



A TRADUÇÃO AUTOMÁTICA COMO TECNOLOGIA DE GUERRA¹

Jacqueline LÉON²
(Laboratoire d'Histoire des Théories Linguistiques, CNRS
Université Paris Cité, Université Sorbonne Nouvelle)

Tradução de Jorge Viana de MORAES
(USP – Universidade de São Paulo)

RESUMO: Este artigo examina a fase inicial e institucional da automatização da linguagem, que se manifestou na Tradução Automática (TA) no início dos anos 1950 e, posteriormente, na Linguística Computacional no início dos anos 1960. O estudo argumenta que a instauração dessa nova orientação para as ciências da linguagem (c. 1948–1966) foi um evento abrupto, impulsionado por uma intervenção institucional massiva no financiamento e na direção da ciência após a Segunda Guerra Mundial. Em particular, a TA foi inicialmente estabelecida nos Estados Unidos como uma tecnologia de guerra, com vastos recursos estatais alocados para atender a demandas estratégicas e políticas urgentes do pós-guerra. Em seguida, e sobre as fundações desta iniciativa, a Linguística Computacional foi igualmente e abruptamente promovida como a “nova linguística”. A análise foca em como esse contexto de origem, focado em tecnologia e estratégia militar, moldou e impôs um novo horizonte de retrospectiva para as ciências da linguagem.

PALAVRAS-CHAVE: Tradução Automática (TA). Tecnologia de Guerra. Linguística Computacional. Warren Weaver. Yehoshua Bar-Hillel. Relatório ALPAC (Automatic Language Processing Advisory Committee). História da Ciência.

LA TRADUCTION AUTOMATIQUE COMME TECHNOLOGIE DE GUERRE

¹ Este artigo é uma versão reformulada do capítulo 1 de Léon (2015) e do capítulo 2 de Léon (2021).

² Jacqueline Léon é Diretora Emérita de Pesquisas no *Centre National de la Recherche Scientifique* (CNRS) na França. Possui doutorado em Linguística e seus principais tópicos de pesquisa incluem a história da automação das ciências da linguagem, tradução automática e o processamento de linguagem natural. Possui graduação em Inglês - Université Paris 7 (1970), mestrado em Linguística - Université Paris 7 (1971) e doutorado em Linguística - École des Hautes Études en Sciences Sociales (1996). E-mail: <jacqueline.leon@u-paris.fr>.

RÉSUMÉ: La première phase de l'automatisation du langage s'est constituée en deux temps: la traduction automatique (TA) au début des années 1950 et la linguistique computationnelle au début des années 1960. Cette automatisation s'est accompagnée de l'instauration brutale d'un nouvel horizon de rétrospection pour les sciences du langage qui s'est effectuée sur une période très courte d'une quinzaine d'années (1948-1966). Elle a été la conséquence d'une intervention massive des institutions qui ont pris des décisions drastiques d'orientation et de financement de la science dans le sillage des bouleversements mondiaux de la Seconde Guerre mondiale. C'est ainsi que la Traduction Automatique, en tant que technologie de guerre, a été instituée aux États-Unis par des instances d'état qui y ont consacré des moyens considérables pour répondre à une demande stratégique et politique propre à cette fin de Seconde Guerre mondiale. À sa suite, et sur ses cendres, la Linguistique Computationnelle a été érigée comme « nouvelle linguistique » de façon tout aussi brutale. L'analyse se concentre sur la manière dont ce contexte d'origine, axé sur la technologie et la stratégie militaire, a façonné et imposé un nouvel horizon de rétrospection pour les sciences du langage.

MOTS-CLES: Traduction Automatique (TA). Technologie de Guerre. Linguistique Computationnelle. Warren Weaver. Yehoshua Bar-Hillel. Rapport ALPAC (Automatic Language Processing Advisory Committee). Histoire de la Science.

INTRODUÇÃO

A primeira fase da automatização da linguagem constituiu-se em dois momentos: a tradução automática (TA) no início dos anos 1950 e a linguística computacional no início dos anos 1960. Essa automatização foi acompanhada pela instauração brutal de um novo *horizonte de retrospectão* para as ciências da linguagem, que ocorreu num período muito curto de cerca de quinze anos (1948-1966). Foi a consequência de uma intervenção maciça das instituições que tomaram decisões drásticas de orientação e de financiamento da ciência, na sequência das perturbações mundiais da Segunda Guerra Mundial. Foi assim que a tradução automática, enquanto tecnologia de guerra, foi instituída nos Estados Unidos por instâncias estatais que lhe dedicaram meios consideráveis para responder a uma exi

gência estratégica e política própria do fim desta Segunda Guerra Mundial. Na sua sequência, e sobre as suas cinzas, a linguística computacional foi erigida como uma “nova linguística” de forma igualmente brutal.

Três relatórios e duas personalidades excepcionais, Warren Weaver e Yehoshua Bar-Hillel, presidiram a esta implementação do novo horizonte de retrospecto em nível institucional:

1949 [1955] : *Translation*, Warren Weaver;

1960: “The present status of automatic translation of languages”, Yehoshua Bar-Hillel;

1966: *Language and Machines. Computers in Translation and Linguistics*, Comitê Consultivo de Processamento Automático de Linguagem (ALPAC) do Conselho Nacional de Pesquisa.

As condições dessa implementação levantam a questão de saber como um domínio científico-técnico pode ser imposto do exterior, por instituições, e sem uma ancoragem disciplinar específica em centros universitários. Pode-se até perguntar se tal domínio, pensado acima de tudo como uma tecnologia, pode transformar-se numa disciplina autônoma, ou se está condenado a desaparecer caso os resultados se revelem insuficientes. Para dar conta deste momento de instauração, que se pode qualificar de evento de uma forma histórica específica, optar-se-á por uma história institucional que por vezes assume a forma de uma narrativa. Esse evento constituirá um “marco de memória disciplinar” (Puech, 2008) para a linguística computacional e para o TAL.

1. A TA: UMA HISTÓRIA CURTA?

Segundo o relato feito por Martin Joos (1956) sobre o primeiro volume coletivo dedicado à TA, publicado em 1955, as primeiras discussões sobre tra

dução automática (TA) ocorreram já em 1943, no centro de criptografia do *Signal Intelligence Service*, instalado em Arlington Hall, Virginia. Essas discussões, muito frequentes e substanciais, opunham dois grupos: de um lado, aqueles que desenvolviam e utilizavam computadores para analisar “textos misteriosos”; de outro, aqueles que traduziam textos de uma língua para outra, uma tarefa mais rotineira e bem conhecida. Enquanto os tradutores defendiam o seu método, bem mais sutil, baseado na intuição e, portanto, parcialmente inconsciente, os criptoanalistas desafiavam-nos a tornar inteligível o processo de tradução humana, de modo a possibilitar sua mecanização. O prestígio, observa Joos, estava claramente do lado da matemática e das máquinas: “In other words, it was mathematics and the machines that enjoyed the higher prestige; or in other words again, those transactions took place in the matrix of today’s dominant culture”³ (Joos, 1956, p. 293) – frase premonitória da virada para a automatização-matematização da linguagem.

As primeiras experimentações só começam efetivamente após a guerra, por iniciativa de Warren Weaver, tanto na Grã-Bretanha, a partir de 1948, quanto nos Estados Unidos, em 1949⁴. Os primórdios da história da TA, se a considerarmos prioritariamente como uma história institucional, são marcados por três relatórios (Weaver, Bar-Hillel, ALPAC), que delimitam uma periodização em três tempos: um início, um meio e um fim.

O memorando *Translation*, de Warren Weaver, em 1949 (Weaver 1955), impulsionou a criação de centros de tradução automática, em grande parte nas universidades, com a missão de produzir em série traduções de textos ci-

³ Tradução francesa (J. Léon): « En d’autres termes, c’étaient les mathématiques et les machines qui jouissaient du plus grand prestige; en d’autres mots encore, ces transactions eurent lieu dans la matrice même de la culture dominante d’aujourd’hui ». Tradução para o português (J. V. Moraes) “Em outras palavras, eram as matemáticas e as máquinas que gozavam de maior prestígio; em outras palavras ainda, essas transações tiveram lugar na própria matriz da cultura dominante de hoje”.

⁴ Sobre os primórdios da TA, ver Locke et al. (1955) e Hutchins (1986, 2000a).

entíficos do russo para o inglês. Bar-Hillel é nomeado o primeiro pesquisador em tempo integral em TA, em 1951, no MIT. Quando os grupos de TA começam, com investimentos maiores ou menores, as suas experiências, a legitimidade da TA ainda não está assegurada: erros, tradução imperfeita, problemas linguísticos mal avaliados. É o que transparece nas conclusões do primeiro colóquio organizado no MIT em 1952 por Bar-Hillel, que recomenda uma tradução humana assistida por computador, em vez de uma tradução inteiramente automatizada (Bar-Hillel 1953a). Assim, apesar do sucesso midiático da primeira demonstração em computador em 1954⁵, que será seguida da criação de novos centros de TA, os financiadores começam a duvidar da eficácia dessa nova tecnologia. Em 1958, Bar-Hillel será encarregado de elaborar um relatório sobre a TA para a NSF (*National Science Foundation*); ele investiga a vintena de equipes existentes no mundo: 10 centros nos Estados Unidos, 7 na URSS, 2 na Grã-Bretanha e 1 na Itália⁶.

A argumentação de Bar-Hillel leva a um julgamento taxativo:

- (i) a TA suscitou investimentos enormes, em termos de recursos financeiros e humanos;
- (ii) a “Fully automatic high quality translation” (tradução de altíssima qualidade inteiramente automatizada), defendida por vários grupos de TA, é uma ilusão e um objetivo irrazoável, mesmo para textos científicos;
- (iii) pouquíssimos problemas linguísticos foram resolvidos, e os mais difíceis permanecem sem solução;

⁵ A primeira demonstração pública de TA em computador ocorreu em janeiro de 1954, em Nova York, em uma máquina IBM. Foi organizada pelo grupo de TA da Georgetown, dirigido por Leon Dostert. Trata-se da tradução de algumas frases do russo para o inglês com auxílio de um léxico bilíngue russo-inglês de 150 termos e de uma “sintaxe operacional” composta de seis operações que regulavam o parsing. Paul Garvin, da Georgetown University, e Peter Sheridan, da IBM, foram seus principais idealizadores (Dostert, 1955).

⁶ Para uma história detalhada dos primórdios da tradução automática, ver Hutchins (1986a, 2000a).

(iv) o único objetivo “razoável”, para uma comercialização de tradução científica de boa qualidade, é a tradução assistida por computador, com pelo menos uma intervenção humana sobre a saída da máquina, ou seja, a pós-edição; “razoável” significando ao mesmo tempo possível em termos de viabilidade técnica e científica, e competitivo em termos de custo em comparação com a tradução humana.

Esse relatório causa grande repercussão quando de sua publicação em 1960 (Bar-Hillel 1960). Redigido por alguém tão respeitado quanto Bar-Hillel, lança suspeita e descrédito sobre todos os grupos de TA. A seguir, o papel institucional é assumido pela criação, em 1964, do comitê ALPAC, que põe fim aos financiamentos. O relatório *Language and Machine. Computers in Translation and Linguistics*, publicado em 1966, terá consequências importantes não apenas para o destino da TA, mas também para o estatuto e a reorganização da linguística nos Estados Unidos. O ALPAC examina três pontos, com base em testes específicos:

- (i) as necessidades de tradução nas agências governamentais e na comunidade científica;
- (ii) o atendimento dessas necessidades pelos serviços de tradução;
- (iii) as vantagens e desvantagens da TA, em comparação com a tradução humana, em termos de custo e qualidade.

Em conclusão, o ALPAC apresenta um conjunto de recomendações. Não há escassez de tradutores nos Estados Unidos; ao contrário do que se chegou a afirmar, a oferta de tradutores é maior do que a demanda. A literatura científica traduzida automaticamente, sem pós-edição, é muito difícil de ler e chega a ser fonte de erros. Além disso, todos os depoimentos atestam que as saídas de TA com pós-edição são mais caras do que a tradução humana. O comitê conclui que não é

útil gastar somas consideráveis de dinheiro para mecanizar uma pequena indústria, já em declínio econômico, que emprega menos de 5.000 pessoas. O que se considera positivamente, em contrapartida, são as ferramentas automatizadas de apoio à tradução. Com base em experiências conduzidas na Alemanha e em Luxemburgo, onde a União Europeia começava a desenvolver bancos terminológicos – estando, nesse ponto, à frente dos Estados Unidos –, o ALPAC recomenda o desenvolvimento de glossários automáticos, bancos terminológicos, consulta automática a dicionários bilíngues, bem como a PAC (*computerized publishing* – publicação assistida por computador), que começava a apresentar resultados muito positivos, sobretudo na imprensa.

Do ponto de vista da história das ciências da linguagem, o relatório do ALPAC contém outro elemento importante. Ele confere legitimidade à linguística computacional como uma “nova linguística” fundada na gramática gerativa e transformacional de Chomsky, em substituição à TA.

Pode-se objetar a essa história curta (1949-1966) que a TA não se encerrou com o relatório do ALPAC. Centros, especialmente na Europa, e em particular na França, são criados após a publicação do relatório Bar-Hillel. Mesmo nos Estados Unidos, houve locais em que as experimentações prosseguiram após o relatório Bar-Hillel e mesmo após o relatório do ALPAC. É o caso dos centros da Bunker-Ramo, da *Georgetown University* e da IBM, que tinham ambições industriais e comerciais. Paul Garvin, autor das regras gramaticais da primeira demonstração em computador de 1954, continua trabalhando com TA no centro da *Bunker-Ramo* até 1969, ano em que os recursos são cortados por causa da guerra do Vietnã. Em 1965, duas das quatro máquinas de tradução construídas pela IBM (Alps e Mark II), baseadas numa tecnologia es

pecífica (a memória fotocópica⁷), ainda estão em operação em serviços governamentais norte-americanos.

Após 1966, contudo, a TA perde sua ambição de tecnologia de ponta autônoma. Ela se torna uma parte do TAL e permanece por muito tempo eclipsada pela linguística computacional e pela documentação automática⁸. No plano teórico, é a linguística computacional que, na esteira e sobre as cinzas da TA, inaugura a virada de automatização-matematização das ciências da linguagem. É o relatório do ALPAC, em 1966, que a legitima como “nova linguística”.

2. A TA COMO TECNOLOGIA DE GUERRA

Essa periodização curta, em três tempos bem delimitados, corresponde na realidade à história da TA como tecnologia de guerra, enraizada nas ciências e na cultura de guerra e constitutiva do novo horizonte de retrospectão.

2.1 As ciências da guerra, segundo Dahan e Pestre (2004)

Vários traços caracterizam as ciências da guerra⁹:

(i) a interpenetração entre ciência e engenharia. Vemos grandes matemáticos tornarem-se engenheiros, como John von Neumann, quando se dedicou ao projeto dos primeiros computadores digitais, ou, inversamente, engenheiros que também são teóricos e projetistas de sistemas de tratamento de sinal, como Claude Shannon na Bell. Lembra-se também do matemático Alan

⁷ As primeiras máquinas de traduzir, *The USAF Automatic Language Translator Mark I* (1958) e *The USAF Automatic Language Translator Mark II* (1964), foram construídas nos Estados Unidos pela Telemeter-Magnetics Inc. para a Força Aérea norte-americana. Utilizavam a memória fotocópica, de Gilbert King, que combinava altíssima capacidade de armazenamento com acesso muito rápido (Léon, 1992; Hutchins, 2000b).

⁸ Para uma periodização da história da TA e do TAL até os dias atuais, ver o anexo abaixo.

⁹ Segundo Pestre, a cultura de guerra, comum aos Estados Unidos e ao Reino Unido, surge no fim da década de 1930 para aperfeiçoar o sistema de detecção por radar que protegia a Grã-Bretanha diante da ameaça nazista.

Turing, envolvido tanto na decodificação da máquina Enigma quanto no desenvolvimento dos primeiros computadores;

(ii) a cultura científica de guerra, herdada diretamente do segundo conflito mundial e perpetuada na Guerra Fria, é uma cultura da urgência e da mobilização permanente. Os cientistas-engenheiros são movidos pela crença de que a ciência deve ser capaz de resolver todos os problemas;

(iii) o desaparecimento das restrições financeiras para esses cientistas, aos quais a guerra ofereceu possibilidades ilimitadas de inovar;

(iv) a onipresença do Estado. Observa-se uma “nacionalização” generalizada das ciências no século XX, em particular a partir de 1945, com a criação de grandes organismos nacionais de pesquisa. Nos Estados Unidos, isso se traduz no financiamento crescente, pelos militares, da pesquisa que, antes da guerra, era financiada pela indústria. Esse tipo de financiamento da pesquisa se desenvolverá de modo maciço no pós-guerra sob o nome de “military-industrial-academic complex” (complexo militar-industrial-acadêmico), expressão cunhada pelo presidente Eisenhower em 1961 e retomada pelo senador J. *William Fulbright* em 1968 (Leslie, 1993; Giroux, 2007);

(v) o estabelecimento de uma “abordagem operacional” (operational research ou OR)¹⁰. Tomando como modelo os militares, os cientistas-engenheiros generalizaram uma abordagem operacional comum: definir com precisão o objetivo, que deve ser único e cuja realização deve poder ser acompanhada e medida; criar um grupo de intervenção que reúna todas as competências possíveis; analisar a situação assegurando-se a contribuição das ciências, das ciências da engenharia e das ciências sociais. No pós-guerra, a Rand Corpora-

¹⁰ A noção de pesquisa operacional foi desenvolvida pelos britânicos durante a Segunda Guerra Mundial para avaliar e aumentar a eficácia de novas armas, como bombardeiros, mísseis de longo alcance, torpedos e radares (Fortun *et al.* 1993, Abella, 2008).

tion, primeiro “think tank” do período, é quem melhor dá continuidade a essa cultura científica de guerra. A Rand se propõe a desenvolver a “vida racional”, segundo a qual sistemas técnicos e humanos (sistemas de transporte, um exército em campanha ou uma sociedade) são conjuntos coordenados, otimizáveis em função de metas e critérios comuns. Essa vida racional implica substituir o trabalho humano por uma automatização generalizada. Seguindo os princípios da cibernética, essa automatização consiste em apagar a fronteira entre homem e máquina por meio da algoritmização;

(vi) o MIT: o nervo das ciências da guerra. Fundado no século XIX, o MIT (*Massachusetts Institute of Technology*) está muito à frente no que diz respeito à concentração e nacionalização da ciência, pelo menos no que toca a verbas de origem militar. Durante a guerra, é sede da seção D2 “fire control”, encarregada, entre outras missões, da defesa antiaérea, dos radares e dos servomecanismos, reunindo grandes cientistas na interface entre ciência e engenharia: Norbert Wiener, John von Neumann, Claude Shannon, Vannevar Bush e Warren Weaver, entre outros. É no MIT que Bar-Hillel é nomeado, em 1951, para dirigir o centro de TA do *Research Laboratory of Electronics*;

(vii) o desenvolvimento de novos conceitos e de novas ferramentas. Essa interação entre cientistas e engenheiros levou, em contrapartida, ao desenvolvimento de conceitos e ferramentas que reforçaram ainda mais essa tendência. Assim, novas teorias (cibernética, teoria da informação) e disciplinas mais tradicionais (lógica, matemática, física), em interação inédita com as ciências da engenharia (telecomunicações, criptografia, eletrônica ou defesa antiaérea), conduziram à construção de novos dispositivos tecnológicos, como radares e computadores eletrônicos. No que nos interessa, são a teoria da informação, a lógica, a estatística e a probabilidade, o cálculo numérico e a crip

tografia que servirão de horizonte de retrospecto para a constituição da TA como primeira aplicação não numérica dos computadores eletrônicos.

Assim, a TA não é criada ex nihilo, ela possui seu próprio horizonte de retrospecto. São as duas personalidades que impulsionaram e orientaram as pesquisas em TA, Warren Weaver e Yehoshua Bar-Hillel, que irão contribuir para estabelecer esse novo horizonte, do qual a linguística, é importante notar, não faz parte.

2.2 Warren Weaver (1894-1978)

A figura de Warren Weaver é absolutamente central no desenvolvimento dessa cultura científica de guerra. Matemático, especialista em probabilidade, ele também é apaixonado por engenharia mecânica e elétrica e começa lecionando matemática no então recém-fundado *Caltech* (posteriormente, *California Institute of Technology*) e depois na Universidade de Wisconsin, onde fez seus estudos. A partir de 1931, e até o fim de sua carreira, dirige a Divisão de Ciências Naturais da *Rockefeller Foundation*, onde promove programas de pesquisa em biologia experimental e molecular, tanto nos Estados Unidos quanto na Europa.

Durante a Segunda Guerra Mundial, a convite de Vannevar Bush, o inventor do computador eletrônico analógico no MIT, Weaver integra o poderosíssimo *Office of Scientific Research and Development*, o coração do dispositivo militar-científico. Ali, ele dirige a seção D2 “fire control” e cria o *Applied Mathematics Panel* (um dos componentes do *Office of Scientific Research and Development*), que reagrupa várias centenas de matemáticos para realizar estudos demandados pelo esforço de guerra e pela defesa em domínios extremamente variados, incluindo computadores, programação, estatística, cibernéti-

ca derivada dos sistemas de defesa etc. Após a guerra, em 1945, ele retorna à *Rockefeller Foundation* e participa de várias instituições governamentais voltadas à organização e ao financiamento da pesquisa. São *organismos civis*, como a NSF (*National Science Foundation*) e a *American Association for the Advancement of Science*, ou militares, como o *Naval Research Advisor Committee*. Além disso, é um dos membros fundadores da *RAND Corporation* (*Research and Development*), explicitamente criada em 1945-1946 pela Força Aérea dos Estados Unidos para ajudá-la a vencer guerras¹¹, reunindo lógicos e matemáticos lógicos (entre eles John von Neumann) e economistas oriundos do *Applied Mathematical Panel*.

Weaver sabe que é mais um facilitador, um divulgador, do que propriamente um pesquisador (Weaver 1970). Esse é o papel que desempenha na promoção da teoria da informação. Solicitado por Chester Barnard, então presidente da Rockefeller Foundation, ele redige uma apresentação da teoria da informação para não especialistas no volume coassinado com Shannon e publicado em 1949. A partir de então, a teoria da informação, até então restrita a alguns especialistas em telecomunicações, passa a ter maior difusão e desperta o interesse de cientistas em diversas disciplinas.

É Warren Weaver, ator central da cultura de guerra, dotado ao mesmo tempo de uma visão global das ciências e de grande talento como incentivador, quem terá a ideia de promover a tradução automática. A grande questão do pós-guerra é a Guerra Fria, na qual, dentro da competição científica e militar com os soviéticos, a tradução de textos científicos russos para o inglês ocupa um lugar

¹¹ “Its charter was clear: ‘Project RAND is a continuing program of scientific study and research on the broad subject of air warfare with the object of recommending to the Air Force preferred methods, techniques and instrumentalities for this purpose’” (Abella, 2008 p. 14). Tradução: “Sua missão era clara: ‘O Project RAND é um programa contínuo de estudo e pesquisa científica sobre o amplo tema da guerra aérea, com o objetivo de recomendar à Força Aérea métodos, técnicas e instrumentos preferenciais para esse fim’”.

crucial. Ao escrever seu memorando em 1949, a TA correspondia, para Weaver, ao tipo de objetivo definido pela RAND, no rastro da “cultura de guerra” e da abordagem operacional: automatização de um trabalho até então especificamente humano; definição de um objetivo único; produção em série de traduções científicas e técnicas; mobilização de recursos humanos e técnicos colossais para atingir o objetivo, privilegiando soluções formais em detrimento de soluções humanas. Além disso, Weaver é movido pela ideia de facilitar a comunicação entre cientistas nesses tempos de internacionalização da ciência. É assim que ele intitula o prefácio do primeiro volume coletivo sobre experiências de TA “The new tower”, em que remete ao mito de Babel (Locke *et al.*, 1955).

Warren Weaver utiliza os grandes meios para colocar em prática seu projeto de TA. Em nome da *Rockefeller Foundation*, propõe ao britânico Andrew Donald Booth, diretor do laboratório de informática do *Birkbeck College* (Universidade de Londres), ajudar os britânicos a obter um computador, com a condição de desenvolverem aplicações não numéricas, em particular a tradução automática.

Seu memorando *Translation*, sob aparência modesta – Weaver afirma em nota ser um neófito na matéria – é na realidade uma poderosa máquina de guerra. Ele o envia a cerca de duzentas pessoas, das quais uma trintena muito influente, que sabem perfeitamente que, se se lançarem com ele na TA, a organização e os financiamentos virão em seguida. Assim, em 1949, Weaver impulsiona pesquisas em três universidades norte-americanas, inclusive o MIT (no *Research Laboratory of Electronics*), onde Bar-Hillel será contratado em 1951, a Universidade de Washington, cujo centro de TA é dirigido pelo sinólogo Erwin Reifler (1903-1965), e a UCLA. Em 1958, haverá cerca de uma dúzia de centros de TA, entre os quais a *RAND Corporation*, que se tornará um dos maiores

grupos de TA (dirigido por David Hays e Abraham Kaplan). Os financiamentos da TA vêm da Força Aérea, da CIA e da NSF.

3. *UMA LINGUÍSTICA PARA ENGENHEIRO*

No âmbito da cultura científica de guerra, a linguística não tem lugar. O papel dos linguistas limita-se essencialmente ao ensino de línguas para militares. Poucos participam dos desenvolvimentos tecnológicos. No MIT, por exemplo, não há departamento de linguística antes de 1964, de modo que é no *Electrical Engineering Department* que o primeiro doutorando de Chomsky, Robert Lees, obtém seu PhD, em 1963.

Não é, portanto, a linguística que serve de referência teórica ou metodológica para a TA; e, quando Weaver levanta alguns dos problemas linguísticos colocados pela TA (expressões idiomáticas e palavras compostas, polissemia, ordem das palavras), é para descartá-los em seguida. Nos textos técnicos e científicos, escopo que ele reserva à TA, esses problemas são poucos e podem ser negligenciados. A TA não é uma disciplina, é uma tecnologia destinada a produzir em série traduções economicamente rentáveis. As questões linguísticas são, portanto, consideradas não prioritárias em comparação aos problemas ligados à arquitetura e à limitação da capacidade das máquinas: limitação de memória, logo, de velocidade e armazenamento. São essas as tarefas a que se dedicam os primeiros experimentadores, entre os quais encontramos, claro, especialistas em línguas vivas (chinês, línguas eslavas, espanhol) e tradutores (Leon Dostert), mas sobretudo engenheiros (Victor Yngve, Anthony Öttinger). Cabe ressaltar que os dois pioneiros, os britânicos Andrew Booth e Richard Richens, encorajados por Weaver a entrar nesse campo, não são nem um nem o outro especialistas em línguas: o primeiro é físico (cristaló

grafo) e o segundo, biólogo (especialista em plantas). São “inventores” que a máquina liberta de toda história e de todo enraizamento no passado, isto é, de qualquer horizonte de retrospectão.

Ao criar uma linguística para a máquina (*machine translation linguistics*, ver Reifler, 1955), os engenheiros negam à linguística qualquer legitimidade na automatização da tradução. Assim, ao contrário das convenções da lexicografia, os primeiros dicionários bilíngues automáticos são construídos a partir das formas abreviadas, e não dos lemas. As restrições gráficas impostas pela digitalização da informação levam os experimentadores a fabricar dicionários de raízes e terminações que não obedecem aos critérios funcionais de ordem fonética ou histórica dos gramáticos. Produzem-se, assim, “falsos” radicais, chamados bases de palavras, “falsas” desinências e redefinem-se afixos (no caso do russo, por exemplo). Em suas experiências de 1948, Booth e Richens (1955) definem a base como a parte mais longa, comum às diferentes formas de uma mesma entrada, que não seja ambígua. Por exemplo, para o verbo francês *saisir*, eles recomendam a escolha do radical *saisi*, de preferência a *sai-*, que tem parte comum com *savoir*. Em seu dicionário automático (russo-inglês), Cettinger (1955) propõe dois falsos radicais para o russo *okn* e *okon*, quando, para os gramáticos, trata-se apenas de variantes de uma mesma raiz. Os pesquisadores do Instituto de Matemática Steklov, em Moscou, acabam, na esteira dos americanos, usando os mesmos métodos. Enquanto o inglês possui apenas seis desinências *-ing*, *-ed*, *-er*, *-est*, *-th*, *-s*, os russos acrescentaram uma desinência “falsa”, o *-e* de *love*. A “falsa” desinência *e*, combinada com a forma *lov-* armazenada no dicionário, permite identificar as formas *love-s*, *lov-ing*, *lov-ed*, a partir das desinências comuns.

O mesmo se dá com a morfossintaxe. Reifler (1955), seguido por Micklesen (1956), define *operational form classes* (classes de formas operacionais) para a TA. Na linguística estrutural (Bloomfield 1933, p. 265), a definição das *form classes* é funcional. São as funções que determinam, de modo dinâmico, os padrões distribucionais em que uma forma linguística pode aparecer¹².

A elaboração de *form classes* para a TA obedece a dois princípios, um linguístico e outro técnico, que não se aplicam simultaneamente. O princípio linguístico consiste em estabelecer as classes de base das formas interativas e, em seguida, modificar essas classes em função das exigências tecnológicas. Essas classes básicas de formas interativas são potencialmente constituintes imediatos de primeiro nível e foram denominadas *mutual pinpointers* (*to pinpoint*: localizar, identificar com precisão) por Reifler. Ele dá o exemplo do artigo *den*, em alemão, que pode ser acusativo masculino singular ou dativo plural. Se é seguido de *Männern*, que é apenas dativo plural, a ambiguidade é eliminada: *den* e *Männern* desempenham aqui o papel de *mutual pinpointers*. Se dois grupos de formas são *mutual pinpointers* em uma dada construção, criam-se duas *form classes* separadas. Isso implica que, na fase automática, estabeleçam-se *form classes* mais amplas se incluírem *mutual pinpointers*, ao passo que, na linguística estrutural, tal subdivisão depende apenas da análise. Micklesen (1956, p. 346) oferece o seguinte exemplo. Para a linguística, formas como *doing* e *swimming* pertencem à classe dos verbos. Já para a TA, devido à função de localização recíproca (*mutual pinpointing*) que podem exercer, ora como verbos, ora como substantivos, por exemplo em *writing letters is*

¹² Bloomfield (1926) oferece dois exemplos, “Richard saw John” e “The man is beating the dog”, para estabelecer dois postulados (32, 33) destinados a definir as *form classes*:

“32. Def. As posições em que uma forma ocorre são suas funções. Assim, a palavra *John* e o sintagma *the man* têm as funções de ‘ator’, ‘meta’, ‘predicativo do sujeito’, ‘objeto de preposição’ e assim por diante.
33. Def. Todas as formas que têm as mesmas funções constituem uma *form-class*. Exemplos de *form-classes* do inglês são: noun-stems [radicais de substantivo], number-affixes [afixos de número], object expressions [expressões de objeto], finite verb expressions [expressões de verbo finito].” (Bloomfield 1926, p. 159)

very difficult, é preciso constituir com esses deverbais em *-ing* uma classe à parte, uma *operational form class*. As *operational form classes* não correspondem às classes resultantes da análise distribucional; elas incluem instruções de localização recíproca e desempenham uma função dinâmica no processo de TA. Essa distorção da análise morfológica suscita acaloradas discussões entre linguistas estruturalistas e experimentadores de TA.

Para estes últimos, a análise morfológica torna-se completamente subordinada à técnica, e criam-se objetos linguisticamente artificiais, sem que se respeite uma coerência teórica. Assiste-se, assim, à negação do horizonte de retrospectiva dos neobloomfieldianos que haviam dedicado várias décadas a descrever a morfofonologia das línguas. Os experimentadores descartam, sem uma crítica de fundo, os métodos e os resultados dos linguistas, simplesmente sob o argumento de que os trabalhos existentes não são adaptados à máquina. Com certo cinismo, reconhecem que os critérios que utilizam não são os da gramática, mas, se os critérios dos linguistas não coincidem com os seus, podem prescindir deles.

Esse estado de espírito persistirá e ainda persiste. Como lembra Melby (1992), citando os conselhos que recebeu, em 1972, de David Hays, um dos pioneiros da TA, não será encontrando o bom modelo formal em linguística que se resolverão os problemas da TA, mas sim oferecendo trabalho sólido, baseado em uma linguística simplificada. Essa também é a intuição de Peter Toma, o idealizador do Systran, que, ao contrário dos cientistas da época, não acredita que a linguística possa fornecer uma solução adequada ao processamento da linguagem por computador. Ele está convencido de que o tratamento da linguagem deve ser adaptado às possibilidades do computador, e não o inverso (Loffler-Laurian 1996). Pode-se igualmente citar a indignação suscitada por

uma nota publicada em um número de 1995 da revista *Traitement automatique des langues*, relatando as declarações de Frederick Jelinek, que, ao dirigir na IBM a equipe de pesquisa em reconhecimento de fala, teria afirmado: “cada vez que eu demito um linguista da minha equipe, o desempenho do nosso sistema melhora em 10%” (TAL, 1995, p. 69, nota 2)¹³.

4. A TA: UMA TECNOLOGIA IMPERFEITA

4.1 Uma tradução aproximada e grosseira

A TA é uma tecnologia, sem dúvida, mas uma tecnologia imperfeita. Weaver reconhece de antemão certos limites, sobretudo a incapacidade da máquina em traduzir textos literários que contêm elementos “ilógicos”, como emoção ou intuição. Ele menciona, a esse respeito, a posição de Norbert Wiener, a quem tentara convencer em uma troca de cartas em 1947 (Weaver, 1955). Wiener mostra-se cético quanto à possibilidade de TA, pois, segundo ele, traduzir uma língua para outra é passar de uma cultura a outra, algo que uma máquina é incapaz de fazer.

Weaver tem consciência de que uma tradução perfeita é ilusória e de que, mesmo no caso de traduções científicas e técnicas, é impossível garantir ausência total de erros. Textos científicos são, sem dúvida, mais fáceis de traduzir, menos ambíguos, com menor grau de polissemia, mas a tradução não será perfeita. Em compensação, têm a vantagem de poder ser traduzidos em massa; além disso, uma tradução aproximativa e grosseira, por exemplo palavra por palavra, deve bastar para um uso prático. Se as saídas de uma tradução palavra por palavra não são muito legíveis, são suficientemente compreen

¹³ Ver a resposta da redação de *TAL* intitulada “efficacité du TALN et linguistes” no número 1996-1 (p. 162), que relata as tensões entre industriais e linguistas.

síveis para permitir que os cientistas selecionem os artigos que merecem uma tradução humana.

Além disso, admite-se um certo percentual de erro dentro de um uso prático. Esse argumento é de peso. Ele foi constantemente utilizado pelos atores da TA e, posteriormente, do TAL, ainda hoje, embora o estatuto desses erros seja ambíguo. Retomando os argumentos de Loffler-Laurian (1996), pode-se questionar o significado de uma tradução “boa em 80%”. Isso quer dizer que 80% das frases se assemelham a frases escritas por humanos? Ou que 80% das palavras receberam um equivalente correto? Ou ainda que a tradução foi julgada aceitável por 80% das pessoas consultadas? Trata-se de uma questão fundamental para a TA — e para o TAL em geral.

Tal argumento coloca a questão de uma técnica que seria falível por de finição, algo bastante singular. Com efeito, quando mísseis guiados, radares ou outras tecnologias desenvolvidas no âmbito das ciências da guerra não funcionavam, buscava-se aperfeiçoá-los. Para os opositores das conclusões do ALPAC, esse é precisamente um dos argumentos utilizados para demonstrar que não se deu tempo suficiente para que a TA se aperfeiçoasse e provasse seu valor (Josselson, 1971).

4.2 Descompasso entre resultados medíocres, ambição dos atores e sucesso público

À imperfeição dos resultados, ou mesmo a sua mediocridade, e ao desaparecimento da esperança de criar uma máquina de traduzir, vem somar-se um terceiro fator de decepção: o descompasso entre o sucesso público da TA e as ambições moderadas de seus protagonistas. O entusiasmo suscitado pelas máquinas eletrônicas e a ilusão de seu potencial infinito já aparecem em 1949,

quando o New York Times publica um artigo sobre a máquina de traduzir intitulado “Electric brain able to translate foreign languages is built” (ver Hutchins 1997, p. 203). Esse entusiasmo suscita um tal eco no público que alguns pesquisadores se veem obrigados a desmentir as fantásticas promessas anunciadas pela imprensa (ver o desmentido de Koutsoudas, 1956)¹⁴.

Em 1954, a primeira demonstração em computador evidencia mais uma vez esse descompasso (ver Gordin, 2015): o paradoxo do sucesso público da TA, amplamente midiaticado, e das vãs esperanças assim alimentadas contrasta com a convicção da maioria dos atores de que a FAHQT (*Full Automatic High Quality Translation*) é uma ilusão e de que é preciso concentrar esforços na tradução assistida (por computador ou por humanos) e na análise sintática. Esse paradoxo será fatal para a TA. Vítima de seu próprio sucesso e das expectativas infundadas que suscita, ela será duramente criticada pelos relatórios de Bar-Hillel e do ALPAC, que põem fim às experiências.

5. CONCLUSÃO

A tradução automática, enquanto tecnologia de guerra, não nasceu do nada. As propostas de Weaver situam a TA no cruzamento de um feixe de disciplinas (lógica, matemática, estatística e probabilidade, neurologia e cibernética), das quais a teoria da informação é o elemento unificador e universal, mas do qual a linguística não faz parte. Entretanto, os primeiros experimentadores, alguns dos quais sequer pertencem a esse conjunto de disciplinas, ignoram, em nome da máquina, todo horizonte de retrospectiva. Isso tem um cus-

¹⁴ Costuma-se citar, frequentemente, para ridicularizar a fragilidade dos sistemas de TA, a tradução do inglês para o russo da passagem bíblica: “The spirit is willing, but the flesh is weak” (“o espírito é forte, mas a carne é fraca”), que, novamente traduzida do russo para o inglês, daria “The whisky is strong, but the meat is rotten” [“O uísque é forte, mas a carne está podre”] ou “The ghost is a volunteer but the meat is tender” [“O fantasma é voluntário, mas a carne é macia”]. Na realidade, tratar-se-ia de um erro de tradução humana, relatado por um jornalista (em 1956) e atribuído à máquina (ver Hutchins 1995 para detalhes).

to: uma das consequências de instaurar um novo domínio de pesquisa por uma decisão brutal é a necessidade de legitimá-lo. Assim, os atores desse período sentiram-se obrigados a uma autocongratulação permanente, num esforço constante de legitimação, desde o surgimento do novo campo e muito antes do relatório do ALPAC. Uma das maneiras de lidar com esse déficit de legitimação é inscrevê-lo na história. Desde a década de 1950, os experimentadores em TA passam a produzir introduções supostamente “históricas”, que na realidade são balanços retrospectivos com viés autocongratatório. Esses textos – o primeiro publicado já em 1955 – são introduções de colóquios ou de volumes coletivos, redigidos pelos próprios atores da TA (Locke e Booth, 1955; Dostert, 1957; Booth, 1958; Delavenay, 1959). Eles indicam que os atores elegeram explicitamente o início das experiências de TA como “marco de memória disciplinar” (Puech, 2008), embora esse marco, por causa do ALPAC, venha posteriormente a ser definitivamente marcado pelo opróbrio (*the infamous report*, segundo Hutchins, 1996). Ademais, encontramos, com a TA – cujos centros, em grande parte, estão localizados em departamentos de línguas de universidades –, diante de uma situação complexa.

Por um lado, a elaboração de uma linguística para a máquina pelos engenheiros representa uma anulação explícita do horizonte de retrospectão dos linguistas neobloomfieldianos, e não simplesmente um esquecimento (Auroux 1987) que seria necessário à inovação. O novo horizonte de projeção, estabelecido pela cultura de guerra e unificado pela teoria da informação, encontra-se, assim, desconectado de seu passado – ou até mesmo sem passado. Muitos atores da TA pertencem às disciplinas envolvidas e participaram da elaboração dessa nova cultura. A questão para os linguistas, ausentes dessa fase de criação, será a de se apropriar desse novo horizonte de retrospectão para transfor

má-lo em horizonte de projeção. A linguística computacional constitui uma primeira etapa dessa apropriação.

ANEXO: PERIODIZAÇÃO DOS DESENVOLVIMENTOS DA TA APÓS 1966

Pode-se prosseguir a história da tradução automática (TA) para além de 1966 e propor quatro períodos que conduzem à fase atual¹⁵:

(i) O primeiro período (1966-1980) é marcado pela sobrevivência e pela “força bruta” dos grandes sistemas de tradução automática, dos quais o mais notável é o Systran. O Systran é criado em 1967 por Peter Toma (nascido na Hungria em 1925), na esteira de vários sistemas de tradução automática desenvolvidos sucessivamente por Toma em computadores IBM. Em 1958, ele trabalha na Universidade de Georgetown no sistema linguístico GAT (*General Analysis Technique*) e no programa informático SERNA. Em 1961, cria a empresa *Computer Concepts*, onde desenvolve o Autotran, utilizando a linguagem Fortran, que se torna Systran em 1962. Ele continua a trabalhar no Systran na Alemanha e na Califórnia, onde cria sua própria empresa Latsec Inc. em 1968. Obtém um financiamento da NASA (*National Aeronautics and Space Administration*) em 1972. A Comissão Europeia, assim como a Euratom, adotaram o Systran como base para suas operações de tradução automática em 1976 (Toma 2000), colaboração que perdurou até o início dos anos 2000¹⁶. Do ponto de vista metodológico, o Systran é, inicialmente, um “sistema por transferência” herdado do “sistema direto” de Georgetown¹⁷. Desenvolvido originalmente como um sistema de tradu-

¹⁵ Para uma visão sintética da história das técnicas em TA, ver Léon (2006a e 2006b).

¹⁶ A colaboração foi encerrada após um processo ganho em 2010 pela Systran contra a Comissão Europeia, acusada de ter violado os direitos autorais e o *know-how* da empresa francesa de software Systran.

¹⁷ Três tipos principais de sistemas foram desenvolvidos ao longo dos anos pioneiros da tradução automática: (i) a tradução direta, ou tradução palavra por palavra, baseia-se essencialmente em um dicionário bilíngue (ver, por exemplo, o *Pidgin English* de Richens e Booth, 1955); (ii) o método por transferência baseia-se em um dicionário bilíngue e em programas de tradução por análise sintática. A análise da língua-fonte e a síntese da língua-alvo são relativamente independentes das pares de línguas fonte-alvo. Esses programas foram desenvolvidos no quadro da constituição da análise sintática automática como domínio autônomo; (iii) o método de língua intermediária baseia-se na análise semântica (para mais informações, ver Hutchins, 1986a, 1986b e Léon, 2021).

ção automática russo-inglês, outras pares de línguas são acrescentados a partir dos anos 1970: chinês-inglês, alemão-inglês, francês-inglês etc. Hoje, cerca de 55 línguas são contempladas por um *software* que funciona *on-line*.

(ii) O segundo período (1980-1990) é marcado pela aplicação da inteligência artificial simbólica à TA, pela virada japonesa e pela automatização da comunicação.

Os resultados do processamento automático de línguas (TAL) e da TA baseados na análise sintática revelam-se cada vez mais decepcionantes. O isomorfismo de representação postulado entre as línguas naturais e as linguagens lógico-algébricas formais é colocado em questão. Os sistemas baseados em IA assumem então a dianteira e utilizam uma representação do conhecimento fundada em lógicas não clássicas, topologia, semântica e pragmática. Esse período caracteriza-se pela integração de técnicas de inteligência artificial em sistemas de TA, tais como sistemas especialistas e bases de dados inteligentes. Esses sistemas recorrem a estratégias e técnicas mistas, combinando inteligência artificial e procedimentos sintáticos, que não existiam entre as décadas de 1960 e 1980. No entanto, esses sistemas híbridos ainda são fortemente questionados pela tradução automática assistida por computador, em particular pelos sistemas interativos de TA desenvolvidos nessa época (ver Nirenburg 1993 para um estudo de referência). A tradução assistida por computador oferece, de fato, ferramentas aperfeiçoadas aos tradutores, tais como dicionários eletrônicos bilíngues e bases de dados terminológicas.

Por outro lado, o desenvolvimento dos microcomputadores e dos softwares de processamento de texto no início dos anos 1980, bem como a democratização de seu uso, abre uma nova era para a TA, a do “marketing”. A TA passa então a ser utilizada para traduzir manuais de instruções, especificações de produtos e de máquinas. A virada “mercadológica” é impulsionada pelos japoneses, desejosos de introduzir seus produtos nos mercados mundiais no quadro de seu projeto de “Quinta Geração”. Seu objetivo é equipar cada microcomputador com

um sistema de TA, preparando assim, a longo prazo, o futuro de uma sociedade baseada na informação. Na esteira do projeto japonês, o Reino Unido lança o projeto Alvey e a Europa implementa o projeto de pesquisa transnacional *Esprit*. A partir dos anos 1980, a TA torna-se, portanto, parte integrante da engenharia linguística e inscreve-se no movimento mais amplo de automatização da comunicação que afeta o conjunto da sociedade.

(iii) Terceiro período (1990-2010): *data turn*, avanço tecnológico e uso generalizado de métodos estatísticos e probabilísticos.

Com o desenvolvimento tecnológico inédito dos computadores e o aparecimento dos microcomputadores, assiste-se ao retorno dos métodos empíricos. Esse renascimento é suscitado pelo sucesso dos métodos estocásticos no tratamento de sinais e no reconhecimento de fala. A tendência geral desses anos foi substituir o processamento “baseado em modelos formais”, funcionando de maneira dedutiva a partir de conhecimentos especializados operacionalizados por regras (autômatos, gramáticas, formalismos lógicos etc.), por métodos empíricos “baseados em dados”, funcionando de maneira indutiva e recorrendo a métodos estatísticos e probabilísticos em vez da expertise dos modeladores. Durante uma década (1990-2000), métodos linguísticos baseados em regras alternam com métodos probabilísticos de tradução automática, com o surgimento de métodos híbridos.

Além disso, grandes corpora de dados textuais tornam-se disponíveis, de modo que a memória de tradução e os corpora alinhados convertem-se nos principais componentes da tradução assistida por computador. A memória de tradução implica corpora bilíngues alinhados (ou sincronizados) já existentes, como os Hansards canadenses, o corpus francês-inglês das leis parlamentares canadenses, ou os textos institucionais da União Europeia traduzidos em múltiplas línguas. O alinhamento, isto é, o estabelecimento de correspondências entre frases, é realizado segundo critérios exclusivamente estatísticos, por exemplo explorando o fato de que frases longas na língua de partida tendem a ser traduzidas por frases longas na língua de chegada, e vice-versa. En-

quanto as abordagens baseadas em IA simbólica se fundavam em regras, as abordagens empíricas baseiam-se em estatísticas e evidenciam o papel crucial do léxico, em particular das colocações lexicais, para resolver constrangimentos locais, tais como ambiguidades semânticas. Além disso, surgem os primeiros sistemas de tradução on-line, como o *Babelfish*, desenvolvido pela Systran em 1998.

No início dos anos 2000, a TA, assim como o conjunto do TAL, é fortemente influenciada pelo data mining, que representa o estado da arte em tecnologia de sistemas de bases de dados. Novas tarefas de exploração de textos começam a aparecer (Manning & Schütze 1999). Os sistemas de recuperação de informação passam a dispor de capacidades de “perguntas-respostas” que aceitam questões formuladas em “linguagem natural”. A classificação consiste em associar automaticamente um texto a um rótulo dentre uma lista de rótulos possíveis; por exemplo, softwares de correio eletrônico distinguem automaticamente os “spams”. A extração de informação consiste em preencher campos de formulários em função do conteúdo de um texto.

(iv) A partir de 2010: desenvolvimentos recentes em TA e TAL: adoção generalizada do *Deep Learning* (Léon 2025).

A partir da década de 2010, o termo “Deep Learning” (DL) – em português, aprendizagem profunda – substitui o de “redes neurais” e se generaliza com o ressurgimento espetacular do interesse pelas abordagens conexionistas centradas em sistemas de Deep Learning. Distinguem-se duas abordagens principais em TAL:

(i) A abordagem “*bag of words*” (saco de palavras) reduz um texto ao conjunto de suas “formas de palavras”, sem levar em conta a ordem em que aparecem. A palavra perde assim seu caráter linguístico e torna-se uma unidade matemática. Isso permite transformar um corpus de textos em uma matriz de números, aplicando diretamente procedimentos já consolidados para tarefas de exploração de dados.

(ii) A abordagem por anotação conserva a ordem das palavras, segmentando o texto em unidades etiquetadas de maneira apropriada.

A partir de 2018, a tendência ao aprendizado profundo em TAL intensifica-se ainda mais com o surgimento dos grandes modelos de linguagem (LLM – *large language models*). Dois LLM maiores, provenientes de dois dos maiores centros de pesquisa privada em IA, OpenAI (Microsoft) e DeepMind (Google), foram apresentados ao público: BERT (*Bidirectional Encoder Representations from Transformers*) e GPT (*Generative Pre-trained Transformer*). Trata-se de um domínio em fortíssima expansão. Desde agosto de 2025, já se chegou à versão GPT-5. Enquanto em 2014 os sistemas de tradução automática apenas conseguiam oferecer traduções aproximativas para certos pares de línguas, os LLM atuais fornecem traduções de altíssima qualidade para numerosos pares de línguas, em particular para línguas bem documentadas como o inglês e o francês.

REFERÊNCIAS

- ABELLA, A. **Soldiers of Reason: The Rand Corporation and the Rise of the American Empire**. Orlando: Harcourt Inc., 2008.
- BAR-HILLEL, Y. **The Present State of Research on Mechanical Translation**. *American Documentation*, v. 2, p. 229-236, 1953a.
- BAR-HILLEL, Y. The Present Status of Automatic Translation of Languages. In: ALT, F. C. (ed.). **Advances in Computers**. London: Academic, 1960. v. 1, p. 91-141.
- BLOOMFIELD, L. A Set of Postulates for the Science of Language. **Language**, v. 2, p. 153-164, 1926.
- BOOTH, A. D. The history and recent progress of machine translation. In: BOOTH, A. D. et al. (ed.). **Aspects of Translation**. London: Secker and Warburg, 1958. p. 88-104.
- BOOTH, A. D.; RICHENS, R. H. Some Methods of Mechanised Translation. In: LOCKE, W. N.; BOOTH, A. D. (ed.). **Machine Translation of Languages**, 14 Essays. Cambridge, MA; New York: MIT/Wiley, 1955. p. 24-46.
- CONWAY, F.; SIEGELMAN, J. **Dark Hero of the Information Age. In search of Norbert Wiener the Father of cybernetics**. New York: Basic Books, 2005.



DAHAN, A.; PESTRE, D. (ed.). **Les sciences pour la guerre (1940-1960)**. Paris: Éditions de l'EHESS, 2004.

DELAVENAY, E. **La machine à traduire**. Paris: PUF, 1959.

DELAVENAY, E. **An Introduction to Machine Translation**. London: Thames & Hudson, 1960.

DOSTERT, L. The Georgetown-IBM experiment. In: LOCKE, W. N.; BOOTH, A. D. (ed.). **Machine Translation of Languages**, 14 Essays. Cambridge, MA; New York: MIT/Wiley, 1955. p. 124-135.

DOSTERT, L. Brief Review of the History of Machine Translation Research. **Monograph Series on Languages and Linguistics**, n. 10, p. 3-10, 1957.

FORTUN, M.; SCHWEBER, S. S. Scientists and the Legacy of World War II: the case of operations research (OR). **Social Studies of Science**, v. 23, p. 595-642, 1993.

GIROUX, H. A. **The University in Chains: Confronting the Military-industrial-academic Complex**. Boulder: Paradigm Publishers, 2007.

GORDIN, M. D. **Scientific Babel: How Science Was Done Before and After Global English**. Chicago; London: The University of Chicago Press, 2015.

HUTCHINS, W. J. **Machine Translation, Past, Present, Future**. Chichester: Ellis Horwood, 1986a.

HUTCHINS, W. J. Direct Translation Systems Since 1965. In: HUTCHINS, W. J. (ed.). **Machine Translation, Past, Present, Future**. Chichester: Ellis Horwood, 1986b. p. 209-223.

HUTCHINS, W. J. 'The whisky was invisible', or persistent myths of MT. **MT News International**, n. 11, p. 17-18, 1995.

HUTCHINS, W. J. ALPAC: The (In)famous Report. **MT News International**, n. 14, p. 9-12, 1996.

HUTCHINS, W. J. From First Conception to First Demonstration: the nascent years of Machine Translation, 1947-1954. **A Chronology. Machine Translation**, v. 12, n. 3, p. 192-252, 1997.

HUTCHINS, W. J. Gilbert W. King and the IBM-USAF Translator. In: HUTCHINS, W. J. (ed.). **Early Years in Machine Translation**. Amsterdam: Benjamins, 2000b. p. 171-176. (SiHoLS, 97).

JOOS, M. Review of “Machine Translation of Languages: Fourteen Essays, ed. by William N. Locke and A. Donald Booth 1955”. **Language**, v. 32, n. 2, p. 293-298, 1956.

JOSSELSOHN, H. H. Automatic Translation of Languages Since 1960: a linguist’s view. **Advances in Computers**, v. 11, p. 1-58, 1971.

KOUTSOUDAS, A. Report from the International Conference on Mechanical Translation (MIT, 20 October 1956). **Machine Translation**, v. 3, n. 2, p. 34, 1956.

LANGUAGE AND MACHINES. Computers in Translation and Linguistics. A report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences. In: NATIONAL RESEARCH COUNCIL. **Language and Machines. Computers in Translation and Linguistics**. Washington, DC, 1966.

LÉON, J. De la traduction automatique à la linguistique computationnelle. Contribution à une chronologie des années 1959-1965. **Traitement Automatique des Langues**, v. 33, n. 1-2, p. 25-44, 1992.

LÉON, J. La traduction automatique I: les premières tentatives jusqu’au rapport ALPAC. In: AUROUX, S. et al. (ed.). **Handbücher zur Sprach- und Kommunikationswissenschaft**. Berlin: Walter de Gruyter, 2006a. v. 3, p. 2767-2774.

LÉON, J. La traduction automatique II: développements récents. In: AUROUX, S. et al. (ed.). **Handbücher zur Sprach- und Kommunikationswissenschaft**. Berlin: Walter de Gruyter, 2006b. v. 3, p. 2774-2780.

LÉON, J. **Histoire de l’automatisation des sciences du langage**. Lyon: ENS Éditions, 2015.

LÉON, J. **Automating Linguistics**. Cham: Springer, 2021.

LÉON, J. Machine Translation, Computational Linguistics and Natural Language Processing, History of-. In: VAN HAL, T. (ed.). History of Linguistics. In: NESI, H.; MILIN, P. (ed.). **International Encyclopedia of Language and Linguistics**. 3. ed. Amsterdam: Elsevier, 2025. Disponível em: <https://doi.org/10.1016/B978-0-323-95504-1.00754-7>.

LESLIE, S. W. **The Cold War and American Science: the military-industrial-academic complex at MIT & Stanford**. New York: Columbia University Press, 1993.

LOCKE, W. N.; BOOTH, A. D. (ed.). *Machine Translation of Languages*, 14 Essays. Cambridge, MA; New York: MIT/Wiley, 1955.

LOFFLER-LAURIAN, A.-M. **La traduction automatique**. Lille: Septentrion, 1996.
MELBY, A. The Translator Workstation. In: NEWTON, J. (ed.). **Computers in Translation, A Practical Appraisal**. London: Routledge, 1992. p. 147-165.

MICKELSEN, L. R. Form Classes: structural linguistics and mechanical translation. In: HALLE, M. et al. (ed.). **For Roman Jakobson. Essays on the Occasion of his Sixtieth Birthday 11 Oct 1956**. The Hague: Mouton, 1956. p. 344-352.

NIRENBURG, S. (ed.). **Progress in Machine Translation**. Amsterdam: IOS Press, 1993.

OETTINGER, A. G. The Design of an Automatic Russian-English Technical Dictionary. In: LOCKE, W. N.; BOOTH, A. D. (ed.). **Machine Translation of Languages**, 14 Essays. Cambridge, MA; New York: MIT/Wiley, 1955. p. 47-65.

PUECH, C. Qu'est-ce que faire l'histoire du "récent"? In: DURAND, J.; HABERT, B.; LAKS, B. (ed.). **Congrès mondial de Linguistique Française – CMLF08**. Paris: Institut de Linguistique Française, 2008.

REIFLER, E. The Mechanical Determination of Meaning. In: LOCKE, W. N.; BOOTH, A. D. (ed.). **Machine Translation of Languages**, 14 Essays. Cambridge, MA; New York: MIT, 1955. p. 136-164.

TOMA, P. From Serna to Systran. In: HUTCHINS, W. (ed.). **Early Years in Machine Translation**. Amsterdam; Philadelphia: John Benjamins, 2000. p. 135-145.

WEAVER, W. Translation. In: LOCKE, W. N.; BOOTH, A. D. (ed.). **Machine Translation of Languages**, 14 Essays. Cambridge, MA; New York: MIT/Wiley, 1955. p. 15-23.

WEAVER, W. **Scene of Change. A Lifetime in American Science**. New York: Charles Scribner's Sons, 1970.