

A tradução automática na didática da tradução

Desafios e oportunidades

Machine translation in translation teaching

Challenges and opportunities

João Brogueira

(Centro de Estudos de Comunicação e Cultura - Universidade Católica
Portuguesa)

RESUMO

Nos últimos anos, avanços tecnológicos importantes a nível da Tradução Automática Neuronal aproximaram estes sistemas da promessa de traduzibilidade universal, agora amplificados por versões sucessivas de ferramentas generativas, como o ChatGPT. Estas tecnologias terão necessariamente efeitos ao nível da formação e da didática da tradução e também das línguas. No presente trabalho, toma-se como base dois exercícios realizados em aula para analisar os desafios e oportunidades criados pela TA e o seu impacto na didática da tradução, bem como a importância da literacia em TA como elemento estruturador da formação de tradutores.

Palavras-Chave: tradução automática, didática da tradução, inteligência artificial

ABSTRACT

Over the last few years, technological advances in Neural Machine Translation have brought these systems closer to the promise of universal translatability, now amplified by successive releases of generative tools, like ChatGPT. These technologies will necessarily have an impact in terms of translation training and teaching, and language teaching. In this paper, two in-class exercises are used to analyze the challenges and opportunities created by MT and its impact in translation teaching, as well as the importance of MT literacy as a structural element in translator training.

Keywords: machine translation, translation teaching, artificial intelligence

Introdução

A 30 de novembro de 2022, a empresa OpenAI disponibilizou ao público o ChatGPT, um *chatbot* baseado em grandes modelos de linguagem originalmente pensado para conversar com os utilizadores sobre os mais diversos temas, podendo os utilizadores definir diversos parâmetros como o tema, a língua, a profundidade ou a extensão da conversa. Cedo, porém, se verificou que esta nova tecnologia também era competente em tradução automática (TA) e poderia representar um importante contributo para uma maior acessibilidade da tradução. Em contrapartida, não têm faltado vozes a clamar a redundância da profissão de tradutor.

Na verdade, a inovação tecnológica tem acompanhado a prática profissional da tradução. Desde meados do século XX que se tenta usar o computador para traduzir, quer autonomamente, quer como ferramenta auxiliar do tradutor. Com a adoção da tradução

automática neuronal (NMT) em meados da década passada, é inegável um salto qualitativo na TA. Em face desta evolução, muitos são os docentes de Tradução que se veem confrontados, por um lado, com a necessidade de reformular métodos de ensino e, por outro, com a dificuldade de motivar os alunos para aprender um conjunto de técnicas que os computadores desempenham, aparentemente, cada vez melhor. O presente trabalho tem como objetivo dar um contributo prático para analisar os desafios e oportunidades criados pela TA e o seu impacto na didática da tradução, em particular com o português europeu como língua de chegada.

Estruturado em três partes, o artigo começa por apresentar brevemente o estado da arte e uma súmula da evolução histórica da TA. Em seguida, expõem-se dois exercícios realizados em sala de aula, com tradução automática, visando explorar o potencial e as limitações das ferramentas que usam esta tecnologia no seu estado atual. Por último, é abordado o tema da literacia em tradução automática como forma de capacitação para a competência tecnológica, em complemento das competências clássicas de tradução.

Estado da arte

Nos últimos anos, avanços tecnológicos importantes a nível da NMT aproximaram estes sistemas da promessa de traduzibilidade universal, agora amplificados por versões sucessivas de ferramentas generativas.

Não é exagerado falar-se de uma nova era para a tradução e da redefinição do papel do tradutor ou intérprete (humano). Na verdade, um qualquer telemóvel mediano com acesso de banda larga à internet já se assemelha, em funcionalidade, a equipamentos como o Starfleet Universal Translator, um dispositivo de tradução universal, que na série de ficção científica *Star Trek (O caminho das estrelas)* surge no século XXII.

Em qualquer ponto do planeta podemos usar um telemóvel para traduzir texto, incluindo a captação de imagens e conversão em texto para posterior tradução. Em alguns sistemas mais sofisticados, é inclusivamente possível fazer a síntese de voz de entrada e de saída e, desse modo, ter acesso a uma forma rudimentar de interpretação automática. Nos últimos anos, muitos têm sido os termos que têm classificado o último avanço tecnológico. Desde *ciber* a *algoritmo*, passando pela *nuvem*, o mais recente é a *inteligência artificial*, que na verdade é conceito vetusto e com uma história partilhada com a tradução automática.

Retorno ao passado: uma ameaça octogenária

A ideia de uma língua comum a todas as pessoas é tão antiga como a Bíblia, tendo sido revisitada nos séculos XVII e XVIII por filósofos como Leibniz, Descartes e Wilkins (Walker, 1972).

Contudo, é no século XX que surgem as primeiras máquinas idealizadas para traduzir textos, havendo registo de patentes na França e na União Soviética em 1933 (Somers, 2003). O final da Segunda Guerra Mundial traz um novo impulso ao objetivo de desenvolver dispositivos de tradução automática (TA), respondendo a necessidades de âmbito militar. Warren Weaver, vice-presidente da Rockefeller Foundation, publicou em 1949 um memorando no qual estabelecia uma analogia entre as técnicas de criptografia e a tradução (Austermühl, 2001). Dois anos depois, Yehoshua Bar-Hillet começou a fazer investigação em TA, a tempo inteiro, no Massachusetts Institute of Technology (MIT) e foram criados grupos de investigação noutros países, incluindo a URSS, a Grã-Bretanha e a França. A Guerra Fria e a necessidade de intercetar rapidamente as mensagens do inimigo deram o primeiro grande impulso à investigação neste domínio.

A primeira avaliação dos resultados do investimento de dinheiros públicos norte-americanos em investigação no domínio da TA foi o relatório ALPAC, publicado em 1966, que tecia considerações bastante negativas sobre o potencial da TA. Entre elas, concluía que era mais lenta, menos precisa e duas vezes mais cara do que a tradução humana, tinha procura reduzida e havia mais potencial na tradução auxiliada por computadores (Somers, 2003). Em resultado desta avaliação desfavorável, verificou-se uma redução das verbas destinadas à TA e um quase abandono da investigação neste domínio.

Nas décadas de 1970 e 1980, novos catalisadores reavivaram a investigação em TA fora dos EUA e da URSS, em particular fatores de ordem política como a política de bilinguismo do Canadá e as políticas de multilinguismo das Comunidades Europeias, cujas necessidades de tradução cresciam a cada alargamento, e fatores de ordem económica, como o crescimento das exportações japonesas e a necessidade de criar documentação adaptada aos mercados de destino. Neste período, os sistemas de TA incorporaram conceitos da linguística estruturalista e das ciências da computação, tornando-se largamente baseados em regras (RBMT), que combinavam regras e dicionários para analisar o texto de partida e redigir o texto de chegada.

Perante a dificuldade em utilizar estes sistemas de uma forma generalizada em textos abertos, os investigadores centraram as atenções em aplicações específicas e usaram «sublínguas». Com recurso a linguagem controlada, fazendo um uso restritivo de estruturas e vocabulário e aplicando regras rigorosas de criação dos textos originais, os computadores são

mais eficientes no processamento de linguagem natural e, conseqüentemente, na tradução. O primeiro grande sucesso desta abordagem surge com o Sistema Méteo (Universidade de Montreal), iniciado na década de 1970, que tinha como objetivo a tradução de boletins meteorológicos de inglês para francês (Bowker, 2023).

Nas décadas que se seguiram, a TA continuou a pautar-se por uma evolução relativamente lenta, em especial fora de ecossistemas altamente especializados e concebidos de raiz para a tradução automática (ou seja, textos redigidos com estruturas sintáticas e léxico restritos, e expurgados de elementos culturais complexos). Foi já no século XXI que se deu um importante salto tecnológico e qualitativo com a NMT.

Aproveitando a elevada disponibilidade de textos em várias línguas na internet e fortes avanços na capacidade de computação, estes sistemas assentam em aprendizagem automática e grandes modelos de linguagem, que permitem adquirir competências específicas e/ou gerar texto com pouca ou nenhuma intervenção humana. Para a tarefa específica de tradução, «[a] ferramenta contém uma rede neuronal artificial que consulta este *corpus* de treino e, com base nos padrões identificados nos dados de treino, aprende a traduzir novos textos» (Bowker, 2023: 96).

Em 2016, com a implementação da tecnologia NMT no Google Tradutor, esta capacidade passou a estar à disposição de qualquer pessoa com acesso à internet e é hoje bem claro que a tradução é uma atividade maioritariamente realizada por não-tradutores. Fazendo as contas ao volume de palavras que o Google Tradutor traduz por dia e comparando com o volume de palavras que se estima ser traduzido diariamente por tradutores humanos, Pym e Torres-Simón concluíram que a tradução humana era, já em 2018, uma mera gota no oceano:

«Google Translate is said to process 143 billion words a day (Wood, 2018). (...) human mediators are producing just under one billion words a day – a mere 0.68% of what Google Translate is reported as processing.» (Pym e Torres-Simón, 2021: 3)

Não obstante todos os avanços tecnológicos e qualitativos, e o potencial da TA para aumentar a produtividade dos tradutores e melhorar a acessibilidade da tradução em geral e em serviços críticos de assistência a populações migrantes deslocadas ou em centros urbanos multilingues, são bem conhecidos alguns custos desta tecnologia. No plano social, teme-se o efeito sobre os profissionais da tradução, em particular a quebra de rendimentos. No plano ambiental, importa salientar que esta tecnologia requer significativos recursos informáticos que consomem muita energia. Pela sua complexidade técnica, estes sistemas tendem a ser opacos e imprevisíveis, o que muitas vezes resulta em textos incoerentes. Por último, no plano linguístico, há a salientar a ocorrência de «alucinações» (uma categoria de erros de TA que

ocorre quando o sistema produz resultados grosseiramente incorretos ou mesmo inexistentes) e, no caso da língua portuguesa, uma grande dificuldade na separação das variantes linguísticas de português de Portugal e português do Brasil. De facto, apesar de a língua portuguesa, no seu conjunto, ser considerada uma língua de elevados recursos, com um grande número de falantes e investigação avançada em processamento de linguagem natural, os efeitos de escala de treinar os sistemas com mais conteúdo em língua portuguesa sobrepõem-se a uma correta separação das variantes linguísticas. O aspeto económico, nesta fase de desenvolvimento, também não será despiciendo.

Perante uma tão profunda transformação na atividade da tradução, é natural que no seu ensino se gerem crises e resistências. Assim aconteceu noutras áreas e ainda hoje o uso da máquina de calcular nas aulas de Matemática ou Física gera acesos debates, embora ninguém coloque em causa que a calculadora é absolutamente fundamental para a realização célere de cálculos avançados.

Num estudo em curso na Faculdade de Ciências Humanas da Universidade Católica Portuguesa sobre a utilização de tradução automática, ferramentas generativas (ChatGPT, etc.) e outras ferramentas linguísticas no âmbito das unidades curriculares de línguas estrangeiras, verificou-se que são universalmente utilizadas pelos alunos (100%). Os alunos comunicaram, nesse mesmo estudo, vincada oposição dos docentes à utilização destas ferramentas.

Concluiu-se que os alunos depositam níveis de confiança médios a elevados nos resultados da TA, em especial em línguas que conhecem mal. Este é um dado preocupante, já que parecem confiar mais quando têm menos capacidade de verificar os resultados. Para os alunos, a qualidade dos textos produzidos por estes sistemas é considerada média. As principais vantagens apontadas são a rapidez, a simplificação das tarefas e a correção de erros.

Aplicações didáticas

Em face deste panorama, em particular de um aparente excesso de confiança nestas ferramentas, desenvolveram-se no âmbito da unidade curricular de Tradução Assistida por Computador (TAC) duas atividades visando explorar desafios e oportunidades criados pela TA.

No primeiro exercício, as alunas tinham de traduzir, a pares, um texto de um manual de utilização de uma impressora com a ferramenta de tradução assistida Phrase (com a opção de tradução automática ativada), outra na ferramenta web Google Tradutor e tinham ainda de criar a sua própria versão.

ID#ICU#	#	Source:(en)#	Target:(pt-pt)#	Comment# ·(read-only)#
		Getting-Started#	Phrase# Primeiros-passos#	Google-translate# # Começando#
	1#			Nossa#Para-começar# #
	2#	Finding-a-location-for-your-laser-printer#	Encontrar-um-local-para-a-sua-impressora-laser#	Encontrar-um-local-para-sua-impressora-a-laser# Encontre-um-local-para-a-sua-impressora-a-laser#
	3#	Place-the-laser-printer-on-a-flat,·clean-and-dust-free-surface,·in-a-dry-location,·and-out-of-direct-sunlight.#	Coloque-a-impressora-laser-·numa·superfície·plana,·limpa·e·sem·poeira,·num·local·seco,·e·sem·luz·solar·direta.#	Coloque-a-impressora-a-laser·em-uma·superfície-plana,·limpa-e-sem-poeira,·em-um·local-seco-e-longo-da-luz-solar-direta.# Coloque-a-impressora-a-laser-numa-superfície-plana,·limpa-e-sem-pó,·num-local-seco,·e-sem-contacto-direto-com-a-luz-solar.# #
	4#	Allow-at-least-22-cm-clearance-from-the-back-of-the-laser-printer-for-the-paper-to-travel.#	Deixe-pelo-menos-22-cm-de-distância-da-parte-traseira-da-impressora-laser-para-que-o-papel-viaje.#	Deixe-pelo-menos-22-cm-de-espaço-na-parte-de-trás-da-impressora-a-laser-para-que-o-papel-se-desloque.# Permita-pelo-menos-22-cm-de-espaço-na-parte-de-trás-da-impressora-a-laser-para-que-o-papel-entre..#

Figura 1. Matriz de tradução com as três versões.

De seguida, tinham de fazer uma análise contrastiva das três versões, comentando os aspetos positivos e negativos das três traduções. O exercício foi realizado numa aula com a duração de 80 minutos. Outras alunas usaram o DeepL na comparação.

Através das secções realçadas na Figura 1, é fácil ver o tipo de erros apontados (e também alguns casos de sobrecorreção). Além da matriz elaborada com os resultados em bruto obtidos nos diferentes sistemas e a versão final de cada grupo, as alunas tinham também de elaborar um comentário sobre os resultados observados. Reproduz-se em seguida o comentário de uma das equipas correspondente à matriz reproduzida na Figura 1 (sublinhado do autor):

No que concerne à tradução automática do programa Phrase, considero que é uma boa tradução mesmo que tenha alguns erros e palavras do português do Brasil. No geral, a tradução tem uma boa qualidade e é perceptível.

A tradução automática do Phrase não obtém nenhum problema de significado. Em relação à ortografia e sintaxe da tradução, encontrei algumas palavras em português do Brasil, fora isso as frases estão bem construídas e a mesma pode ser vista como uma boa tradução.

A tradução não é muito consistente, visto que uma vez traduz “botão de ligar e desligar” e logo na frase seguinte está traduzido como “botão on/off”. Logo é perceptível que a tradução não mantém coerência nem consistência.

Ao comparar as traduções automáticas com a minha própria tradução, notei que há algumas semelhanças com ambas as traduções automáticas mas a minha tradução é mais constante e usa termos mais próximos à versão original. (...)

Este tipo de análise capacita as alunas para uma abordagem mais crítica a estas ferramentas e às suas reais capacidades. Deste modo, ficam a conhecer os pontos fortes e os pontos fracos de sistemas ilusoriamente simples, o que lhes permite criar estratégias para abordar futuras traduções com mais segurança e mais capacidade para eliminar as falhas.

Além disso, ganham experiência num tipo de tarefa que irá ganhar cada vez mais relevância no setor da tradução: a pós-edição de tradução automática. Uma tarefa com grandes semelhanças com a tradicional revisão, mas com a preocupação extra de conferir características humanas a um texto traduzido automaticamente e adaptá-lo a eventuais necessidades especiais dos clientes.

Num outro exercício, realizado em casa, individualmente e com apresentação em aula, as alunas foram convidadas a fazer um trabalho em torno da tradução automática. Os temas foram bastante diversos, variando desde uma previsível comparação entre ferramentas (como o Google Tradutor ou o DeepL) à comparação do desempenho na tradução literária entre estas ferramentas e versões traduzidas publicadas. Um caso em particular foi mais longe e tentou explorar o potencial do ChatGPT para criar diferentes versões da mesma tradução em função do público-alvo a que se destina, usando comandos do sistema para modificar os resultados.

No exemplo apresentado *infra*, a aluna usou o ChatGPT (versão 3.5, gratuita) para criar versões portuguesas do texto publicitário de um videogame de grande sucesso. Usando apenas *prompts*, pediu ao ChatGPT para criar versões formais, informais e infantis – em português – do mesmo texto.

A Figura 2 mostra o início do texto original.

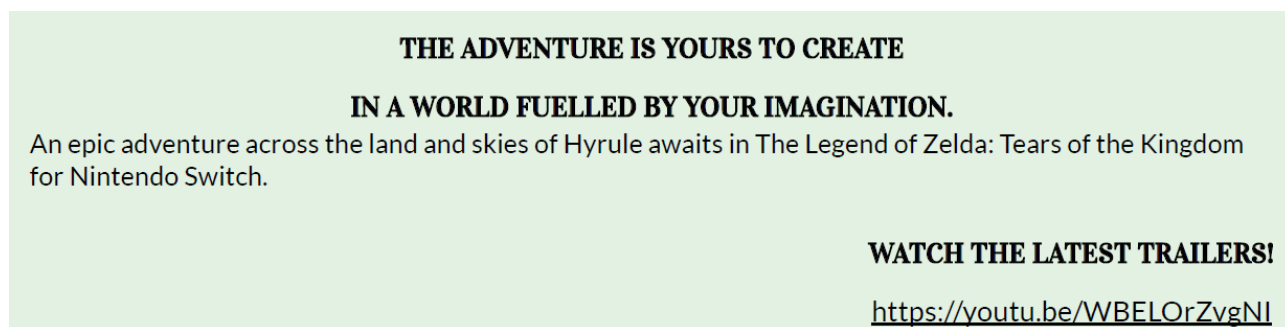


Figura 2. The Legend of Zelda. Excerto do texto original.

Para a primeira versão, a aluna pediu ao sistema para produzir o texto num registo formal, para um público adulto. O resultado, patente na Figura 3, corresponde ao esperado, pese a interferência do português do Brasil no tratamento da contração da preposição e artigo e também na escolha de terminologia para o dispositivo. Trata-se de um bom ponto de partida para a criação de uma versão adaptada ao mercado português.

DESENVOLVA A SUA PRÓPRIA AVENTURA

EM UM MUNDO ONDE A SUA IMAGINAÇÃO PODE VOAR LIVREMENTE.

Embarque em uma jornada épica que o levará por terras e vastos céus de Hyrule em The Legend of Zelda: Tears of the Kingdom para o console Nintendo Switch

Desfrute da exibição dos trailers emocionantes!

Figura 3. The Legend of Zelda. Tradução formal.

No pedido seguinte, a aluna definiu a *prompt* para obter uma versão informal do mesmo texto, dirigida a jovens. O resultado, patente na Figura 4, mostra uma variação de registo compatível com o solicitado. Note-se, em particular, a utilização de vocábulos como «Bora» e «mano». No entanto, nesta mesma versão, o ChatGPT finaliza o texto com uma frase ambígua e gramaticalmente incoerente.

Bora criar a tua própria aventura,

Num mundo onde tu podes soltar a imaginação e voar alto!

Vamos encarar uma jornada épica, mano!

Vamos percorrer as terras incríveis e os céus imensos de Hyrule no jogo The Legend of Zelda: Tears of the Kingdom para o Nintendo Switch. Vai ser demais, não vê a hora de embarcar nessa aventura junto comigo?

Figura 4. The Legend of Zelda. Tradução informal.

Na última versão pedida, dirigida a crianças, é uma vez mais visível a capacidade da ferramenta para modular o discurso, produzindo um texto com rima e um tom característico de histórias infantis, convidando o leitor a juntar-se à aventura. Na Figura 5, a distância face ao original é evidente, mas não deixa de ser uma possibilidade de adaptação do texto.

Vamos criar a tua aventura especial,
Num mundo onde a imaginação é a lei fundamental,
Vamos embarcar numa viagem mágica e emocionante,
A percorrer as terras e os céus de Hyrule, radiante!

No jogo The Legend of Zelda: Tears of the Kingdom,
Vamos voar alto na nossa Nintendo Switch, sem ter medo de nada,
A diversão será imensa, cheia de encanto e alegria,
Vamos juntos explorar e viver essa fantasia!

Figura 5. The Legend of Zelda. Tradução infantil.

Nenhuma destas versões tinha qualidade de texto final. Nestes três casos, o sistema apresenta-se como gerador de ideias que podem depois ser desenvolvidas pelo tradutor.

A avaliação de exercícios que envolvam TA, quer através de ferramentas específicas ou de outras ferramentas com esta capacidade, assenta, por um lado, em critérios habitualmente utilizados nas aulas de tradução (significado, adequação ao público-alvo, qualidade linguística, etc.). No entanto, é de grande importância que nos primeiros exercícios com estas ferramentas se abra espaço a uma análise pormenorizada e crítica dos resultados da tradução automática, designadamente a análise, a categorização e a descrição pormenorizada dos problemas encontrados, assim como das suas principais virtudes. Devem ser especialmente bonificadas as iniciativas de exploração e incorporação de funcionalidades avançadas destas ferramentas como estímulo à descoberta e à adaptação a novos meios.

Este tipo de exercícios pressupõe um maior grau de autonomização da aprendizagem e um modelo de trabalho mais centrado no aluno, sendo por isso indispensável muní-lo de competências reforçadas de análise e decisão.

Literacia em Tradução Automática

Para Lynne Bowker, a simplicidade na utilização da TA atualmente é sugestiva de um «aspeto mágico» (2023: 97), que leva os utilizadores a crer que a tradução é fácil e universal. Com efeito, a TA está amplamente acessível e também os alunos de tradução a utilizam, com melhores ou piores intuítos e com maior ou menor ingenuidade.

Importa, por isso, promover a literacia em TA no âmbito do ensino da tradução e, possivelmente, também no ensino das línguas. As competências clássicas devem ser reforçadas com a competência tecnológica, nomeadamente o manuseio das ferramentas e a criação de estratégias para o seu uso nos mais diversos tipos de projetos.

Mostrar que a «caixa negra» tem fragilidades, conhecê-las e colmatá-las. Sabendo como funcionam, é mais fácil obter melhores resultados e explorar mais alternativas. Um bom conhecimento do potencial e das limitações destas ferramentas permite usá-las apenas quando for adequado. Este conhecimento, apesar de ser desenvolvido dentro do campo da tradução, deve ser exportado para fora deste setor. Na verdade, como vimos anteriormente, a maior parte da utilização da tradução automática não advém de tradutores humanos. Bowker (*ibid.*) enuncia como principais objetivos da literacia em TA: melhorar a compreensão do funcionamento destas ferramentas, avaliar quando e onde existem condições para usá-las com êxito e aprender a trabalhar com elas de forma mais eficiente para obter resultados melhores.

Conclusão

As novas tecnologias de tradução automática e transformadores generativos pré-treinados (GPT) têm um enorme potencial ao nível da acessibilidade da tradução e podem também ter um significativo custo social para os profissionais da tradução, sobretudo os que trabalham nos domínios mais suscetíveis a estas tecnologias e que não se adaptem a elas. Neste último aspeto, é possível encontrar semelhanças com o que aconteceu há cerca de 30 anos quando se massificaram as ferramentas de tradução assistida por computador (CAT), assentes em memórias de tradução. Apesar de uma grande resistência inicial, foram gradualmente aceites e, ao invés de promoverem uma saída maciça de profissionais do setor, foram essenciais para a expansão do mercado da tradução e para a absorção de novos profissionais. A nível do ensino superior, após alguma resistência inicial, são hoje poucos os cursos que não preveem unidades curriculares específicas sobre ferramentas de tradução.

Há, porém, uma diferença significativa entre a TA/GPT e as CAT: enquanto as CAT eram ferramentas desenvolvidas para tradutores e para a prática profissional da tradução, a TA/GPT tem um campo de aplicação muito mais vasto e difuso, o que exigirá um esforço significativo de acompanhamento e exploração das funcionalidades para selecionar as mais úteis a cada tarefa de tradução e promover a sua didatização. É, todavia, previsível que comecem a surgir versões mais especializadas para profissionais da tradução.

No entanto, por ora, existem práticas como os exemplos aqui apresentados que se pode adotar na sala de aula para sensibilizar os alunos para as limitações da tecnologia e capacitá-los para explorar o potencial destas ferramentas, promovendo a literacia em tradução automática.

BIBLIOGRAFIA FINAL

- Austermühl, Frank (2001). *Electronic Tools for Translators*. Manchester: St. Jerome.
- Bowker, Lynne (2023). *De-Mystifying Translation: Introducing Translation to Non-Translators*. London: Routledge.
- Pym, Anthony; Torres-Simón, Ester (2021). Is Automation Changing the Translation Profession?. *International Journal of the Sociology of Language*, 2021, 270, pp. 39-57.
- Somers, Harold [Ed.] (2003). *Computers and Translation*. Amsterdam: John Benjamins Publishing Co.
- Walker, D. P. (1972). Leibniz and Language. *Journal of the Warburg and Courtauld Institutes*, 35, 1, pp. 294–307.