

# Língua portuguesa e tecnologia: o futuro é agora



Ana Paula Laborinho e António Branco

Os avanços da Inteligência Artificial têm sido impressionantes, sobretudo na sua aplicação à Tecnologia da Língua. Este progresso é baseado na aprendizagem automática com os chamados Grandes Modelos de Linguagem, como o GPT-3 ou o ChatGPT, de que tanto se tem falado recentemente.

Estes modelos são redes neuronais artificiais que aprendem ao ajustarem os pesos das ligações entre os neurónios para melhorarem o nível de acerto durante o treino. As tarefas de treino consistem em, a partir dos respetivos contextos de ocorrência, prever palavras omitidas.

Estas redes são gigantescas – o GPT-3, por exemplo, tem 175 mil milhões de ligações entre neurónios. Captam regularidades linguísticas quando treinadas em processos computacionais maciços, sobre volumes colossais de dados linguísticos, texto ou áudio. No caso do GPT-3, foram usadas 500 mil milhões de palavras no treino, cujo custo de computação é estimado em mais de 4 milhões de dólares.

Uma vez treinados, estes modelos podem ser usados em outras tarefas linguísticas com

um nível inédito de qualidade, como por exemplo tradução, conversação, transcrição de fala e legendagem, geração de texto e fala, análise do conteúdo e extração de informação, etc. Quando integrados em sistemas mais vastos, estão a transformar os diagnósticos e cuidados de saúde, os serviços financeiros e jurídicos, os jogos e o entretenimento, o ensino, a criatividade e a cultura, etc.

Devido ao tamanho dos modelos, estas tarefas de processamento estão disponíveis remotamente como serviços *online*, como é o caso dos motores de busca, e não como os corretores ortográficos de instalação local nos nossos dispositivos. Devido à dimensão dos recursos para a aprendizagem, no imediato esses serviços são disponibilizados pelo oligopólio das *bigtechs*, que se contam pelos dedos de uma mão, com capacidade de acesso aos colossais volumes de computação e dados necessários para o treino.

Em consequência, na era digital a utilização da linguagem – com outros seres humanos, organizações, serviços ou dispositivos artificiais – não voltará a ser feita sem esta intermediação tecnológica pervasiva e profunda, que processa os atos de comunicação e acede ao seu significado.

Temos suficiente experiência com os motores de busca de informação, por exemplo, e com os seus pressupostos e impactos, para intuir as consequências desta intermediação tecnológica na utilização quotidiana da própria linguagem. Intermediação tecnológica, em geral, gera um rasto digital de dados pessoais fora do nosso controlo. Intermediação tecnológica incessante da linguagem e comunicação

humanas, em particular, afunilada num pequeno oligopólio mundial, gera riscos alarmantes para as soberanias individuais e coletivas.

Impactos indesejáveis de tecnologias emergentes mitigam-se com mais e melhor tecnologia, não com menos. A dispersão do fornecimento destes serviços é crucial para debelar a ameaça que a sua concentração constitui. A resposta encontra-se assim na promoção de um ecossistema de inovação que, em alternativa, permita atempadamente banalizar o acesso aos recursos necessários para a apropriação e exploração da Tecnologia da Linguagem pelo maior número possível de indivíduos e organizações, privadas e públicas, pequenas e grandes, nacionais e internacionais.

Face aos volumes de dados e custos de computação que é necessário fazer convergir



**Se não acrescentarmos às políticas de língua uma aposta clara na sua preparação tecnológica, o português perderá importância e tenderá no limite a ser substituído por outras línguas**

e disponibilizar para propiciar tal ecossistema, e perante o mais relevante interesse público em causa, esta é uma incumbência, nova e urgente, para os Estados democráticos, isoladamente e em cooperação.

A língua portuguesa, com 250 milhões de falantes em quatro continentes, é uma das grandes línguas internacionais de projeção global. Os indicadores apontam para o seu crescimento até final do século com a maioria dos falantes no continente africano. Contudo, se não acrescentarmos às políticas de língua clássicas uma aposta clara na sua preparação tecnológica, perderá importância e tenderá no limite a ser substituída por outras línguas. Por essa razão, importa congregarmos esforços para que haja um Plano de Preparação Tecnológica da Língua Portuguesa, o que terá efeitos em diferentes domínios, incluindo a economia.

Em Espanha, há mais de uma década que se trabalha no Plano de Impulso das Tecnologias da Linguagem: entre 2015 e 2020, recebeu uma dotação de 89 milhões de euros, reforçado com uma dotação adicional de 334 milhões de euros até 2025, agora no âmbito da iniciativa Nova Economia da Língua.

Várias iniciativas europeias podem ser aproveitadas – se acreditarmos que a língua portuguesa é um ativo que não podemos desperdiçar.

**Directora em Portugal da Organização de Estados Ibero-Americanos para a Educação, a Ciência e a Cultura; director-geral da Portulan Clarin Infra-Estrutura de Investigação para a Ciência e Tecnologia da Linguagem**