telefone afinal ganhou no século XX o *status* de mais uma demonstração de propriedade privada de uma tecnologia de comunicação.

Isso não é novidade! Afinal, foi o que já ocorreu com os aparelhos de rádio, que foram diminuindo de tamanho até se tornarem portáteis e à pilha. A mesma lógica prevaleceu para os televisores (de móveis na sala para fruição familiar aos aparelhos individuais nos quartos de cada membro da família).

Na atual transformação digital, conseguir “linha” equivale a “acessar a rede”, as “operadoras” e seus sistemas de conexão analógica assumiram a forma de gigantescos conglomerados empresariais (as Big Tech são o Big Brother). Na pesquisa mais recente sobre uso da internet no Brasil, quase 60% da população continua sem acesso significativo à rede mundial.

Conseguir interagir na rede não é apenas uma questão de acesso físico ou posse de um aparelho (do celular aos sensores e câmeras onipresentes aos óculos para imersão em realidades estendidas).

As falhas de conexão de ordem educacional, cognitiva, territorial e identitária criam filas de espera, linhas de fronteira e filtros de acesso que impõem novas exclusões.

Embora acelerado e globalizado, o acesso e a partilha de saberes ainda são desiguais e extremamente individualizados, a comodidade da tecnologia portátil e privada paga o preço de permitir apenas um acesso insuficiente de cada cidadão a uma internet mais criativa e colaborativa.

A perspectiva de uma esfera pública digital e democrática corre riscos.

Com a complexidade crescente dos sistemas e de suas formas de controle e governança, o acesso é também condicionado pelo

letramento digital necessário para que se alcance uma emancipação cidadã por meio da vida em rede.

A violência cibernética é uma dimensão essencial da guerra e da paz, espelha outras violências e exclusões físicas e simbólicas.

É preciso reconfigurar a iconomia das redes digitais para articular identidade, inteligência e inovação de modo transparente, pró-ODS e democracia.

As competências para interagir com bases de dados, trabalhar e socializar são cada vez mais complexas e assimétricas, aumentam os riscos de exclusão, manipulação e passividade mesmo quando o acesso é individual, portátil e “customizado” (ou seja, sujeito a um algoritmo de fidelização do cliente).

As dimensões do corpo, do fazer tangível, do participar na rua e na cidade, parecem amortecidas por bolhas virtuais de esquecimento do corpo e da memória, culto ao ódio e incapacidade para a multiplicação dialógica.

As redes digitais são uma tessitura feita de linhas de conexão praticamente infinitas e infinitamente expostas a falhas de conexão física, social e cultural.

As linhas e as redes articulam desde o cotidiano de cada indivíduo até as inúmeras configurações de uso coletivo ou colaborativo e de acesso a bens, serviços e direitos.

Os espaços e tempos multiplicam-se em *lives*, encontros em salas de reunião virtuais e, cada vez mais, metaversos e novas formas de imersão e representação.

Mas... quais são as contradições e potenciais gerados por linhas de conexão em alta velocidade capazes de configurar bolhas de consumo, ideologia e entretenimento condenando cada indivíduo ao isolamento na massa e na multidão?

Como fazer uma reengenharia crítica dessas tecnologias dos pontos de vista cognitivo, epistemológico e existencial, vital, experimental? Quais as novas formas do brincar e da coexistência criativa entre diferentes identidades?

É possível inteligência sem identidade? É aceitável o identitarismo sem inteligência? Quais os poros do sistema e quais

epistemologias podem abrir caminho para reconfigurar as relações entre o humano, a cultura e a natureza?

Quem é sujeito, quem é objeto nas esferas e bolhas digitais?

O que significa estar conectado, ou seja, *on-line* (“em linha”) e ao mesmo tempo ser excluído, confinado, segregado, monitorado, monetizado e certificado por “operadoras” de governança, transparência, publicidade e ética opacas, quando não são explicitamente postas a serviço da perpetuação da desigualdade?

Diante de rupturas sociais e políticas, entre a consolidação e a universalização da Inteligência Artificial, quais são as situações *on-line* que reproduzem ou agravam condições de insalubridade, trabalho indigno, ameaças à saúde mental e emergência de novas formas de repressão e vigilância, racismo, negacionismo climático e xenofobia?

Em que medida as práticas de “cancelamento” *on-line* recriam em forma complexa e incontrolável uma nova versão do *panopticum* estudado por Foucault?

Quais modelos de legitimidade *on-line* são associados a novas formas de estar “na linha” ou, como na expressão em português, “andar na linha”?

As linhas do jogo democrático

No Brasil, tornou-se conhecida a expressão “jogar dentro das quatro linhas” da Constituição, aproximando-se também o alinhamento à garantia de que são seguidas as regras do jogo. Questões de identidade (local, étnica, racial, religiosa, política, cultural ou econômica e informacional) colocam-se nos interstícios e nas dobras dos algoritmos supostamente automáticos, neutros e inteligentes.

Mas também se multiplicam as denúncias de manipulação, apagamento ou radicalização de identidades “em linha” (*on-line*).

Estar alinhado a valores grupais ou mesmo tornar-se dependente das redes como quem é apanhado em um emaranhado de linhas desenhado para aprisionar, alienar ou matar tornam-se condições em que identidade e inteligência estão ambos contrapostos e muitas vezes indissociáveis de novas formas de inteligência.

Em “Identidades x Inteligências: Artificiais”, pesquisadores, lideranças comunitárias e do mundo da arte popular discutem essas questões e se colocam a partir de perspectivas identitárias no contexto da transformação digital.

Do quilombo ao *kibutz*, da aldeia ao templo, da cidade ao campo, questões de ordem étnica, cultural, política e religiosa produzem linhas cruzadas, interferências humanas em universos fechados por algoritmos, vigilância, violência (material e simbólica) e segregação (territorial, digital e material).

É preciso reescrever as linhas de programação que definem os modelos de linguagem usados pela Inteligência Artificial para retomar o horizonte da esperança lúdico, participativo e imune aos exageros do cartesianismo colonial.

Palavras-chave: iconomia; capitaloceno; edugamificação; jogos; linguagem.

INTELIGÊNCIA ARTIFICIAL E LINGUAGEM

JOÃO CARLOS PEREIRA DA SILVA
Universidade Federal do Rio de Janeiro

No final de novembro de 2022, foi lançado o programa ChatGPT (*Chat Generative Pre-Trained Transformer*), que de imediato ganhou as manchetes e atraiu grande interesse do público, tornando-se o aplicativo de internet que conseguiu atingir mais rapidamente (em apenas dois meses) a marca de 100 milhões de usuários ativos. Na esteira de seu sucesso, outras grandes empresas de tecnologia lançaram suas próprias Inteligências Artificiais Generativas, que são programas capazes de gerar textos, áudios e imagens (e agora até vídeos).

Grande parte do sucesso desse tipo de ferramenta junto ao público vem da sua capacidade de *bater papo (chat)* com as pessoas e de produzir textos sobre os mais variados assuntos. Tal sucesso se refletiu no grande número de reportagens que apresentavam casos de uso dessas ferramentas. Elaborar relatórios

ou textos acadêmicos, preparar uma apresentação, pedir dicas de investimentos, conversar sobre seus problemas e até preparar sermão são alguns exemplos do que foi relatado na imprensa. O escopo dessas reportagens se mostrou tão amplo que até a saída e o retorno do CEO da empresa que criou o ChatGPT mereceu um acompanhamento semelhante ao de um *reality show*.

A Inteligência Artificial é uma área da computação que teve seu início na década de 1950. Em uma definição ingênua, podemos dizer que a Inteligência Artificial é a inteligência produzida pelo ser humano, e não pela natureza. Isso nos leva à questão do que é a inteligência (natural). Embora não exista um consenso entre os especialistas do que seja inteligência, aprender pela experiência, adquirir e reter conhecimento, raciocinar, planejar, identificar semelhanças e diferenças entre objetos, ser capaz de estabelecer relacionamentos, compreender e se expressar através da chamada linguagem natural, identificar e resolver problemas são algumas das habilidades que a caracterizam. Portanto, uma Inteligência Artificial para ser considerada inteligente deverá ter estas habilidades. Mas esta não é, de fato, uma resposta definitiva, uma vez que ainda há controvérsias sobre se uma máquina pode replicar a inteligência humana.

A pergunta se seria possível construir uma máquina capaz de pensar foi feita pelo matemático inglês Alan Turing (TURING, 1950), que é considerado um dos pais da Computação e da Inteligência Artificial. Se você assistiu ao filme “O Jogo da imitação” (2014), conheceu a participação de Turing no esforço para quebrar os códigos secretos nazistas que eram enviados através das máquinas Enigma durante a Segunda Guerra Mundial.

Curiosamente, o título do filme é uma referência ao jogo proposto por Turing (TURING, 1950), no qual ele discute se uma máquina pode *pensar*. Neste jogo temos três jogadores: A, B e C. O jogador C fica em um local diferente dos outros jogadores e se comunica com eles através de textos datilografados (funciona se você usar mensagem de texto via celular). O jogador C deve identificar qual é o gênero dos jogadores, sendo que o jogador A responde às perguntas

tentando enganar o jogador C, enquanto o jogador B responde às perguntas tentando ajudar C na identificação de quem é quem. Turing propôs substituir o jogador A por uma máquina e mudar a pergunta “As máquinas podem pensar?” pela pergunta “Será que o jogador C erra com a mesma frequência de quando o jogo é jogado com duas pessoas?”. Hoje, este jogo é conhecido como Teste de Turing, e um programa capaz de *imitar* um ser humano seria aprovado neste teste.

Dos vários aspectos interessantes existentes na proposta de Turing, chama a atenção o fato de que máquinas pensantes só seriam reconhecidas como tal se fossem capazes de estabelecer uma comunicação conosco em nosso idioma. A linguagem nos permite estabelecer uma forma de nos comunicarmos uns com os outros e é frequentemente citada como um diferencial com relação às outras espécies animais. Ao lado da área de Jogos (como o xadrez) e de Prova Automática de Teoremas, o Processamento de Linguagem Natural foi uma das primeiras áreas de investigação da Inteligência Artificial.

Na década de 1960, surgiram os primeiros trabalhos de sucesso na área de Processamento de Linguagem Natural. O primeiro programa de computador que permitia uma comunicação através da linguagem natural entre humanos e computadores era chamado ELIZA (WEIZENBAUM, 1966) e foi desenvolvido por Joseph Weizenbaum, do Massachusetts Institute of Technology (MIT). O funcionamento do programa era bastante ingênuo: um texto era digitado e o programa procurava identificar a existência de certas palavras-chaves. Quando uma palavra-chave era encontrada, uma regra associada a essa palavra era aplicada de forma a transformar a frase apresentada em uma frase de resposta. Caso nenhuma palavra-chave fosse encontrada, alguma resposta prévia ou uma frase mais generalista era gerada. Esse programa foi o precursor dos atuais sistemas de atendimento automático, hoje tão comuns em diversas empresas. Outro trabalho relevante foi o sistema de resposta a perguntas chamado SHRDLU (WINOGRAD, 1972). O sistema também permitia ao usuário instruir o programa a realizar tarefas

(como empilhar e desempilhar blocos) em um ambiente virtual controlado chamado *mundo dos blocos*.

Os desenvolvedores desses sistemas precisavam explicitar todo conhecimento necessário para a resolução do problema: construir uma gramática capaz de verificar se uma frase estava escrita corretamente, além de produzir frases gramatical e semanticamente corretas; construir uma lista de palavras-chaves; identificar as palavras do texto pertencentes à lista de palavras-chaves; ser capaz de identificar eventuais mudanças no contexto das conversas etc. Os desafios eram tão grandes que os sistemas propostos ficavam limitados a temas específicos ou modelos muito simples, inviabilizando que tais programas fossem aplicados a situações do mundo real. Assim, era necessário mudar a maneira como esses sistemas eram construídos. Ao invés de *dizer* à máquina o que ela deveria fazer, era necessário que a máquina *aprendesse* o que fazer.

A área de Aprendizado de Máquina é uma subárea da Inteligência Artificial dedicada ao desenvolvimento de programas que dão ao computador a capacidade de aprender. As Redes Neurais Artificiais (RNA), cujo primeiro modelo foi proposto por McCulloch e Pitts (McCULLOCH, W.; PITTS, W., 1943), são um desses modelos computacionais e, como o nome indica, são inspirados nas Redes Neurais Biológicas. A primeira RNA, chamada MADALINE (Multiple ADaptive LINear Elements), foi aplicada a um problema do mundo real em 1959, quando, através de um filtro adaptativo, era capaz de remover ruídos de uma linha telefônica.

Para você ter uma ideia de como uma RNA aprende, imagine que você quer ensiná-la a reconhecer letras do alfabeto. O que você vai precisar é de um conjunto de imagens de letras escritas por diferentes pessoas e um rótulo (gabarito) indicando qual letra está escrita em cada imagem. Cada imagem é apresentada à rede neural, que responderá dizendo qual letra foi apresentada. Inicialmente, a chance de sua rede errar será grande, pois ela nunca viu uma letra, mas esses erros são usados para ajustar as conexões da RNA. Estas conexões entre os neurônios, chamadas de pesos, são representadas por valores numéricos que indicam quão forte

é uma conexão entre dois neurônios. Assim, à medida que você vai apresentando as imagens (e as mesmas imagens são apresentadas diversas vezes), a rede vai ajustando os pesos e aumentando sua acurácia, fazendo com que ela aprenda quais são as características que fazem um “A” ser um A, um “B” ser um B etc. É importante chamar a atenção para o fato de que se seu conjunto de imagens for formado apenas por letras bem escritas, sua rede terá grande dificuldade de identificar os garranchos. Por isso a importância de sua base ter diversidade, conter diferentes formas e estilos de escrita para cada uma das letras do alfabeto.

Para ensinar um computador a gerar um texto coerente, precisamos de algo mais potente do que a rede que usamos para aprender a reconhecer letras. Precisamos do chamado aprendizado profundo, que tem uma estrutura bem mais complexa que nossa rede anterior e nos permite capturar características bem mais abstratas do que aquelas que diferenciam um “A” de um “O”. Antes de vermos de onde vêm os dados textuais que serão utilizados para ensinar um computador a gerar textos, vamos, por meio de uma brincadeira, dar uma ideia de como os chamados Modelos de Linguagem geram esses textos.

Imagine cinco crianças alinhadas. Você vai dizer algo (uma palavra ou frase) e, a partir do que foi dito, as crianças devem contar uma história. Só que a história deve ser contada de uma maneira específica: cada criança poderá falar apenas uma palavra por vez, uma após a outra. Depois que a quinta criança falar a quinta palavra, a vez retorna para a primeira criança que fala a sexta palavra e assim por diante, até que a história tenha terminado.

Nesse jogo, as cinco crianças fazem o papel do Modelo de Linguagem. A palavra ou frase que você diz estabelece o contexto ou tema a ser abordado. Baseado no que foi dito e na experiência e/ou conhecimento de cada criança sobre o assunto, a história terá um desenvolvimento diferente. Por exemplo, se a palavra dita for *praia*, uma primeira sequência de palavras possível poderia ser formada por: *ontem, fui, à, praia, de*. A primeira criança teria agora de dizer a sexta palavra. Perceba que além do contexto inicial, essas

cinco palavras abrem algumas possibilidades de qual poderia ser esta sexta palavra. Por exemplo, *bicicleta*, *Copacabana* ou *chinelo* completam a frase de maneira satisfatória e provavelmente seriam ditas com muita frequência. Opções talvez não tão comuns poderiam ser *branco* ou *cachorro*. Se uma delas for dita, qual palavra você diria em seguida? E se a sexta palavra for *mim*, você conseguiria continuar a história?

Já se sua frase inicial for *gravidade quântica*, ou você vai acabar com a brincadeira, uma vez que as crianças não sabem o que é isso, ou elas poderão dizer “O que é gravidade quântica?” e esperar que você dê mais informações a respeito, embora eu acredite que isso não vá ajudar muito. Provavelmente, você terá que substituir as cinco crianças por cinco físicos que entendam do assunto, o que pode ser um pouco difícil de encontrar. Colocar cinco adultos com nenhum ou pouco conhecimento de Física pode gerar histórias que até tenham algum sentido, mas com muitas informações erradas a respeito do tema.

Desta forma, o desenrolar do jogo está fortemente vinculado às pessoas escolhidas e a seus conhecimentos e/ou experiências de vida. Pessoas com uma mesma visão de mundo tenderão a contar as histórias que reflitam esta visão. Selecionar um grupo mais diverso, com experiências distintas, muito provavelmente fará com que as histórias sejam mais ricas e variadas. De certa forma, é isso o que ocorre com os Modelos de Linguagem, tendo seu desempenho vinculado ao tipo de informação usada em seu processo de aprendizado.

Para que o processo de aprendizado profundo seja efetivo, precisamos de um grande volume de dados textuais. No caso da construção dos Modelos de Linguagens usados pelas Inteligências Artificiais Generativas, a Wikipédia é uma das principais fontes de dados textuais. Sendo um repositório de textos construído de forma colaborativa, ela apresenta certas características que devem ser consideradas quando usamos seus artigos para construir tais modelos.

A primeira é a quantidade de textos por idiomas. Dados da própria Wikipédia indicavam que, em dezembro de 2023, existiam 6.752.910 milhões de artigos em inglês (idioma com maior número de artigos) e 1.114.259 milhão de artigos em português, fazendo com que nosso idioma ocupasse a 18ª posição entre todos os idiomas da Wikipédia, com aproximadamente 2.5% do total de artigos. Portanto, quanto mais artigos existirem para um determinado idioma, mais informações estarão disponíveis, o que permite a construção de modelos de linguagem mais abrangentes.

O segundo ponto é a qualidade dos textos. É provável que você já tenha procurado algo na Wikipédia e tenha encontrado um artigo em português com pouca ou nenhuma informação, tendo que recorrer ao artigo correspondente em inglês ou em algum outro idioma. Portanto, há uma variação no valor informativo que é fortemente dependente do idioma e dos colaboradores que alimentam a Wikipédia. Outro fator a considerar é que existe um desbalanceamento dos assuntos presentes na Wikipédia, na qual alguns tópicos, como música ou esporte, tendem a ter uma maior quantidade de artigos. Aqui também há uma variação dos tópicos que é dependente do idioma considerado. Além disso, o perfil dos colaboradores, com maior presença de pessoas jovens e do sexo masculino, com residência em países desenvolvidos, resulta em uma visão de mundo enviesada e com pouca diversidade. Caímos em um problema semelhante ao nosso jogo de contar histórias, que ficam dependentes de quem são os escolhidos para contá-las. Nosso aprendizado será limitado.

Todo este trabalho de coleta dos dados é bastante complexo e não pode ser feito uma única vez. Novas informações sobre fatos até então desconhecidos surgem a todo momento e alimentam estas fontes primárias de dados textuais, fazendo-se necessária uma nova coleta de informação para que os modelos de linguagem possam ser atualizados. Antes de 2020, a covid-19 não existia e um modelo treinado sem esta informação não poderia dizer nada sobre ela. Sendo um evento global há uma tendência a que ele seja registrado em diversos idiomas. Mas não podemos dizer o mesmo quanto ao

registro de eventos, pessoas ou elementos culturais de menor alcance, que ficam restritos a alguns idiomas ou nem são registrados.

Uma nova coleta de dados é necessária e implicará em refazer todo o custoso processo de documentação e curadoria dos dados para que possa ser usado em um novo treinamento do modelo. Visando fomentar maior equidade de conhecimento através da proposição e do desenvolvimento de soluções que visam diminuir, entre outras, lacunas culturais, geográficas e linguísticas existentes, a Wikipédia possui uma iniciativa chamada Wikipedia Diversity Observatory. Sendo bem-sucedida, poderá levar à redução dos problemas aqui apontados.

Além da Wikipédia, outras fontes de coleta de dados textuais são as redes sociais, *blogs* e repositórios de dados como o Common Crawl¹, que coleta dados da internet e os disponibiliza para pesquisas. Em particular, no caso das redes sociais, o tratamento e a curadoria dos dados são mais complexos. Fatores como grupos de usuários que preferem certas mídias sociais, as diferentes formas de se expressar (com a utilização de *emojis*, frases e/ou palavras abreviadas), a influência do grupo no comportamento do indivíduo, as mudanças de comportamento ao longo do tempo e a facilidade de distribuição de informações não factuais e/ou com viés racial, de gênero, orientação sexual, religioso e cultural devem ser considerados quando os dados coletados destas fontes são utilizados na construção de Modelos de Linguagem. Uma das primeiras críticas feitas às Inteligências Artificiais Generativas foi justamente o fato de que diversos textos produzidos não tinham embasamento em fatos, sendo observada até a geração de referências bibliográficas que não existiam. É o Modelo de Linguagem gerando desinformação.

Fica claro, portanto, a influência que os colaboradores e usuários destes sistemas têm nos resultados obtidos pelas Inteligências Artificiais Generativas quando os dados textuais usados no aprendizado não forem feitos de forma criteriosa. O fato de ser possível coletar um grande volume de informações textuais não garante a diversidade e a qualidade da informação. Esse é um dos grandes desafios para o aprimoramento desses sistemas, sendo necessário

um investimento significativo nas pessoas que atuam como produtores, moderadores e curadores dos textos e no desenvolvimento de ferramentas de Inteligência Artificial que permitam que este tratamento ocorra de forma automática ou semiautomática.

E o próprio modelo, uma vez construído, não representa o fim dos problemas. Se você faz uma pesquisa na internet, ao apresentar suas respostas em uma determinada ordem, a ferramenta de busca acaba induzindo seu comportamento e a informação que você irá consumir. Pense em quantas vezes você clicou em um *link* que não estivesse na primeira página de resposta de sua ferramenta de busca. Muito provavelmente você clicou nos dois ou três primeiros *links* e não é por outro motivo que os *links* pagos aparecem nestas posições.

Temos então um ciclo que se retroalimenta: dados produzidos pelas pessoas são coletados e utilizados na construção das Inteligências Artificiais Generativas, que produzem informações consumidas pelas pessoas para a produção de novos dados que serão capturados e utilizados em um novo processo de treinamento desses modelos. Esse processo nos levará a uma homogeneização do conteúdo produzido e consumido, restringindo ainda mais a participação de grupos sociais e ideias com pouca representatividade.

Todos esses problemas estão relacionados à construção destes Modelos de Linguagem. Mas existem também preocupações com o avanço da Inteligência Artificial de forma mais ampla e seus impactos socioeconômicos que já ocorrem e são de enorme magnitude. Listas de profissões condenadas à extinção em um curto período de tempo (em até cinco anos), o elevado gasto energético despendido no processo de aprendizado e questões que envolvem, por exemplo, desinformação, direitos autorais e privacidade de dados são alguns dos problemas apontados. A necessidade de regulamentação legal destas tecnologias já está ocorrendo, sendo a União Europeia líder na implementação de políticas de desenvolvimento e controle. No Brasil, que é fortemente influenciado pela legislação europeia, ainda estamos atrasados neste trabalho.

É importante ressaltar, no caso dos Modelos de Linguagens e das Inteligências Artificiais Generativas, que o resultado obtido é

apenas o reflexo de nossa sociedade. Elas não criam preconceitos e vieses, elas apenas os expõem; elas não amplificam espontaneamente preconceitos e vieses, mas as pessoas podem usá-las para isso. Criar desinformação se torna uma tarefa mais fácil, mas a própria Inteligência Artificial pode nos ajudar a evitar esse problema.

O momento que vivemos é apenas a sequência natural do desenvolvimento tecnológico que a humanidade tem experimentado. Considere como ponto de partida a invenção da máquina de imprensa de Gutenberg, que ampliou e facilitou a circulação de conhecimentos e ideias (boas e más). Se tivéssemos seguido o pensamento do monge beneditino Filippo de Strata, que no século XV nos alertou sobre como textos impressos poderiam corromper “corações suscetíveis” (PETTEGREE, 2010), talvez hoje não estivéssemos enfrentando esses problemas nem tivéssemos o progresso que tivemos.

Ou a invenção do rádio, que permitiu a criação e a divulgação de (des)informação, incluindo pessoas que não tinham acesso à (des)informação escrita, ou pela impossibilidade de acesso, ou que não eram capazes de ler. O impacto desta tecnologia pode ser sentido pela população norte-americana quando da transmissão em 1938 de uma adaptação radiofônica do romance *A guerra dos mundos* de H. G. Wells, e de maneira mais dramática no papel que ela desempenhou na propagação de ideias que nos levaram à Segunda Guerra Mundial. De certa forma, o que passamos hoje são situações semelhantes às já vividas, só que hoje em uma escala muito maior.

A visão de máquinas inteligentes que ameaçam a humanidade é um tópico recorrente na ficção científica. Parece pouco provável hoje que a Inteligência Artificial possa ser capaz de acabar com a humanidade, tarefa que parece estar sendo mais bem desempenhada pela inteligência natural. O que se deve ter em mente é que a Inteligência Artificial é apenas uma ferramenta, termo que usei com frequência neste texto, e o que importa no final é como esta ferramenta será usada por nós. Usada de forma inteligente, pode nos levar a um mundo muito melhor. Tudo vai depender da inteligência natural.

Palavras-Chave: Inteligência Artificial; Processamento de Linguagem Natural; Aprendizado de Máquina; Modelos de Linguagem.

Referências

McCULLOCH, W.; PITTS, W. A Logical Calculus of Ideas Immanent in Nervous Activity. **Bulletin of Mathematical Biophysics**, 5 (4), 1943, p. 115-133.

PETTEGREE, Andrew. **The Book in the Renaissance**. Yale University Press, 2010.

TURING, Alan. Computing Machinery and Intelligence. **Mind**, v. 49, n. 236 (1950): 433-460. Disponível em: <https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf>.

WEIZENBAUM, Joseph. ELIZA-A Computer Program for the Study of Natural Language Communication Between Man and Machine. **Communications of the ACM**, v. 9, n. 1 (jan. 1966): 36-35. Disponível em: http://www.universelle-automation.de/1966_Boston.pdf.

WINOGRAD, Terry. Understanding natural language. **Cognitive Psychology**, v. 3, n. 1, 1972, p. 1-191. Disponível em: <https://www.sciencedirect.com/science/article/pii/0010028572900023?via%3Dihub>.

Notas

[1] Disponível em: <https://commoncrawl.org/>.

GOVERNO DO ESTADO DE SÃO PAULO

GOVERNADOR
Tarcísio Gomes de Freitas

VICE-GOVERNADOR
Felício Ramuth

SECRETÁRIA DE ESTADO DA CULTURA,
ECONOMIA E INDÚSTRIA CRIATIVAS
Marília Marton

SECRETÁRIO EXECUTIVO
Marcelo Assis

CHEFE DE GABINETE
Daniel Scheiblich Rodrigues

COORDENADORA DA UNIDADE DE
PRESERVAÇÃO DO PATRIMÔNIO MUSEOLÓGICO
Karina Santiago

DIRETORA DO GRUPO TÉCNICO DE
COORDENAÇÃO DO SISTEMA ESTADUAL
DE MUSEUS
Sofia Gonçalves

DIRETORA DO GRUPO DE PRESERVAÇÃO DO
PATRIMÔNIO MUSEOLÓGICO
Mirian Midori Peres Yagui

DIRETORA DO NÚCLEO DE APOIO
ADMINISTRATIVO
Regiane Lima Justino

EQUIPE TÉCNICA DA UNIDADE DE PRESERVAÇÃO
DO PATRIMÔNIO MUSEOLÓGICO
Angelita Soraia Fantagussi
Dayane Rosalina Ribeiro
Eleonora Maria Fincato Fleury
Kelly Rizzo Toledo Cunegundes
Luana Gonçalves Viera da Silva
Marcia Pisaneschi Sorrentino
Marcos Antônio Nogueira da Silva
Roberta Martins Silva
Tayna da Silva Rios
Thiago Brandão Xavier

MUSEU DA LÍNGUA PORTUGUESA

DIRETORA EXECUTIVA
Renata Vieira da Motta

DIRETORA ADMINISTRATIVA
E FINANCEIRA
Vitória Boldrin

DIRETORA TÉCNICA
Roberta Saraiva

CENTRO DE REFERÊNCIA
Camila Chagas Aderaldo
Cecilia Farias
Janaina Lopes
Leonardo Arouca
Luiza Victória Brito Magalhães

**Dossiê II Seminário
Viagens da Língua: Língua
e tecnologias**

COORDENAÇÃO EDITORIAL
Camila Aderaldo e Cecilia Farias

ORGANIZAÇÃO
Cecilia Farias

REVISÃO
Daniela Uemura

CAPA, PROJETO GRÁFICO
E DIAGRAMAÇÃO
Patrícia Yamamoto

IDENTIDADE VISUAL
Beto Cavalcante

**Centro de Referência do
Museu da Língua Portuguesa**

+55 11 4470 1507
centrodereferencia@mlp.org.br

Praça da Luz
01120-010 São Paulo-SP
museudalinguaportuguesa.org.br

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Seminário Viagens da Língua (2. : 6-8 dez. 2023 :
São Paulo, SP)

Dossiê II Seminário Viagens da Língua : língua e tecnologias /
organização Cecília Farias ; coordenação Camila Chagas Aderaldo. --
São Paulo : Museu da Língua Portuguesa, 2024.

Vários autores.

ISBN 978-65-999470-3-2

1. Inteligência artificial 2. Língua portuguesa - Estilo
3. Linguística - Congressos 4. Preservação 5. Seminários
I. Farias, Cecília. II. Aderaldo, Camila Chagas. III. Título.

24-211883

CDD-418

Índices para catálogo sistemático:

1. Linguística : Congressos 418

Catalogação na fonte: Eliane de Freitas Leite - Bibliotecária - CRB 8/8415

Este livro foi composto por Söhne e
impresso em papel offwhite 90 g/m²
por Cocktail em junho de 2024.

Museu da Língua Portuguesa

II Seminário Viagens da Língua: Língua e tecnologias



Patrocínio



JOHN DEERE



Apoio



Gestão



Patrocínio Máster Reconstrução

Concepção e Implantação

Realização



MINISTÉRIO DA CULTURA



