

António Branco

Universidade de Lisboa, Portugal

Programas em Rede para o Desenvolvimento Tecnológico das Línguas

Com o advento da era digital, as línguas enfrentam um novo choque tecnológico, que volta a impor novas exigências à sua utilização, promoção e sobrevivência. À semelhança de choques tecnológicos anteriores, os falantes serão atraídos pelas línguas com melhor promessa de oportunidades, por aquelas línguas com as quais é possível interagir com novos serviços, dispositivos, jogos, robôs, realidades aumentadas, universos simulados, etc.

Preparação tecnológica dos falantes

Para tanto, os falantes precisam de estar preparados tecnologicamente. Precisam de ter acesso à *internet*, a dispositivos computacionais, a aplicações informáticas e a demais dispositivos, tanto aos que são importantes agora, como àqueles que o virão a ser no futuro. E esse é um desafio importante, para o qual há um amplo reconhecimento social e político, e um arsenal de tipos de respostas em termos de políticas públicas.

Essa é uma condição necessária, mas de forma alguma suficiente, para a continuidade das línguas na era digital. É crucial que, não só os falantes, mas as línguas, elas próprias, estejam preparadas tecnologicamente. Porém, esse é um desafio da maior importância que está ainda por ser

amplamente reconhecido social e politicamente, e por maioria de razão por ser amplamente enfrentado através de políticas públicas.

Haver páginas *web* e vídeos numa dada língua, haver aplicações cujos menus estão nessa língua, haver redes sociais em que falantes dessa língua a usam quando comunicam entre si, haver dicionários dessa língua em linha, etc., não é, nem sequer aproximadamente, ter essa língua preparada tecnologicamente.

Preparação tecnológica das línguas

Para uma língua estar preparada tecnologicamente para a era digital, é necessário que os seus falantes a possam usar — tal e qual como fazem quando a usam entre si — para comunicar com robôs, com agentes e serviços artificiais, com acervos de informação e com dispositivos computacionais, assim como para comunicar com os falantes de outras línguas com a intermediação desses serviços. É necessário que, por via de processamento semântico profundo, os dispositivos computacionais consigam obter uma representação do significado veiculado por essa língua, de forma a suportarem toda uma gama de novas possibilidades e de serviços avançados — alguns já concretizáveis, vários apenas imaginados e inúmeros ainda por imaginar.

Para uma língua estar preparada tecnologicamente para a era digital, é preciso sobre ela levar a efeito, de forma sistemática e diferenciada, investigação científica que aprofunde conhecimentos e capacidades, em que se baseiam todos estes avanços tecnológicos, sobre os quais se desenvolvam propostas inéditas de inovação e de construção de valor.

Mas também aqui, essa é uma condição necessária, mas de forma alguma suficiente. Do ponto de vista do interesse comum, é crucial que a preparação tecnológica das línguas para a era digital assegure a cidadania dos seus falantes e da sua cultura na sociedade da informação. Esse é, porém, um

desafio da maior importância que carece de ser devidamente equacionado, e que por isso está por ser amplamente reconhecido social e politicamente, e por maioria de razão por ser enfrentado.

Preparação tecnológica para a cidadania digital

Para que a preparação tecnológica sirva a cidadania digital dos falantes de uma língua na sociedade de informação, é preciso levar a efeito investigação científica sobre essa língua com uma intensidade que se aproxime da intensidade como, até hoje, só foi feita para a língua inglesa. De outra forma, o crescente efeito de atração para uma esfera linguística digital global e unipolar, em acelerada convergência monolíngue para o inglês, nunca chegará a ser suficientemente contrariado.

Para além de uma intensidade acrescida, essa preparação precisa de ser feita com recursos para avanços de ponta que se aproximem dos que, até hoje, têm estado ao alcance sobretudo das *bigtech*, sediadas nos EUA. É preciso que os recursos científicos e tecnológicos de base e os conjuntos de dados relevantes se encontrem disponíveis para as restantes empresas e iniciativas de inovação, para o sector público, para a esfera criativa e cultural, para as comunidades de investigação, e para todos os interessados em geral.

Tal é necessário para que se possam exercer as oportunidades e as capacidades de explorar tais dados e recursos endógena e autonomamente, e assim evitar e derrubar barreiras à concorrência e à inovação.

Tal é necessário para que o meio básico de comunicação e bem comum maior, que é a linguagem, não acabe apropriado por monopólios de intermediação, como aconteceu com a busca de informação na web, e assim evitar a corrosão da soberania.

Tal é necessário para que se possa contribuir para se quebrar com a concentração dos dados privados dos utilizadores,

e assim evitar a degradação da liberdade e a diminuição da cidadania na sociedade da informação.

Painel sobre Línguas, Tecnologia e Inovação

No exíguo espaço da presente nota, reservado para uma breve súpula do nosso painel, não caberá o aprofundamento destas temáticas nem o desenvolvimento sobre os respetivos desafios e respostas. O objetivo aqui é tentar “unir os pontos”, é procurar dar encadeamento e ajudar à compreensão dos temas essenciais que as riquíssimas contribuições dos palestrantes trouxeram para este painel de discussão sobre “Línguas, Tecnologia e Inovação”, que tive a honra e o privilégio de coordenar, e que fez parte da conferência CILPE ocorrida em Brasília, em fevereiro de 2022.

Ainda menos caberá aqui substituir ou reduzir as abordagens e propostas avançadas pelos palestrantes, cujas apresentações pelas suas próprias penas se encontram nas páginas vizinhas deste volume. O objetivo é, por isso, de servir de convite e estímulo à leitura e compreensão dessas contribuições e do produtivo painel de discussão que ajudaram a construir:

- **Mario Tascón**, jornalista e especialista nos novos média digitais, contribuiu para o painel da perspectiva do jornalismo, com uma problematização crítica e abrangente sobre as dinâmicas da sociedade de informação e do seu impacto nas línguas e respetivas culturas.
- Ativista junto do povo Enlhet, **Hannes Kalish** trouxe o testemunho das tensões e desafios acrescidos que a era digital representa para línguas e povos indígenas, com populações minoritárias e tecnologicamente menos favorecidas.
- **Olga Juan**, responsável pelas tecnologias aplicadas no Instituto Cervantes, assinalou a importância da prepara-

ção tecnológica dos falantes, assim como o papel crucial da tecnologia na promoção das línguas e da sua aprendizagem por falantes não nativos.

- Nesse propósito, foi secundada pela comunicação de **Mauro Luiz Rabelo**, do Ministério da Educação e Cultura do Brasil, em que foi realçada a importância de políticas públicas neste capítulo, incluindo junto dos próprios falantes nativos das línguas.
- **Renata Vieira** ajudou a compreender com maior profundidade as virtualidades da tecnologia da linguagem e a necessidade da preparação tecnológica das línguas para a era digital. Ela é investigadora científica e perita no processamento computacional das línguas.
- **Christopher Shulby** é Diretor de Aprendizagem Automática na *startup* Defined.AI, cujo modelo de negócio se baseia na uberização da anotação linguística de dados, cruciais para o treino de sistemas de processamento de fala e texto. Deu testemunho das soluções de negócio que são possíveis com a exploração inovadora da tecnologia da linguagem.
- Da forma como neste campo a inovação está a transformar o modo de utilizar as línguas e de comunicar com falantes de outras línguas, deu conta **Vasco Pedro**, CEO da *startup* Unbabel, cujo modelo de negócio assenta na uberização do trabalho de tradução, crucial para a adaptação ágil dos negócios aos diferentes mercados linguísticos e sua internacionalização.
- Consultor na área das políticas públicas para o setor das telecomunicações e das tecnologias da informação,

Jorge Fernando Negrete equacionou as relações entre iniciativa privada e interesse público, ajudando a colocar em destaque como a preparação tecnológica das línguas dever ser concretizada em termos que contribuam para a salvaguarda das culturas e do interesse comum dos seus falantes.

- **German Rigau i Claramunt** é investigador científico em processamento computacional da linguagem. Deu conta dos desafios para as políticas públicas colocados pela tecnologia da linguagem, assim como do “Plano de Estímulo para as Tecnologias da Linguagem” (2015-2020) do Governo de Espanha, de cujo desenho foi um dos principais arquitetos.
- Pioneiro e Professor Emérito em Inteligência Artificial, **Helder Coelho** é um profundo conhecedor das comunidades de investigadores científicos ibero-americanos e dos seus laços de cooperação, que ao longo das décadas passadas ajudou a estabelecer e para os quais contribuiu. Deixou-nos a sua expectativa informada sobre como, com a mobilização da OEI-Organização de Estados Ibero-americanos e da sua conferência CILPE, tais laços podem continuar a ser aprofundados.

Como expressei acima na parte inicial da presente nota, na minha perspetiva, o alinhamento destas contribuições e do debate que eles proporcionaram durante a conferência aponta no sentido de que, do ponto de vista das políticas públicas, há que trazer para o reconhecimento social e político, como desafio existencial maior na defesa e promoção das línguas, a urgente necessidade da sua preparação tecnológica para a era digital em termos que contribuam para a salvaguarda do interesse comum e da soberania, e também das culturas e dos direitos e garantias dos seus falantes.

Programas de desenvolvimento tecnológico

Na anterior edição da CILPE, de 2019, recebi a distinção de poder participar, dessa feita, como um dos palestrantes do painel sobre “Línguas e Sociedade Digital”. Como princípio de resposta a este desafio, assinei a necessidade, oportunidade e urgência de um Plano de Desenvolvimento Tecnológico para a Língua Portuguesa na Era Digital.¹⁴ Desde então, e atendendo ao aprofundamento desse debate na presente edição da CILPE de 2022, essa proposta apenas ganhou uma ainda mais diferenciada justificação, e mais ampla aplicação ao leque das línguas e suas variantes no âmbito da OEI-Organização de Estados Ibero-Americanos.

Numa perspetiva abrangente, de cima para baixo, tal plano deverá ser suficientemente ambicioso para incluir, entre outros, capítulos relativos à cooperação internacional estratégica, à investigação e formação interdisciplinar avançada, à capacitação científica e tecnológica, à modernização administrativa, e à inovação especializada. Do prisma oposto, numa perspetiva operativa, de baixo para cima e dando prioridade às necessidades mais urgentes primeiro, tal plano deve acautelar, entre outros aspetos, o desenvolvimento de coleções de dados linguísticos e de modelos de linguagem que funcionam como blocos básicos e abertos para o desenvolvimento de soluções e negócios inovadores. Deve contemplar a reunião, curadoria e distribuição pública e aberta de dados, singelos e anotados linguisticamente, entre os quais se incluem coleções de texto e de áudio com fala suficientemente grandes, provenientes em grande medida dos acervos do setor público, devidamente anonimizados. Deve contemplar também o desenvolvimento e distribuição aberta de grandes modelos de linguagem e de ferramentas de processamento da linguagem,

14 António Branco, 2021, “Plano de Desenvolvimento Tecnológico da Língua Portuguesa: necessário, oportuno e urgente”, In Ibero-América: Uma Comunidade, Duas Línguas Pluricêntricas, OEI-Organização de Estados Ibero-Americanos, Imprensa Nacional, pp. 173–177, ISBN 978-972-27-2979-6. Disponível também em <http://www.di.fc.ul.pt/~ahb/pubs/2021BrancoB.pdf>

para cujo treino e teste essas coleções de dados são necessárias, e para cujo desenvolvimento, os elevados custos dos recursos a mobilizar se encontram muito longe de estar ao alcance das dotações usuais das equipas de investigação e do orçamento de pequenas e médias empresas.

Aproveitei na altura também para indicar exemplos de iniciativas alinhadas com esse propósito, com um caso saliente vindo do país de Cervantes. O Governo de Espanha encontrava-se a executar um “Plano de Estímulo para as Tecnologias da Linguagem”, para o período de 2015 a 2020, com um montante atribuído de 89 milhões de euros de financiamento público.

Recentemente, a 1 de março de 2022, em data já posterior à do nosso painel na CILPE, foi aprovado pelo Governo de Espanha um “Projeto Estratégico para a Recuperação e Transformação Económica da Nova Economia da Língua”, com uma dotação de 1.100 milhões de euros de financiamento público. Dois dos seus seis eixos visam apoiar a preparação tecnológica das línguas para a cidadania digital, dando continuidade ao anterior “Plano de Estímulo” que havia decorrido até 2020.¹⁵

Programas em rede

Desejavelmente, em próximas edições da CILPE, o objetivo já não deveria ser trazer para o reconhecimento público, como desafio existencial maior na promoção das línguas e das suas variantes, a necessidade da sua preparação tecnológica para a era digital. Idealmente, o objetivo deverá ser já o de discutir a melhor forma de compor e articular em rede os programas individuais de desenvolvimento da tecnologia da linguagem dos países da OEI-Organização de Estados Ibero-Americanos, para assim se procurar otimizar custos e resultados com a sua potenciação recíproca, partilhada e cooperativa.

15 <https://planderecuperacion.gob.es/como-acceder-a-los-fondos/pertes/perte-nueva-economia-de-la-lengua>