

UNIVERSIDADE PRESBITERIANA MACKENZIE

LUCIANA DEBONIS

UM ESTUDO HISTORIOGRÁFICO DA EVOLUÇÃO DA
TRADUÇÃO AUTOMÁTICA

São Paulo

2021

LUCIANA DEBONIS

UM ESTUDO HISTORIOGRÁFICO DA EVOLUÇÃO DA
TRADUÇÃO AUTOMÁTICA

Dissertação apresentada ao Programa de Pós-Graduação em Letras da Universidade Presbiteriana Mackenzie, como requisito parcial à obtenção de título de Mestre em Letras.

ORIENTADOR: Prof. Dr. Ronaldo de Oliveira Batista

São Paulo

2021

D287e Debonis, Luciana.
Um estudo historiográfico da evolução da tradução automática /
Luciana Debonis.
132 f. : il. ; 30 cm

Dissertação (Mestrado em Letras) – Universidade Presbiteriana
Mackenzie, São Paulo, 2021.
Orientador: Ronaldo de Oliveira Batista.
Referências bibliográficas: f. 129-132.

1. Tradução automática. 2. História da tradução automática. 3.
Mecanização da tradução. 4. Historiografia da linguística. I. Batista,
Ronaldo de Oliveira, *orientador*. II. Título.

CDD 418.02

Bibliotecária Responsável: Andrea Alves de Andrade - CRB 8/9204


LUCIANA DEBONIS

UM ESTUDO HISTORIOGRÁFICO DA EVOLUÇÃO DA
TRADUÇÃO AUTOMÁTICA

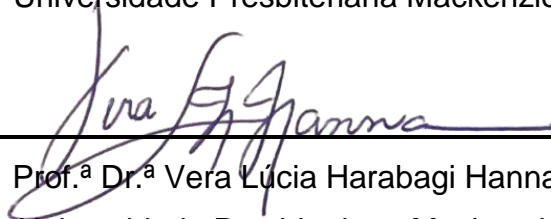
Dissertação apresentada ao Programa de Pós-Graduação em Letras da Universidade Presbiteriana Mackenzie, como requisito parcial à obtenção de título de Mestre em Letras.

Aprovada em 04 de fevereiro de 2021.

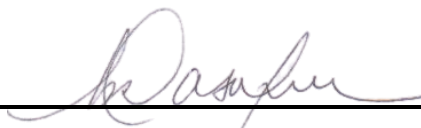
BANCA EXAMINADORA



Prof. Dr. Ronaldo de Oliveira Batista
Universidade Presbiteriana Mackenzie



Prof.ª Dr.ª Vera Lúcia Harabagi Hanna
Universidade Presbiteriana Mackenzie



Prof.ª Dr.ª Nancy Casagrande
Pontifícia Universidade Católica de São Paulo

FOLHA DE IDENTIFICAÇÃO DA AGÊNCIA DE FINANCIAMENTO

Autor: Luciana Debonis

Programa de Pós-Graduação *Stricto Sensu* em Letras

Título do Trabalho: Um estudo historiográfico da evolução da tradução automática

O presente trabalho foi realizado com o apoio de ¹:

- CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico
- FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo
- Instituto Presbiteriano Mackenzie/Isenção integral de Mensalidades e Taxas
- MACKPESQUISA - Fundo Mackenzie de Pesquisa
- Empresa/Indústria:
- Outro:

¹ **Observação:** caso tenha usufruído mais de um apoio ou benefício, selecione-os.

Aos meus pais, Divino e Zélia, por me inspirarem a realizar este trabalho.

AGRADECIMENTOS

Ao Prof. Dr. Ronaldo de Oliveira Batista, por me guiar pelos caminhos da historiografia compartilhando seu conhecimento intelectual de maneira fascinante.

À Prof.^a Dr.^a Vera Lúcia Harabagi Hanna e à Prof.^a Dr.^a Nancy Casagrande, pelos importantes apontamentos e sugestões de melhorias neste trabalho durante o exame de qualificação.

À Prof.^a Dr.^a Neusa Maria Oliveira Barbosa Bastos e à Prof.^a Dr.^a Maria Mercedes Saraiva Hackerott, por compartilharem suas visões a respeito do andamento da pesquisa no decorrer deste trabalho.

Às amigas Olívia Carvalho, Adriana Ribeiro Mendes, Jéssica Máximo Garcia e Gabriela Zhongxin Zhong, pelo apoio no decorrer do curso do programa de mestrado que possibilitou construirmos novos laços de amizade.

Ao meu primo Bruno Bezerra de Araújo, por sua ajuda no processo inicial de elaboração do projeto de pesquisa desta dissertação para ingresso no PPGL da Universidade Presbiteriana Mackenzie.

À minha prima-irmã, Patricia da Costa, e ao meu primo-sobrinho, Guilherme da Costa Souza, por serem forças motoras neste projeto de vida.

Aos meus pais, Divino e Zelia Debonis, e minha irmã, Silvana Debonis, pelo apoio incondicional no decorrer da realização deste trabalho, em especial, no momento em que tive uma intercorrência médica.

RESUMO

A tradução automática (TA) é uma modalidade de tradução que serve de recurso para facilitar o estreitamento de barreiras linguísticas. O surgimento dos primeiros experimentos de máquinas de tradução foram na década de 1940 nos EUA, os quais investiram em estudos nessa área por quase duas décadas, até serem atingidos por um corte nesses investimentos no final dos anos de 1960. A retomada das pesquisas aconteceu apenas a partir dos anos de 1980 e inaugurou um novo padrão de estudos científico no campo. Na década seguinte, a abertura do mercado de tradução técnica promovida pela oferta de ferramentas para automatização da tradução impulsionou o progresso das pesquisas. Desde a virada do milênio, os estudos em TA demonstraram avanços significativos em virtude da evolução tecnológica que permitiu o desenvolvimento de ferramentas de TA com base em *deep learning* e inteligência artificial (IA). Para a reconstrução dessa narrativa historiográfica, este trabalho fundamentou-se no conceito de programas de investigação proposto por Swiggers (1981) e grupos de especialidades teóricas de Murray (1994), a fim de estabelecer os programas de investigação da TA. A partir de uma reflexão epistemológica sobre a natureza da tradução do ponto de vista das operações da linguagem (OUSTINOFF, 2003), que inclui a natureza cultural do processo tradutório, chegamos à fundamentação dos métodos de mecanização da linguagem (AUROUX, 1998). Com o objetivo de construir uma interpretação historiográfica sobre os processos de TA para estabelecer uma postura crítica a seu respeito, foram definidos critérios de análise com base em parâmetros internos (incidência da língua na tradução adotada nos modelos de TA) e parâmetros externos (contexto histórico e social em que circulavam as ideias linguísticas que impactaram os programas da TA). A análise crítica busca evidenciar as rupturas teóricas e metodológicas ocorridas em cada programa de investigação da TA, bem como observar suas mudanças do ponto de vista de visão de língua que determinaram o avanço dos modelos de TA.

Palavras-chave: Tradução automática. História da tradução automática. Mecanização da tradução. Historiografia da linguística.

ABSTRACT

Machine translation (MT) is a type of translation, which serves as a resource to facilitate the narrowing of language barriers. The emergence of first experiments with translation machines was in the 1940s in the USA, which had invested on studies in such area for almost two decades, until be impacted by a cut on investments in the late 1960s. The research resumed only in the 1980s and opened space for a new standard on scientific studies in the area. In the following decade, the opening of technical translation market promoted by offering of translation automation tools drove the progress of researches. Since the turn of the millennium, the studies on MT have demonstrated significant advances due to technological progress, which enabled the development of MT tools based on deep learning and artificial intelligence (AI). For reconstruction of such historiographical narrative, this work based on the concept of research programs proposed by Swiggers (1981) and theoretical specialty groups from Murray (1994), in order to establish the MT's research programs. As of an epistemological approach on the nature of translation from the language operations point of view (OUSTINOFF, 2003), which includes the cultural nature of translation process, we arrive at foundation of language mechanization methods (AUROUX, 1998). In order to build a historiographical interpretation of MT processes to establish a critical thinking about them, it was defined analysis criteria based on internal parameters (language incidence in translation adopted by MT models) and external parameters (historical and social context where linguistic ideas that influenced MT programs have moved around). The critical analysis seeks to highlight the theoretical and methodological disruptions that occurred on each MT's research program, as well as look at their changes from language view approach that determined the progress of MT models.

Keywords: Machine translation. History of machine translation. Translation mechanization. Historiography of linguistics.

LISTA DE ILUSTRAÇÕES

Diagrama 1: Tradução intralingual	41
Diagrama 2: Tradução interlingual	42
Diagrama 3: Decodificação e Recodificação.....	43
Diagrama 4: Representação de ATN	50

LISTA DE TABELAS

Tabela 1: Dados históricos	37
Tabela 2: Dados do modelo de base estatística.....	38
Tabela 3: Dados do modelo de base neural e <i>deep learning</i>	38
Tabela 4: Resumo das fontes de dados (corpus).....	39
Tabela 5: Etapas do desenvolvimento da análise	39
Quadro 1: Grupos de pesquisa americanos nas décadas de 1950 e 1960	59
Quadro 2: Grupos de pesquisa fora dos EUA nas décadas de 1950 e 1960.....	60
Quadro 3: Grupos de pesquisa fora dos EUA nas décadas de 1970 e 1980.....	63
Quadro 4: Primeiros sistemas comercializados nos anos de 1980.....	65
Quadro 5: Principais <i>CAT tools</i> comercializadas na década de 1990	68
Quadro 6: Grupos de pesquisa no Japão na década de 1990	70
Quadro 7: Grupos de pesquisa de destaque no Eurotra até a década de 1990	71
Quadro 8: Novos grupos de pesquisa nos EUA na década de 1990.....	72
Quadro 9: Empresas norte americanas de destaque no mercado de TA nos anos 2000	81
Quadro 10: Programas de TA <i>open-source</i> a partir dos anos 2000	81

LISTA DE ABREVIATURAS

ALPAC	<i>Automatic Language Processing Advisory Committee</i>
ATN	<i>Augmented Transition Network</i>
CAT	<i>Computer Assisted Translation</i>
DTS	<i>Descriptive Translation Studies</i>
EA	Enunciado alvo
EF	Enunciado fonte
EBMT	<i>Example-based Machine Translation</i>
FAHQMT	<i>Fully Automatic High Quality Machine Translation</i>
HL	Historiografia da Linguística
IA	Inteligência Artificial
LC	Língua de chegada
LC	Linguística Computacional
LP	Língua de partida
LSP	<i>Language Service Provider</i>
NMT	<i>Neural Machine Translation</i>
PBMT	<i>Phrase-based Machine Translation</i>
PLN	Processamento de Linguagem Natural
RBMT	<i>Rule-based Machine Translation</i>
RNA	Redes Neurais Artificiais
SMT	<i>Statistical Machine Translation</i>
TA	Tradução Automática
TALN	Tratamento Automático das Línguas Naturais

TAUM	<i>Traduction Automatique de l'Université de Montréal</i>
ThTs	<i>Theoretical Translation Studies</i>
TM	<i>Translation Memory</i>
WBMT	<i>Word-based Machine Translation</i>

SUMÁRIO

INTRODUÇÃO	14
CAPÍTULO I: FUNDAMENTOS TEÓRICO-METODOLÓGICOS DA PESQUISA	19
1. A Historiografia da Linguística (HL): o referencial teórico.....	19
1.1 Programas de investigação.....	23
1.2 Grupos de especialidade.....	24
2. Estudos da Tradução: um complemento para o referencial teórico.....	26
3. Tradução Automática: os primeiros anos	32
4. Procedimentos metodológicos da pesquisa.....	34
4.1 Parâmetros de análise	35
4.2. Material de análise: seleção e coleta	36
4.3 Etapas de análise da pesquisa.....	39
CAPÍTULO II: EPISTEMOLOGIA DA TRADUÇÃO.....	40
1. Operações da tradução.....	40
2. Tradução e cultura	45
3. Mecanização da linguagem	48
CAPÍTULO III: PROGRAMAS DE INVESTIGAÇÃO NA HISTÓRIA DA TRADUÇÃO AUTOMÁTICA.....	54
1. Programa de mecanização	56
1.1 Grupos e áreas de pesquisa.....	58
1.2 A retomada do interesse pela TA.....	64
2. Programa de matematização.....	65
2.1 Grupos e áreas de pesquisa.....	70
2.2 As limitações do modelo de base estatística.....	72
3. Programa de operacionalização.....	76
3.1 Grupos e áreas de pesquisa.....	80
3.2 As limitações do modelo de base neural	82
CAPÍTULO IV: ANÁLISE DOS PROGRAMAS DE INVESTIGAÇÃO DA TRADUÇÃO AUTOMÁTICA.....	84
1. Programa de mecanização	85
1.1 Parâmetros internos.....	85
1.2 Parâmetros externos.....	92
2. Programa de matematização.....	96
2.1 Parâmetros internos.....	96
2.2 Parâmetros externos.....	101
3. Programa de operacionalização.....	108
3.1 Parâmetros internos.....	108
3.2 Parâmetros externos.....	112
CONSIDERAÇÕES FINAIS.....	122
GLOSSÁRIO	127
REFERÊNCIAS BIBLIOGRÁFICAS.....	129

INTRODUÇÃO

Língua e tecnologia parecem orbitar em dimensões opostas, mas estão mais conectadas do que nunca em nosso tempo. O avanço tecnológico tem atingido áreas que antes pareceriam pouco prováveis de serem alcançadas e temos presenciado significativo progresso na otimização do uso da linguagem humana por meio da tecnologia. A globalização é, sem dúvida, um dos principais fatores que motivam tal progresso, visto que a comunicação entre nações precisa ser mais eficiente para promover negociações comerciais e políticas, bem como desenvolvimento científico e cultural.

A Tradução Automática (TA) começa ser idealizada em um contexto onde as negociações da paz mundial se tornavam essenciais para o entendimento das nações num cenário pós-Segunda Guerra Mundial (1939-1945) mas, logo, se torna desacreditada por suas limitações. As pesquisas em torno da Inteligência Artificial (IA) avançaram e hoje aliadas a métodos de *Machine Learning* (Aprendizagem de máquina) e *Deep Learning* (Aprendizagem profunda) proporcionam um mundo de possibilidades no desenvolvimento de ferramentas de TA que varrem qualquer dúvida sobre seu potencial de avanço e eficiência.

As ferramentas de TA disponíveis gratuitamente na Internet têm se mostrado cada vez mais assertivas aos olhos dos usuários desse meio, e esse fenômeno não é apenas uma impressão. As tecnologias empregadas atualmente nesse tipo de sistema têm avançado significativamente a ponto de se tornarem capazes de produzir traduções com níveis de fluência e qualidade comparados à tradução humana em determinados tipos de textos.

Para compreender a trajetória de evolução da TA é necessário entender a formação dos primeiros modelos de Processamento de Linguagem Natural (PLN) desenvolvidos para TA. O primeiro deles era formado por algoritmos que faziam a substituição de palavra por palavra (*word-based*) do texto-fonte no texto-alvo. Esse modelo considerava a língua no nível linguístico do léxico, e empregava técnicas de simples substituição de uma palavra por outra no processo de tradução, o que não se apresentou tão eficiente num primeiro momento.

Porém, novos modelos começaram a ser experimentados e o modelo com base em regras (RBMT, *Rule-based Machine translation*), posteriormente, o modelo com base em frase (PBMT, *Phrase-based Machine Translation*), já representaram progresso nos resultados da TA. Mas foi o desenvolvimento do modelo com base em estatísticas (SMT, *Statistical Machine Translation*) que garantiu o salto significativo na qualidade e eficiência da TA, abrindo caminho para o aprimoramento de processos que resultaram no modelo de base neural (NMT, *Neural Machine Translation*).

O desenvolvimento de ferramentas de TA de base neural representa o que temos de mais atual e avançado em tecnologias para TA. Essas ferramentas começaram a ser desenvolvidas apoiadas pela tecnologia de Redes Neurais Artificiais (RNA), significando uma evolução na maneira como a máquina passa ser treinada e, conseqüentemente, “aprende” a traduzir melhor. A melhoria na qualidade e fluência do texto traduzido é inquestionável, na medida em que abre possibilidades de integração com outros recursos, como a integração com sistemas de reconhecimento de voz, gerando ferramentas para interpretação simultânea.

Nesse contexto de desenvolvimento de tecnologias para TA, esta dissertação de mestrado tem como **tema** a compreensão de aspectos históricos que desenharam a evolução de processos da TA até chegarmos a plataformas atuais de base neural, que refletem o que existe de mais moderno em métodos de automatização da linguagem humana.

Sendo assim, o questionamento que motiva a realização desta dissertação é direcionado nas seguintes **problematizações**: *qual a natureza essencial da evolução histórica das ferramentas tecnológicas de tradução automática e quais são seus impactos na sociedade?*

Partindo dessas perguntas de pesquisa, a dissertação tem como **objetivo geral** desenvolver uma reflexão sobre a contextualização histórica dos processos de TA e, associada a essa tarefa, estabelecer uma discussão para compreender a natureza qualitativa desses mesmos processos a fim de que se estabeleça uma postura crítica a respeito da TA.

Para atingir resultados interpretativos que busquem responder a esse objetivo geral, estabelecem-se os seguintes objetivos específicos de trabalho: a) visitar criticamente episódios da história da tradução, com destaque para o desenvolvimento de teorias, métodos, ferramentas; b) analisar as razões da emergência de mecanismos de tradução automática na história da tradução; c) interpretar

historiograficamente mudanças nos modelos de tradução automática; d) relacionar essas mudanças a aspectos históricos e sociais que possam servir de apoio a interpretações sobre desenvolvimentos históricos nas práticas de tradução; e) elaborar uma síntese interpretativa historiográfica sobre a presença das tecnologias no contexto da história da tradução; f) elaborar uma reflexão sobre o impacto social do uso das ferramentas de tradução na sociedade.

Dada a importante transformação e o avanço tecnológico nos últimos anos nesta área, aprofundar a investigação a respeito da TA motiva a realização deste trabalho a fim de entender os processos que trouxeram a TA ao patamar que hoje ocupa na vida de nós tradutores e dos usuários de maneira geral, bem como possibilita refletir sobre as possíveis trajetórias que ela ainda poderá tomar.

Entende-se que essa motivação aponta para uma **justificativa** da relevância da realização deste trabalho, que se evidencia também pela ausência de uma retomada histórica, a partir de teorias e métodos da Historiografia da Linguística, em termos avaliativos e interpretativos da presença de ferramentas tecnológicas nas práticas da tradução.

Gigantes multinacionais como Google, Microsoft e Facebook investem há anos no desenvolvimento de novas tecnologias para automatização da tradução a fim de melhorar a qualidade da TA disponível em suas plataformas, diminuir o tempo e custo da tradução, assim como atrair mais usuários para suas ofertas de serviços por meio dos recursos de TA disponíveis. Esse investimento também é forte no mercado de tradução técnica e localização de *softwares* que busca constantemente por soluções para reduzir custo e tempo na tradução dos imensos volumes de palavras que transitam nessa indústria.

Desde seu início a TA foi pensada como solução para a necessidade de atender uma demanda por tradução que ultrapassava os limites da capacidade que tradutores ou profissionais que trabalhavam com idiomas poderiam suprir, demanda essa tanto nos estudos científicos como nas relações políticas e que, sem qualquer dúvida, não conseguiria ser atendida sem a TA. Agilidade e redução de custos são as principais razões que levam gigantes multinacionais a continuarem investindo em tecnologias para melhorar a TA. A conectividade global que rompe as barreiras de comunicação entre países impulsiona a necessidade por rapidez no que diz respeito à tradução entre idiomas em tempo real.

Os estudos e o desenvolvimento em larga escala de sistemas de TA começam oficialmente na década de 1980, porém registros mostram os primeiros passos em direção à automatização da tradução ainda entre os anos de 1947 e 1949, por cientistas do departamento de matemática do *Rockefeller Foundation* em Nova Iorque e do Instituto de Tecnologia de Massachusetts (MIT).

No Brasil, os estudos acadêmicos em torno da TA ainda são poucos, especialmente na área de humanidades, e podemos dizer que ainda é necessário despertar o interesse em aprofundar o tema dentro da comunidade científica brasileira. Os estudos sobre TA trata-se de um campo interdisciplinar que pode despertar interesse de diferentes áreas de conhecimento.

Os Estudos da Tradução e a Linguística, tradicionalmente, estudam a TA a partir da perspectiva dos resultados da tradução, enquanto que disciplinas das Ciências Exatas investigam o desenvolvimento de ferramentas e tecnologias para automatização da linguagem. Aliás, o tema da automatização da linguagem é de interesse da Ciência da Computação que hoje abarca disciplinas como a Linguística Computacional (LC) e o Processamento de Linguagem Natural (PLN), compreendendo também um campo interdisciplinar que estuda os processos computacionais de automatização da linguagem humana.

Assim, é necessário levarmos em conta a característica transdisciplinar dos estudos da TA, ou seja, sua abrangência em diferentes áreas de conhecimento tanto dentro das Ciências Humanas como das Exatas. Para a reconstrução de uma história da TA, neste trabalho pretendemos desenvolver uma discussão em relação à inscrição social e temporal da mudança tecnológica ocorrida nas tecnologias de TA desde a década de 1980, em especial, com a recente chegada de plataformas de TA em NMT.

Nesse complexo cenário, de modo a dar conta de sua proposta e reafirmar a justificativa da sua realização, esta dissertação está organizada de modo a conduzir o leitor numa reconstrução da história da TA a partir da visão de uma tradutora, começando desde sua criação no final dos anos de 1940 e seguindo pela cronologia de sua evolução, para alcançar uma análise crítica sobre seu processo evolutivo e possível impacto social até o momento.

Após uma Introdução em que são expostos problematização e circunscrição do tema, objetivos e justificativas, a dissertação apresenta a seguinte estrutura: a) no Capítulo I, indicam-se os pressupostos teóricos que orientam a retomada histórica que

se propõe a realizar nesta dissertação, assim como os procedimentos metodológicos da pesquisa são apresentados; c) no Capítulo II, apresentam-se os fundamentos epistemológicos da tradução e sua natureza cultural numa perspectiva de humanização da tradução; d) no Capítulo III, propõe-se, a partir dos métodos explicitados no Capítulo I, os programas de investigação da tradução automática; e) no Capítulo IV, realiza-se, pela perspectiva historiográfica, a análise crítica da evolução tecnológica da tradução automática a partir de seus programas de investigação definidos no Capítulo III; f) a Conclusão procura responder em forma de síntese interpretativa aos problemas de pesquisa destacados nesta Introdução. Ainda fazem parte do trabalho, um glossário para situar os leitores em algumas expressões e termos e as referências bibliográficas.

CAPÍTULO I

FUNDAMENTOS TEÓRICO-METODOLÓGICOS DA PESQUISA

1. A *Historiografia da Linguística (HL): o referencial teórico*

O termo História em sua definição refere-se ao estudo de fatos passados a partir de um recorte do ponto de vista de um indivíduo sobre um determinado evento, paralelamente, encontramos o termo Historiografia, que se caracteriza pelo estudo analítico, ou seja, pela interpretação desses fatos. A diferenciação entre os termos nem sempre é consenso, porém neste estudo trataremos como termos diferentes devido a conceituação de Historiografia da Linguística¹ (HL).

A HL define-se como o estudo analítico da história do pensamento sobre a linguagem e as línguas e toma como objeto de estudo as ideias linguísticas, circunscritas em diferentes recortes temporais que definem o ambiente histórico, social e cultural de circulação do conhecimento (cf. BATISTA, 2013).

As reflexões sobre a HL, neste trabalho, tem o caráter de apresentar e demonstrar sua aplicabilidade na elaboração de uma história da TA. Sendo assim, serão adotados pressupostos e métodos da HL a fim de orientar o desenvolvimento de uma narrativa interpretativa para uma análise crítica da reconstrução do surgimento e da evolução das tecnologias de TA, em especial a NMT.

Esse recorte científico, a HL, consolida-se a partir da década de 1970 com a criação da primeira revista científica especializada - *Historiographia Linguistica* – desenvolvida por Konrad Koerner e, posteriormente, também por trabalhos publicados por vários pesquisadores, entre eles Pierre Swiggers, um dos historiógrafos mais preocupados com questões teórico-metodológicas da área.

A historiografia linguística é o estudo interdisciplinar do curso evolutivo do conhecimento linguístico; ela engloba a descrição e a explicação, em termos de fatores intradisciplinares e extradisciplinares [...], de como o conhecimento linguístico, ou mais genericamente, o *know-how* linguístico foi obtido e implementado. (SWIGGERS, 2010 *apud* BATISTA, 2013, p. 48)

¹ Entende-se *linguística* em sentido amplo e abrangente, levando em conta todo e qualquer estudo sobre línguas e linguagem em qualquer recorte temporal e com diferentes objetivos. O termo não se limita à extensão ao sentido de ciência da linguagem, como organizada a partir do século XX.

A HL torna-se disciplina formal como área de conhecimento dentro das ciências humanas entre as décadas de 1960-70, e no Brasil a recepção e divulgação dos primeiros estudos sobre HL se deu por meio dos trabalhos de pesquisa divulgados por Cristina Altman, que introduziu os pensamentos e trabalhos de Koerner e Swiggers à comunidade acadêmica brasileira.

A denominação do campo - Historiografia da Linguística - também passa por discussões, por encontrar relações com outras formas de nomeação, por exemplo, História das Ideias Linguísticas, História da Linguística, História dos Estudos sobre a Linguagem e História do Conhecimento Linguístico.

Nesta dissertação, adotaremos o uso da expressão “Historiografia da Linguística”, por ser denominação empregada por Konrad Koerner, Pierre Swiggers e Cristina Altman na literatura que serve de embasamento para este trabalho e que, segundo Altman, ocorre com mais frequência na literatura internacional sobre o estatuto da disciplina que “visa [...] à reconstrução da história, ou de parte da história, da teoria e da prática gramatical e lexicográfica, da política linguística e das atitudes em relação à língua” (ALTMAN, 2009 apud BATISTA, 2013, p.17).

A HL caracteriza-se também por sua vocação interdisciplinar que engloba fundamentos da História, Linguística e Filosofia, o que lhe atribui a possibilidade de diferentes vieses para análise dos acontecimentos em um determinado recorte histórico. A HL se propõe a ser uma historiografia problematizadora, ou seja, uma história de problemas, na qual procura-se observar no passado fatos históricos para que “se possam dar os devidos pesos e medidas para os objetos de investigação” (BATISTA, 2013, p. 55).

Nem a língua e nem a linguagem são objetos de estudo da HL, mas sim o conhecimento produzido sobre a linguagem e as línguas (restrito ao campo científico da linguística ou parte de outros campos e formas de conhecimento, mesmo que não científicos e circunscritos a diversos recortes temporais). Como resultado, a HL é observada como uma metaciência (ou metadisciplina), pois seu objeto de estudo remete à reflexão sobre a própria ciência da linguagem (ou sobre as formas de conhecimento sobre línguas e linguagem) e é de caráter histórico, pois se preocupa com a elaboração, circulação e recepção das ideias linguísticas em períodos delimitados pelo pesquisador.

Hoje em dia, o que é normalmente referido como ‘historiografia linguística’ (forma abreviada da designação mais precisa ‘historiografia das ciências da linguagem’, em que o termo ‘ciência’ no sentido estrito da ‘ciência natural’ é claramente evitado) constitui uma investigação metodologicamente informada e a apresentação de acontecimentos passados na evolução da disciplina designada de ‘linguística’ ou ‘ciências da linguagem’. Se a atividade de estabelecer as *res gestae* do estudo da linguagem é chamada de ‘historiografia linguística’, o seu resultado deveria ser designado de ‘história da linguística’. Por outras palavras, a ‘história da linguística’ é o produto e não a atividade de a estabelecer. A permanente discussão em torno de questões relativas ao método, à filosofia e à epistemologia é chamada de ‘meta-historiografia’. (KOERNER, 2014, p. 17²)

O pesquisador da HL é o historiógrafo da linguística. O ponto de vista desse historiógrafo que reconstrói eventos históricos está atrelado a um recorte temporal selecionado, assim, trata-se de uma história seletiva e de objetividade relativa. O historiógrafo faz o recorte, a seleção e a hierarquização dos fatos de acordo com sua observação e interesses interpretativos, sem que com isso se entenda a produção de uma historiografia enviesada ou a serviço da propaganda de determinadas ideias (cf. KOERNER, 2014 para uma longa discussão sobre os tipos de historiografia). O historiógrafo trabalha com evidências (seus documentos históricos) associadas a fatores externos, buscando explicar por que o fato que analisa se tornou relevante no seu recorte temporal, justificando desse modo suas opções seletivas ao elaborar suas interpretações historiográficas.

O trabalho do historiógrafo é analisar e explicar o que foi dito sobre línguas e linguagem. Essa atividade de natureza histórica é realizada por meio da hierarquização de dados, tendo em vista compreender teorias e métodos, filosofias e modos de ensino, produção e circulação de textos, entre outras práticas de interesse para a HL. Isso significa que uma pesquisa em HL deve incluir a busca por fontes diversificadas com o objetivo de encontrar evidências para uma problematização, bem como para determinar o método de análise que será empregado.

A reconstrução da narrativa histórica na HL é conduzida pela perspectiva da visão do historiógrafo encarregado de refletir sobre a “escrita histórica dos estudos da linguagem” e não “apenas descrever um registro histórico da pesquisa sobre linguística” (cf. Koerner, 2014) que, tradicionalmente, escreve uma narrativa de uma história de fatos. A história da linguística precisa de sua parcela relacionada ao registro

² A publicação de 2014 é uma coletânea de textos escritos por Konrad Koerner ao longo de sua carreira, com traduções em português de artigos e capítulos que vão desde 1970 até os anos 2000.

histórico que deve continuar sendo produzida, o que queremos ressaltar aqui é que esse método de descrever História não é o objetivo nem a função da HL.

A percepção crítica da narrativa historiográfica está ancorada à definição de parâmetros internos e externos de análise. Parâmetros internos referem-se à visão geral do objeto, à incidência de análise de abordagens linguísticas (qual(is) unidades(s) linguística(s) é(são) privilegiada(s), a técnicas e procedimentos analíticos; dessa forma, estão relacionados a aspectos dos programas de investigação que delimitam formas de estudo sobre línguas e linguagem. Parâmetros externos referem-se ao contexto histórico e social em que se desenvolvem e circulam ideias linguísticas; assim, a periodização, o clima intelectual e as circunstâncias de institucionalização, bem como, a divulgação dos saberes e da motivação do autor resultam do trabalho dos grupos de especialidades.

O resultado do trabalho de agentes produtores e receptores de ideias linguísticas – reunidos em grupos de especialidades – define eixos de continuidade ou descontinuidade com outras ideias linguísticas anteriores ou mesmo concomitantes a sua elaboração e divulgação. Descontinuidades nas ideias implicam características disruptivas que rompem com movimentos tidos como dominantes em determinada época (sem a conotação de exclusividade), e que se chocam com novos pensamentos críticos, estabelecendo-se assim novas propostas para os estudos sobre a linguagem e as línguas.

Esses fatos disruptivos também fazem parte do objeto de estudo do historiógrafo que busca na reconstrução historiográfica evidências que justifiquem a importância factual de um evento. A história da linguística é permeada de descontinuidades para anunciar novas propostas, mas é função do historiográfico investigar, de maneira cuidadosa, as descontinuidades propostas a fim de buscar fatos que as justifiquem. Koerner (2014) vê a produção de ideias linguísticas como processos de continuidades e descontinuidades ao longo da história da linguística e se opõe ao método de uma história da linguística linear, ou seja, em que uma nova proposta teórica invalidaria e eliminaria teorias anteriores, apenas pelo argumento de serem novas propostas substituindo ideia e padrões antigos.

De maneira oposta também há continuidades históricas, que estão presentes quando observamos ideias que se legitimaram dentro de um campo de conhecimento, por um grupo de especialidade, em que se criou uma tradição de pensamento.

Continuidades e discontinuidades são movimentos de um percurso dinâmico, e o destaque a um ou outro é resultado da escolha interpretativa do historiógrafo.

Na sequência, apresentamos os dois conceitos que serão privilegiadamente empregados na interpretação historiográfica proposta nesta dissertação: programas de investigação e grupos de especialidade.

1.1 Programas de investigação

O conceito de programas de investigação (*research programs*) é proposto por Pierre Swiggers em “*The history writing of Linguistics: A methodological note*” de 1981. Swiggers propõe descrever teorias, filosofias, métodos de estudo da linguagem por meio de programas de investigação que possibilitariam agrupar ideias que compartilhassem do mesmo conceito sobre o objeto de estudo e de mesmos procedimentos metodológicos de análise ao longo do desenvolvimento histórico.

A proposta de Swiggers tem como objetivo agrupar em um conjunto de quatro programas (não necessariamente excludentes entre si) modos diferentes de estudo sobre a linguagem e as línguas. A definição dos programas de acordo com sua área de abrangência é configurada como:

1) programa de correspondência: parte do conceito de que língua é um meio para expressar ideias e pensamentos, assim sua principal característica inclui estudos de pesquisadores que se propõem a investigar a correspondência entre linguagem e pensamento, bem como a segmentação das unidades linguísticas. Entre os linguistas que podemos destacar dentro desse programa está Noam Chomsky.

2) programa descritivista: engloba estudos relacionados aos fatos formais da língua, sobre as formas e estruturas linguísticas. Entre os linguistas que podemos destacar nesse programa está Leonard Bloomfield.

3) programa sociocultural: engloba estudos que relacionam a língua em seu contexto sociocultural, com o objetivo de analisar a variação das formas linguísticas dentro de uma comunidade linguística e no discurso do falante. Assim, caracteriza-se por uma perspectiva interdisciplinar de análise linguística. Entre os linguistas que podemos destacar nesse programa está William Labov.

4) programa de projeção: engloba estudos relacionados à investigação de uma língua natural a partir de sua projeção de um sistema formal. O objetivo de se adequar dentro de um modelo abstrato formal pode incluir fórmulas lógico-matemáticas para

descrever um padrão linguístico. Entre os pesquisadores desse programa estão Jaakko Hintikka e Šaumjan Sebastian Kostantinovič.

Os programas têm como objetivo organizar os diferentes tratamentos dos objetos teóricos da linguística e funcionam como ferramenta de classificação, não normativa (mas de valor inicial descritivo), para que o historiógrafo da linguística possa desenvolver uma reflexão crítica dentro da variedade de modelos e teorias propostas no decorrer do percurso da história da linguística. Dessa forma, seria possível reconhecer, classificar e selecionar “similaridades, sobreposições, abordagens recorrentes e, entre outros, divergências fundamentais de visões e estratégicas” (SWIGGERS, 2017, p. 80) para desenvolver uma análise teórico-metodológica sólida sobre os estudos da linguagem e das línguas.

A partir da definição de programas de investigação proposta por Swiggers (1981), faremos uma interpretação e adotaremos uma nova nomeação específica para os programas de investigação da TA neste trabalho, que servirá de método para o desenvolvimento da análise historiográfica da evolução da TA. A questão a ser delineada mais adiante envolve a descrição detalhada dos programas da TA, a visão de linguagem e língua subjacente, os procedimentos e técnicas de análise adotados.

1.2 Grupos de especialidade

O comportamento de cientistas começa ser tema e interesse de estudo de sociólogos da ciência, formalmente, a partir da década de 1970. Trabalhos como os de Merton (1973) e Kuhn (1962) iniciam uma nova tradição de pesquisa em história da ciência e sociologia da ciência, segundo Stephen Murray em *Theory Groups and the Study of Language in North America* (1994), que define o processo de formação de grupos de especialidade teórica como necessário para compreender etapas da consolidação de mudanças científicas ou de continuidades teórico-metodológicas que estabelecem tradições de pesquisa.

O modelo de Kuhn em *The Structure of Scientific Revolutions* de 1962, analisado por Murray (1994, p. 3), destaca que a legitimação de uma teoria dá-se por meio de um processo social e não pela simples difusão automática de ideias. Nesse processo, uma mudança científica ocorre na “ciência normal”, como o acúmulo contínuo de conhecimento, quando ocorre uma “ciência revolucionária”, com a descontinuidade de ideias; o que provocaria questionamentos sobre as bases de uma disciplina. Esses questionamentos resultam em períodos de crises em que bases

teórico-metodológicas legitimadas como tais não seriam mais validadas como prática científica viável.

Como resultado das crises e conflitos, teóricos e científicos, emerge a necessidade da institucionalização, ou seja, de uma organização para legitimar a teoria/ideia/pensamento e promover sua aceitação social. A institucionalização, segundo Murray (1994), não provaria a veracidade da teoria, porém possibilitaria sua consolidação, elaboração e difusão por meio de um líder organizacional e/ou de um líder intelectual.

Dentro desse modelo, identificam-se as elites universitárias e instituições que seriam grupos responsáveis pela legitimação e difusão de uma teoria no ambiente acadêmico. A formação desses grupos, Murray (1994) propõe, passa por estágios que se iniciam com pesquisadores trabalhando, praticamente, de maneira isolada e que vão gradativamente se agrupando e compartilhando as mesmas ideias e pensamentos, até atingirem os estágios de especialidade acadêmica, em que se reúnem numa organização formal e, posteriormente, se reúnem em uma comunidade científica para trabalhar ativamente na solução de problemas de uma disciplina.

A noção de grupos de especialidade teórica contempla três pré-requisitos mínimos para formação de uma comunidade de pesquisadores: 1) boas ideias; 2) liderança intelectual; 3) liderança organizacional.

Murray (1994) explica que não se forma um grupo de especialidade teórica sem liderança e organização de elementos intelectuais essenciais. Ele destaca funções da liderança intelectual, como definição da base conceitual da linha de pesquisa, explicitação das ideias propostas na pesquisa, validação de trabalhos de outros pesquisadores, elaboração do estatuto do programa e construção de pesquisas modelo como etapas para a legitimação de um grupo:

(1) definição de uma base conceitual para uma linha de pesquisa, (2) explicação das implicações da pesquisa em relação às “boas ideias”, (3) aprovação e validação do trabalho de outros pesquisadores como competente, relevante e/ou como parte da estrutura da pesquisa. [...] 4) elaboração do estatuto do programa, especificando qual pesquisa deve ser conduzida e como ela se adequa à teoria principal, [...] 5) construção de pesquisas modelo, mostrando como a pesquisa deve ser conduzida. (MURRAY, 1994, p. 22, tradução nossa)

A liderança organizacional do grupo será responsável pelas tarefas administrativas, como organização do tempo, busca por fundos de apoio e meios que

facilitem a pesquisa, incluindo atividades de comunicação e divulgação dos resultados da pesquisa validados pela organização intelectual. O processo assemelha-se à organização institucional de uma empresa com suas devidas atividades específicas a cada especialidade de pesquisa.

Segundo Batista (2013, p. 87), o conceito de grupos de especialidade é uma ferramenta para análise dos aspectos externos do objeto de análise e serve para investigar os estágios pelos quais grupos passam em busca de sua legitimidade e atuação em determinada área de pesquisa.

Assim, esta pesquisa buscará compreender a formação dos grupos de especialidade que possibilitaram o desenvolvimento de sistemas de TA, bem como o avanço de seus programas de investigação, a fim de reconstruir a história da TA desde seu surgimento (na década de 1940), chegando a um panorama dos dias atuais em relação às tecnologias atualmente adotadas.

A construção de uma narrativa historiográfica com base na HL se dá pela possibilidade de selecionar recortes históricos para compreender a formação de grupos de especialidade em torno de propostas teóricas e técnicas, em um certo momento, e o entendimento dos programas de investigação, junto aos grupos de especialidades. Essa abordagem mostrará o resultado da legitimação das propostas teóricas e técnicas na comunidade acadêmica e profissional em que o pesquisador associa sua investigação e ideias que circulam e são recebidas em determinado contexto histórico e social.

A seguir, apresentamos também outros elementos teóricos que servirão de apoio aos fundamentos e métodos da HL para o desenvolvimento da proposta de pesquisa desta dissertação.

2. Estudos da Tradução: um complemento para o referencial teórico

A prática da tradução é intrínseca à prática comunicativa, é parte integrante da linguagem que depende de processos para transpor sentido entre idiomas e promover o entendimento entre sistemas linguísticos diferentes. A prática tradutória acompanha a humanidade e torna-se interesse de estudo em diferentes momentos, com diferentes fases de importância e entendimento, assim como o desenvolvimento de teorias ao longo de sua história.

O autor George Steiner em sua publicação *After Babel* (1975) propõe a divisão de quatro períodos da história da tradução, que são definidos pelo surgimento de novas linhas de pesquisa relacionadas aos estudos da tradução, e são eles:

- 1) trabalhos desde Cícero (106-43 a.C.) e Horácio (65-8 a.C.), na Roma antiga, até a publicação de Alexander Fraser Tyler “*Essay on the Principles of Translation*” (1791). Nesse período a principal característica era a prática da tradução como um trabalho empírico, com base na simples experiência do ato de traduzir;
- 2) a partir da publicação de Valery Larbaud “*Sous l’invocation de Saint Jérôme*” (1946) onde presenciamos a evolução de metodologias para a prática da tradução voltadas à questão da teoria e hermenêutica, associada ao desenvolvimento de um vocabulário para os estudos da tradução;
- 3) a partir da década de 1940 (sobrepõe-se ao segundo período) quando começaram as primeiras discussões sobre a mecanização da tradução e com elas pensamentos para introdução de uma teoria da comunicação e da linguística estrutural nos estudos da tradução;
- 4) a partir da década de 1960 quando se introduz uma visão metafísica aos estudos da tradução e interpretação. Nesse período passa a se reconhecer a disciplina numa abordagem interdisciplinar dentro dos estudos da linguagem.

A prática da tradução e o papel do tradutor passa, ao longo da história da tradução, por diferentes fases de entendimento e importância. Segundo Milton (2010, p. 25-26), podemos observar períodos em que a tradução, bem como o papel do tradutor é visto como inferior e sem valor e, em contrapartida, encontramos períodos em que o tradutor é reconhecido por refinar o original.

Durante o período de 1550 a 1650 as imagens predominantes são as relativas a se seguirem os passos exatos do autor, ao tradutor como servo ou escravo, e o trabalho do tradutor como sendo infinitamente inferior ao original – o avesso de uma tapeçaria, ou a luz da vela comparada à luz do sol. [...] Agora, encontramos imagens que retratam o tradutor preservando a “chama” do original, acrescentando algo de si para preservar a “essência” do original. O tradutor deixa de ser um servo; ele agora é um amigo e conselheiro, o qual pode ter acentuada afinidade com o autor. (MILTON, 2010, p. 25-26)

Os conceitos da prática da tradução passam ainda pela discussão sobre qual “linha de abordagem” (BASSNETT, 2002, p. 50) o tradutor teria seguido no processo de tradução. O conceito renascentista ou clássico da abordagem “palavra por palavra”

ou “sentido por sentido” depende da escolha do tradutor. Essa escolha será sempre pessoal, pautada em suas convicções, experiências e no mercado em que atua.

A linha “sentido por sentido” é observada nos clássicos de Horácio e Cícero que buscavam transferir o sentido do texto de partida na língua de chegada. Outro exemplo, seria a tradução de São Jerônimo da Bíblia do hebreu em 390 d.C., a primeira tradução da Bíblia, que também tratava-se de uma tradução na linha do sentido.

Apesar do breve relato histórico, uma vez que não caberia uma abordagem cronológica mais extensa em virtude do objetivo que buscamos neste trabalho, podemos observar a importância e presença dos estudos da tradução desde a Roma antiga. Assim, um longo período foi transcorrido até tornar-se reconhecida como uma disciplina institucionalizada em 1980, consolidando-se como uma disciplina própria, definitivamente, apenas na década de 1990, frente ao avanço tecnológico dos meios de comunicação que proporcionou sólida expansão da globalização.

A partir do reconhecimento dos estudos da tradução como disciplina, foram implantados no Brasil cursos técnicos e de nível superior, seguindo uma tendência internacional. A demanda por traduções especializadas crescente impulsionou o mercado de tradução mundial que motivou a comercialização de ferramentas para auxiliar no trabalho de tradução, com o objetivo de trazer agilidade ao processo e diminuir custos. Discutiremos o período de abertura e expansão do mercado de tradução no capítulo III.

Desde o início da década de 1960 ocorreram mudanças significativas no campo dos estudos da tradução motivadas, em especial, por novos estudos sobre linguagem desenvolvidos por pensadores do Círculo de Bakhtin e linguistas como R. Jakobson e M. Halliday (este último um pouco mais recente, no início da década de 1980), que contribuíram para o estabelecimento de novos critérios na fundamentação de uma teoria da tradução.

Conforme Bassnett (2002, p.12), o estabelecimento dos estudos da tradução como disciplina formal foi fundamentado no trabalho de André Lefevere intitulado “*Louvain Colloquium on Literature and Translation*” de 1976 que propunha a adoção da disciplina com o nome de “Estudos da Tradução”, com o objetivo de estudar “os problemas levantados pela produção e descrição de traduções” e que fosse uma disciplina independente, resultando hoje numa área que “não é um espelho dos

estudos literários, nem uma área específica da linguística, mas sim, um vasto campo complexo com inúmeras ramificações”.

A fundamentação da disciplina representa uma relação importante com a prática, isto é, o exercício prático da tradução tornou-se a base para investigação da teoria. A tradução literária, tradicionalmente, era o principal material de estudo e desenvolvimento da teoria da prática, uma vez que a produção da tradução é o objeto de estudo da disciplina.

A necessidade de sistematizar os Estudos da Tradução seguiu até o final da década de 1980 seguindo o modelo sistêmico de I. Even-Zohar e G. Toury denominado Teoria dos Polisistemas (*Polysystems Theory*). O modelo enfatizava o estudo no sistema de chegada, em que a cultura de chegada seria o objeto de estudo da disciplina. Nesse modelo, era defendida a necessidade do estabelecimento de padrões de regularidade do comportamento tradutório e não aceitava-se “a ideia de que o objeto da teoria da tradução servisse para melhorar a qualidade da tradução” (BASSNETT, 2002, p. 7).

Toury afirma que uma vez que a tradução tem o objetivo básico de atender a uma necessidade na cultura de chegada, é lógico tornar o sistema de chegada como o objeto de estudo. [...] O centro da Teoria dos polisistemas era a ênfase no estudo da poesia da cultura de chegada. Sugeriu-se que seria possível prever as condições pelas quais cada tradução deveria acontecer e presumir quais tipos de estratégias os tradutores poderiam empregar. (BASSNETT, 2002, p. 7, tradução nossa)

A Teoria dos Polisistemas preencheu uma lacuna que existia entre os estudos da linguística e da literatura, uma vez que proporcionou a estrutura interdisciplinar que seria construída dentro dos estudos da tradução. Nesse modelo teórico, a ênfase estava no estudo da poesia da cultura de chegada, focalizando também na tradução literária, mas incluindo outras modalidades como a dublagem, legendagem, tradução de literatura infantil e até de textos de cultura popular e propaganda.

De acordo com Bassnett (2002, p. 16), o postulado definido por Lefevere como objetivo da disciplina dizia ser necessário “estabelecer uma teoria abrangente que pudesse também servir como guia para a produção de traduções” que continuou passando por discussões a fim de atingir novos patamares de abrangência interdisciplinar, resultando em quatro grandes áreas dentro de seu campo de pesquisa

que são: 1) história da tradução; 2) tradução na cultura da língua de chegada; 3) tradução e linguística; 4) tradução e poesia.

A história da tradução, que faz parte da história da literatura, incorpora em seus estudos as teorias da tradução em diferentes épocas, a resposta crítica a traduções, os processos de produção e publicação de traduções, o papel e a função da tradução em um determinado período, o desenvolvimento metodológico da tradução e a análise do trabalho individual de tradutores.

A tradução na cultura da língua de chegada é a área que investiga a tradução além do texto e do autor. Ela considera as influências que abrangem um texto, um autor ou um gênero, bem como a assimilação de termos do texto traduzido na língua de chegada, no sistema e nos princípios de operação de seleção dentro desse sistema.

Em tradução e linguística estuda-se a organização dos elementos linguísticos em comparação à língua de partida e chegada, nos níveis da fonética e morfologia, do léxico e da sintaxe. Nesse campo também são investigados problemas de equivalência linguística, de equivalência entre idiomas e de TA, como também textos não literários.

Em tradução e poesia estão dedicados os estudos da teoria e prática da tradução literária que investiga os problemas na tradução de poesia, textos para teatro e cinema, dublagem e legendagem. A área aborda ainda os estudos dos problemas relacionados à criação de textos poéticos e da inter-relação entre os textos da língua de partida e chegada, assim como do autor com o tradutor.

As teorias da tradução começaram ser cada vez mais estudadas a partir da formalização da disciplina, incluindo também estudos sobre o papel do tradutor, a importância da ética na tradução e os problemas filosóficos na prática da tradução.

Propostas de estudos da disciplina foram surgindo, como o trabalho de James Holmes, "*The Name and Nature of Translation Studies*" de 1972, que representou um grande passo na definição da disciplina que apresentava as teorias da tradução em duas grandes áreas: 1) os estudos descritivos da tradução (*DTS, Descriptive Translation Studies*) e 2) os estudos teóricos da tradução (*ThTs, Theoretical Translation Studies*).

Com base no conceito dos estudos descritos, Pym (1998) propôs estudar a história da tradução a partir de métodos específicos desenvolvidos para a disciplina, numa abordagem de humanização, da busca dos agentes e dos efeitos das traduções

no contexto social e cultural. A proposta incluía três áreas de estudos que compunham o método de pesquisa: 1) a arqueologia da tradução; 2) o criticismo histórico e 3) a explicação (causa vs. efeito de uma tradução).

O objetivo da proposta era atualizar a maneira de se fazer historiografia da tradução, visto que a tradição tentava comprovar a teoria na prática e não investigava a cena em que o processo de tradução foi idealizado, por isso o método incluía a análise de seus agentes. Não é uma questão de negar a importância da teoria mas defender que a reconstrução de uma história da tradução precisa de “um método, com seus próprios conceitos, procedimentos e resultados” (PYM, 1998, p.13).

Para uma abordagem de humanização da história da tradução nos estudos descritivos, Pym (2009, p. 35) sugere primeiro estudar o tradutor para depois estudar o texto. Investigar questões que podem envolver a decisão do tradutor em suas escolhas, engloba desde suas configurações discursivas no momento de produção da tradução até “tensões e negociações que ultrapassam as fronteiras sociais e culturais”.

O segundo fator de importância na abordagem de humanização dos estudos da história da tradução é buscar entender as interculturais profissionais, isto é, a maneira como os profissionais se agrupam para trabalhar num contexto de comunicação *cross-cultural* (entre culturas). Pym (2009, p. 42) procura demonstrar que a intersecção cultural não é homogênea e que não se pode simplificar que tradutores serão sempre mediadores da cultura do texto de partida ou fiéis a cultura do texto de chegada e vice versa. É necessário um mapeamento complexo a fim de identificar as redes de conexão, com o objetivo de alcançar uma abordagem que questione o modelo tradicional de fidelidade ao texto de partida e as abordagens sistêmicas com base no texto de chegada.

Existe uma diferença envolvida, uma diferença fundamental, quando vamos além do foco estreito da língua nos textos. De um lado, um determinado conjunto de questões de pesquisa focalizam-se na reconstrução de perfis sociais, transações financeiras, datas e movimentações, bem como os detalhes arqueológicos que podem unir uma situação social ao invés de apenas assumir um sistema inteiro. [...] Do outro, encontramos tentativas de privilegiar os tradutores como uma voz dentro das traduções e mantê-la aí. (PYM, 2009, p. 37, tradução nossa)

[...] No entanto, as questões que estão sendo feitas, assim como os métodos empregados, parecem revelar algo oculto (a subjetividade do

tradutor é de fato com frequência suprimida) que parece bem óbvio nos discursos extra tradutórios dos tradutores históricos. Parece haver poucas razões do por que a História da Tradução deve estabelecer um criticismo textual complicado, enquanto que poderia alcançar muitos de seus objetivos mais diretamente ao formular perguntas biográficas ou sociológicas, ou olhar de maneira crítica na linguagem dos prefácios, correspondências e nos textos relacionados, ao invés das traduções. (PYM, 2009, p. 37, tradução nossa)

Nos estudos contemporâneos da tradução temos também Lawrence Venuti (2012) que propõe uma abordagem teórica em que divide os modelos de tradução em instrumental e hermenêutica. No modelo instrumental trata-se a tradução como reprodução ou transferência do texto-fonte com base no fator da invariabilidade, enquanto que na hermenêutica trabalha-se com a tradução numa abordagem de determinantes linguísticos e culturais, ou seja, a tradução passa a ser uma interpretação do texto-fonte, com o fator da variabilidade sendo uma de suas características.

Encontramos na história da tradução descontinuidades de teorias e métodos de análise, assim como veremos na história da TA. Essas descontinuidades estão relacionadas às características de formação e desenvolvimento dos programas de investigação da TA que serão apresentados no capítulo III.

3. Tradução Automática: os primeiros anos

Tradução Automática (TA) é uma modalidade de tradução que consiste na automatização de uma parte ou do todo no processo de tradução de uma língua natural para outra. O termo tem origem no inglês “*Machine Translation*” em razão dos primeiros modelos idealizados no território americano terem sido planejados para serem máquinas dedicadas exclusivamente à tradução, no entanto, a modalidade tornou-se popularmente conhecida posteriormente em outros países como “Tradução Automática” pela sua funcionalidade de automatização do processo de tradução. O termo MT (*Machine Translation*) é ainda o mais usado na área, especialmente fora do Brasil, mas adotamos o termo TA nesta dissertação com o objetivo de consolidação de uma terminologia na língua portuguesa.

A comercialização em larga escala de ferramentas de TA começa oficialmente na década de 1980, porém a primeira ideia de tentativa de automatização da tradução surge ainda entre os anos de 1947 e 1949, logo após a Segunda Guerra Mundial (1939-1945).

Em verdade, anos antes em 1933 outro evento marcaria a primeira tentativa de mecanização da tradução, trata-se da construção de duas máquinas que tinham o objetivo de mecanizar dicionários bilíngues. A primeira foi construída pelo francês George Artsrouni e tinha algumas funções gerais, além de incluir um dicionário multilíngue mecânico. Já a segunda, do russo Petr Trojanskij, tinha um dicionário mecânico que objetivamente codificava e interpretava funções gramaticais usando símbolos universais com base no Esperanto. Ambas máquinas tiveram suas patentes reconhecidas na França e Rússia, respectivamente, tornando-se oficialmente os primeiros experimentos reconhecidos nesta área.

No entanto, o avanço mais representativo em direção à TA ocorre por intermédio de Warren Weaver. Professor de matemática e considerado precursor na invenção da TA, Weaver era diretor do departamento de Ciências Naturais do *Rockefeller Foundation* em Nova Iorque, que desenvolvia pesquisas em tecnologia com base em criptografia. Seu memorando enviado ao professor Norbert Wiener do Instituto de Tecnologia de Massachusetts (MIT), em 4 de março de 1947, marcou os primeiros passos rumo ao desenvolvimento de um sistema de TA.

A correspondência questionava a respeito da possibilidade no desenvolvimento de um computador capaz de traduzir, com o objetivo de ajudar a Unesco que enfrentava problemas relacionados à paz mundial após a Segunda Guerra (1939-1945), em razão da dificuldade de comunicação ocasionada pela barreira entre idiomas.

A comparação entre tradução e decodificação feita pelo professor Weaver no memorando enviado tornou famosa sua frase que dizia “quando vejo um texto em russo acredito ter sido escrito em inglês, porém codificado com símbolos estranhos. Então, vou decodificá-lo”. Uma ideia simplista sobre a criação de algoritmos que fariam apenas a substituição de palavras do texto-fonte por palavras no texto-alvo por intermédio de uma máquina.

Dois anos depois, em 15 de Julho de 1949, um novo memorando foi enviado por Weaver ao Instituto de Tecnologia de Massachusetts (MIT) e dessa vez incluía as diretrizes para construção de um computador capaz de fazer a análise dos elementos básicos dos idiomas e que pudesse trabalhar de maneira estatística com esses mesmos elementos, para isso era necessário a ajuda de especialistas em linguística para que os elementos comuns a certos idiomas pudessem ser mapeados.

Andrew Booth assumiu o projeto da construção da máquina com o professor Weaver que resultou no primeiro protótipo de máquina de TA (vale ressaltar que nenhum dos dois tinha conhecimento a respeito das máquinas francesa e russa construídas na década anterior).

A necessidade da busca por um mecanismo de TA aumenta conforme cresce a demanda por tradução de idiomas estrangeiros nas áreas acadêmica, de finanças e diplomática. A ideia do uso da TA surge com a necessidade de alcançar mais velocidade na tradução de textos e aceitava menor qualidade na tradução. Isto porque se tratava de um grande volume impresso que precisava de entendimento, mesmo que de maneira geral, visto que o governo temia potenciais inimigos, uma vez que ainda estavam num período pós-guerra.

Com isso, grupos de pesquisa se debruçaram em estudos para a construção de um computador capaz de traduzir, incentivados por grandes quantias no financiamento de suas pesquisas, tanto por parte de órgãos públicos como de algumas instituições privadas. No capítulo III, explicaremos como ocorreu a formação desses grupos de pesquisa em cada programa de investigação da TA, assim como o impacto de seu trabalho no desenvolvimento da TA e na evolução da área.

4. Procedimentos metodológicos da pesquisa

A pesquisa contempla a análise dos aspectos históricos, sociais, científicos e técnicos que motivaram a evolução tecnológica da TA. Para alcançar resultados satisfatórios a partir dos objetivos delineados para esta dissertação, a HL servirá como base conceitual e metodológica (em associação, quando procedente, com os Estudos da Tradução e da Filosofia da Linguagem), uma vez que possibilita definir ferramentas para análise crítica dos elementos científicos, históricos e sociais, além de sua dimensão metodológica alcançar aspectos teóricos (no qual se podem incluir aspectos técnicos/tecnológicos), temporais e sociais.

A organização de um trabalho historiográfico pode seguir métodos diferentes de acordo com o objetivo do historiógrafo. Neste trabalho adotaremos o modelo de reconstrução da narrativa historiográfica com base nos períodos de continuidade e descontinuidade dos métodos de TA, visto que o desenvolvimento da TA se dá em conjunto com as Ciências Exatas, o que nos remete a um modelo voltado a rupturas, na maior parte das vezes definitiva, em que uma nova tecnologia desenvolvida chega com o objetivo de substituir a anterior.

No entanto, a circulação e recepção de uma nova tecnologia passa por um processo de transição de teoria científica que iremos observar nesta narrativa historiográfica, nos mostrando os momentos de ruptura com padrões de pesquisa tidos como contemporâneos mas questionáveis do ponto de vista prático.

Para a elaboração dessa narrativa, adotaremos os conceitos de programas de investigação proposto por Swiggers (1981) e de grupos de especialidade teórica de Murray (1994), conforme pressupostos teóricos apresentados na seção 1 “A Historiografia da Linguística (HL): o referencial teórico” deste capítulo. Esses conceitos apresentam ferramentas que serão adotadas para mapear os agentes motivadores do desenvolvimento e legitimação da TA. Abrangendo desde sua idealização na década de 1940, passando pelos eventos que ocasionaram ascensão na década de 1950, seguido pelo evento que ocasionou a suspensão das pesquisas americanas no final da década de 1960, com a consequência de duas décadas de paralização dos estudos em TA e retomada definitiva somente a partir dos anos de 1980.

Na sequência, apresentamos os procedimentos metodológicos adotados nesta pesquisa.

4.1 Parâmetros de análise

Seguindo métodos da HL apresentados, entre outros, por Altman (2004), Batista (2013, 2019) e Swiggers (2013), a historiografia que se estabelece nesta dissertação contempla a sistematização, em perspectiva histórica, de questões relacionadas à produção, circulação e recepção de modelos e técnicas de TA ao longo do recorte temporal estabelecido para esta pesquisa, tendo em vista o material selecionado como corpus indicado nas tabelas a seguir.

O corpus a partir do qual serão captados os parâmetros de análise é composto por: a) correspondências históricas com ideias e métodos da TA no início de seu desenvolvimento; b) artigos acadêmicos com os modelos e recursos de TA publicados em revistas científicas; c) anais de congresso com discussões no meio acadêmico em relação ao trabalho dos grupos de especialidade; d) artigos em revistas não-científicas com opiniões sobre a TA no mercado de tradução.

A escolha desse material como fonte de dados para análise dos programa de investigação da TA foi definida a partir de critérios que auxiliassem na reconstrução da história da TA. Com base nos critérios definidos (que serão detalhados na

sequência), priorizamos os documentos acadêmicos (que constituem a maior parte do material de análise) com o objetivo de buscar evidências sobre a incidência da língua e o clima de opinião onde circulavam as ideias linguísticas no desenvolvimento da TA. Adicionalmente, selecionamos alguns documentos diferentes dos acadêmicos, como artigos em revistas não-científicas, que tem a função de colocar em perspectiva as evidências capturadas a partir da observação da produção acadêmica.

Durante o processo de seleção, foram escolhidos os documentos que incluíssem evidências para o desenvolvimento de uma análise crítica dos programas de investigação da TA, e a partir da seleção desse material diversificado foram definidos os seguintes parâmetros de análise:

a) parâmetros internos: visão de tradução adotada (ou subjacente) aos modelos, nível linguístico privilegiado no método (por exemplo: palavra, sentença, significado), técnicas e procedimentos adotados em TA.

b) parâmetros externos: periodização (recorte temporal de elaboração, difusão e uso de ferramentas de TA), clima intelectual de proposição e uso de modelos de TA (relacionado principalmente a questões tecnológicas e econômicas), motivação dos propositores do modelo (como inovação tecnológica e ganhos financeiros), entendidos, assim, como agentes da produção, circulação e recepção de ideias linguísticas.

4.2. Material de análise: seleção e coleta

Segundo Altman (2004, p.52), o historiógrafo determina critérios para seleção da fonte de dados de acordo com o objeto que pretende investigar, resultando em duas categorias de objeto: a) o objeto material e b) o objeto formal.

O objeto material diz respeito aos dados selecionados como referente a partir da observação do historiógrafo, enquanto que o objeto formal está relacionado ao ponto de vista que orientará sua análise a partir do referente observado.

Para este estudo, o objeto formal, que representa o corpus de análise é constituído por correspondências históricas, artigos acadêmicos, anais de congressos e artigos em revistas não-científicas (ver seção 4.1 “Parâmetros de análise”) que foi selecionado e categorizado resultando em três tipos:

1) documentos históricos: dados que evidenciam fatos marcantes no desenvolvimento da TA, tais como as correspondências entre pesquisadores que relatam as primeiras tentativas rumo à automatização da tradução, bem

como estudos sobre os primeiros modelos de TA com base na tradução “palavra por palavra”. Esse material tem como objetivo fundamentar a reconstrução histórica do surgimento da TA. Neste material também está incluído o relatório da ALPAC;

2) documentos sobre funcionalidade: dados que mostram a TA em funcionamento, a partir de estudos de caso, experimentos publicados e discussões em congressos, com o objetivo de ilustrar as diferentes técnicas adotadas no desenvolvimento de modelos e processos de TA.

3) documentos sobre ferramentas: dados que mostram a legitimação de sistemas de TA desde o final dos anos de 1940, incluindo os modelo com base em regras, em exemplos, estatística, neural e *deep learning*.

A partir dessa categorização inicial, selecionamos dois grandes recortes temporais os quais contemplarão três conjuntos de documentos que veremos mais à frente. Esses recortes foram definidos com base no critério de seleção dos fatos marcantes provocados por rupturas históricas que serão discutidas no capítulo III. Assim, os dois grandes períodos que serão contemplados pela seleção do corpus são: 1) o período de 1947 à 1967 e 2) a partir da década de 1980.

O primeiro conjunto de documentos são os documentos históricos que incluem 12 arquivos, contabilizando aproximadamente 220 páginas, constituído por correspondências, artigos acadêmicos e anais de congressos, revistas não-científicas e pelo relatório da ALPAC. A Tabela 1 relaciona os tipos de documento, recorte temporal e quantidade definidos como fonte de dados históricos:

Tabela 1: Dados históricos

Tipo de documento	Ano de publicação	Quantidade
Correspondências	1947, 1948 e 1949	3
Artigos acadêmicos e anais de congresso	1951 a 1960	6
Relatório técnico da ALPAC	1966	1
Revistas não-científicas	1967	2

A partir da década de 1980 iniciou-se o movimento de retomada das pesquisas em TA, que tinham sido interrompidas no final dos anos de 1960 em virtude da publicação do relatório da ALPAC (que será detalhado no capítulo III). A análise dessa retomada é feita com base no segundo conjunto de documentos que visa investigar a evolução para o modelo de TA com base estatística (*SMT, Statistical Machine*

Translation), modelo esse que representa o passo mais próximo em direção ao desenvolvimento do modelo com base neural (*NMT, Neural Machine Translation*) e *deep learning*, esses últimos são as tecnologias em TA mais modernas na atualidade.

Os documentos selecionados neste segundo conjunto são artigos acadêmicos e anais de congressos que mostram as tendências nas pesquisas do modelo de base estatística, assim como o movimento de abertura do mercado de tradução impulsionado pelo cenário daquele período. A Tabela 2 ilustra os documentos selecionados no recorte temporal de 1988 à 2013 como fonte de dados para análise do modelo de base estatística, contabilizando aproximadamente 100 páginas:

Tabela 2: Dados do modelo de base estatística

Tipo de documento	Ano de publicação	Quantidade
Artigos acadêmicos e anais de congressos	1988, 1993, 1994, 1995, 1997, 1999, 2001, 2005, 2013	9

O terceiro conjunto de documentos é constituído por artigos acadêmicos e artigos em revistas não-científicas que foram selecionados em virtude das informações contidas sobre o modelo de base neural, desenvolvido a partir do conceito de RNA e *deep learning*. O desenvolvimento dessa tecnologia começou com mais força a partir da virada do milênio e continua sendo a área de maior interesse dos estudos da TA; o modelo tem sido adotado por corporações desde 2016.

Fazem parte desse conjunto anais de congressos selecionados no recorte temporal de 2002 à 2015 e artigos em revistas não-científicas de 2010 à 2016, totalizando 11 documentos e aproximadamente 70 páginas, conforme relacionados na Tabela 3:

Tabela 3: Dados do modelo de base neural e *deep learning*

Tipo de documento	Ano de publicação	Quantidade
Anais de congresso	2002, 2007, 2010, 2012, 2014 e 2015	7
Artigos em revista não-científica	2010, 2011 e 2016	4

Em resumo, o corpus de análise está composto por três conjuntos de documentos separados nas categorias: 1) dados históricos, 2) dados do modelo de base estatística e 3) dados do modelo de base neural e *deep learning*. Na Tabela 5 temos o resumo geral do material que conta com 32 documentos e aproximadamente 400 páginas.

Tabela 4: Resumo das fontes de dados (corpus)

Tipo de documento	Quantidade
Dados históricos	12
Dados do modelo de base estatística	9
Dados do modelo de base neural e <i>deep learning</i>	11
TOTAL	32

4.3 Etapas de análise da pesquisa

Para a construção de uma narrativa historiográfica da evolução da TA, a pesquisa contemplou os seguintes procedimentos: a) descrição das etapas de uma história da tradução, tendo em vista circunscrever historicamente a presença da TA; b) elaboração de uma reflexão epistemológica sobre a prática da TA; c) análise crítica dos desenvolvimentos históricos da TA, seus processos e mecanismos; d) elaboração de uma síntese crítica e interpretativa de uma história da TA a partir do desenvolvimento das etapas anteriores. E finalmente, seguimos as diretrizes de Swiggers (2013) na definição das três etapas formais do desenvolvimento da análise da evolução da TA, conforme descrito na Tabela 5:

Tabela 5: Etapas do desenvolvimento da análise

Primeira etapa: fase heurística	<ol style="list-style-type: none"> 1. Leitura do material de análise e da bibliografia teórica (tarefa que a rigor se estende ao longo de todo o tempo da pesquisa). 2. Leitura e análise de propostas metodológicas em HL, e nas áreas em que o diálogo interdisciplinar se estabelece. 3. Seleção de aspectos da história da tradução relevantes para a TA. 4. Seleção de movimentos de continuidade e descontinuidade na história da TA.
Segunda etapa: fase hermenêutica	<ol style="list-style-type: none"> 1. Seleção de fragmentos dos documentos que propiciam a análise dos parâmetros internos tais como estabelecidos. 2. Circunscrição da análise de parâmetros internos a aspectos externos, considerando os itens indicados nos parâmetros. 3. Interpretação e análise historiográfica, em termos interpretativos, a partir da correlação entre parâmetros internos e externos.
Terceira etapa: fase executiva	<ol style="list-style-type: none"> 1. Produção de análises interpretativas que pretendem demonstrar os resultados da reflexão e da interpretação sobre uma história da TA. 2. Elaboração do texto da dissertação.

CAPÍTULO II

EPISTEMOLOGIA DA TRADUÇÃO

1. Operações da tradução

O termo *tradução* pode abranger uma série de domínios, desde a noção básica da passagem de uma língua formal para outra, isto é, tradução linguística, extrapolando para outras formas como tradução cultural, tradução audiovisual, tradução de língua de sinais, entre outras. Devido a essa extensa abrangência nas modalidades de tradução torna-se necessário discutir sobre a natureza da tradução e do ato de traduzir, com o objetivo de estabelecer o tratamento do termo *tradução* neste trabalho.

Ao pensar em tradução e sobre o que é tradução não encontramos uma resposta definitiva. Para refletir sobre o que é tradução, qual sua problemática central e sua ressonância social é preciso fazer um exercício de reflexão que problematize acerca do que implica a tradução e o ato de traduzir. Nessa perspectiva, discutiremos sobre tradução a partir da observação das operações que envolvem as modalidades (ou os tipos) de tradução, isto é, o que caracteriza tradução e seus processos.

O conceito de operações da tradução está fundamentado na definição da natureza do texto a ser traduzido, ou seja, o tipo de texto (científico, técnico, jornalístico, literário, etc.), e na diferenciação da função desse texto que o classifica como uma operação linguística ou literária, determinando qual tipo de tradução poderá ser adotada.

Dos três tipos de tradução definidos por Roman Jakobson (BASSNETT, 2002, p. 23), a intralingual, interlingual e intersemiótica, a primeira das operações da tradução é a tradução intralingual ou reformulação que se caracteriza pela “interpretação de signos linguísticos por meio de outros signos linguísticos dentro da mesma língua” (JAKOBSON, 1959 apud OUSTINOFF, 2003, p. 73).

Segundo Oustinoff (2003, p. 73-76), o processo tradutório na tradução intralingual exerce um papel fundamental no exercício da tradução, visto que é necessário primeiro ocorrer uma reformulação do enunciado fonte (dentro da própria

língua de partida) que, apenas depois de concluída, poderá passar para a reformulação na língua de chegada que resultará no enunciado alvo. Entende-se aqui enunciado fonte como o texto-fonte (na língua de partida) e enunciado alvo como o texto-alvo (na língua de chegada).

O Diagrama 1 representa a operação de tradução intralingual, onde ocorre o processo de reformulação do EF (enunciado-fonte) dentro da LP (língua de partida).

Diagrama 1: Tradução intralingual

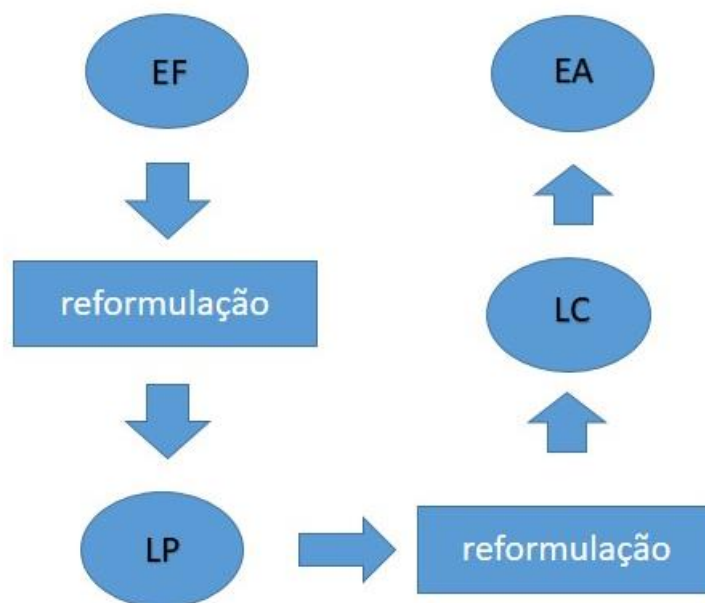


Fonte: autoria própria (2020)

Na tradução interlingual (tradução entre línguas diferentes), o processo de reformulação acontece em duas etapas: a) primeiro no EF (dentro da LP) e b) depois dentro da LC para alcançar o EA.

O Diagrama 2 representa o processo de tradução interlingual quando ocorre as duas etapas de reformulação, onde o EF passa pela primeira reformulação ainda na LP e, em seguida, pela segunda reformulação dentro da LC, resultando no EA que corresponde à tradução.

Diagrama 2: Tradução interlingual



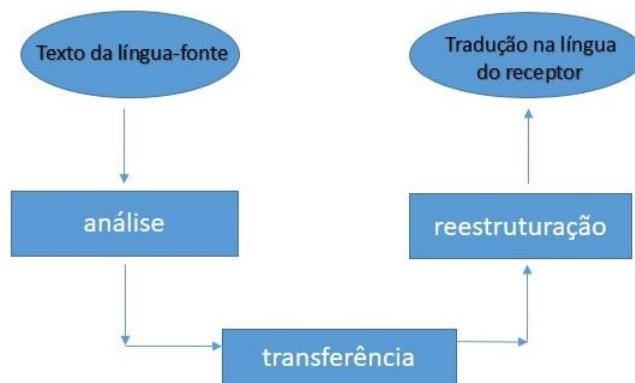
Fonte: autoria própria (2020)

As etapas de reformulação que ocorrem na tradução interlingual acontecem em diretrizes diferentes. Quando a reformulação acontece dentro da LP implica na ordem da compreensão e quando ocorre na LC corresponde à ordem da expressão.

Entretanto, no modelo de processo de tradução interlingual proposto por Eugene Nida (1969, apud BASSNETT, 2002), a reformulação aparece como as etapas de análise e reestruturação, passando por uma fase de transferência de sentido.

O Diagrama 3 representa esse modelo, em que a análise do texto-fonte (na LP) passa pela etapa de transferência de sentido para alcançar a reestruturação (na LC) e concluir-se como uma tradução. Neste caso, a operação de tradução é denominada “decodificação e recodificação”.

Diagrama 3: Decodificação e Recodificação



Fonte: BASSNETT, S. Translation Studies 3rd ed. London and New York: Routledge, 2002. p. 25

A operação de decodificação e recodificação apresentada desse modo aparenta um processo direto e simples, no entanto, não se trata de uma operação matemática que consegue ser resolvida a partir de sua lógica, mas de uma operação linguística complexa.

O processo de decodificação e recodificação serve para exemplificar que os elementos extralinguísticos afetam o texto como um todo (de maneira vertical) e as palavras se articulam (de maneira horizontal), segundo o sistema de relações interdependentes (relações sintagmáticas e relações associadas), existentes em um texto (BASSNETT, 2002, p. 27). A tradução opera num critério que extrapola a linguística unicamente e o processo de decodificação e recodificação acontece como parte da natureza do processo tradutório.

Na tradução intersemiótica o processo de tradução acontece a partir da “interpretação de signos verbais por meio de sistemas de signos não verbais” (BASSNETT, 2002, p. 23) e essa abordagem é responsável por trazer os elementos extra linguísticos às teorias da tradução, o que colabora para destacar as diversas formas de tradução existentes.

Outra operação da tradução é a “transposição ou recategorização”, que consiste na modalidade que opera no significante, ou seja, na forma do signo linguístico. Esse tipo de tradução resulta na substituição de uma categoria gramatical original no texto-fonte por outra categoria gramatical no texto-alvo.

Assim, a operação de transposição atua na recategorização do significado porque o princípio da literalidade, isto é, da tradução direta ou literal não pode ser aplicado naquele texto-fonte. Como resultado temos uma tradução oblíqua que não apresenta uma literalidade estrutural entre texto-fonte e texto-alvo, em virtude de uma incompatibilidade de significantes entre as línguas.

Essa operação é uma das que ocorre com mais frequência dentre todas as operações da tradução, visto que há casos em que a recategorização gramatical é obrigatória para conseguir alcançar o significado pretendido. Como resultado, a transposição pode ser identificada por seguir o princípio da idiomacidade (tradução oblíqua) em detrimento do princípio da literalidade (tradução literal), em virtude das incompatibilidades estruturais das línguas.

Por último, temos a operação de “modulação” que ocorre como uma tradução oblíqua, uma vez que opera no significado, no sentido do signo linguístico. A modulação acaba por afetar o sentido porque muda o ponto de vista em relação ao texto-fonte. São vários os tipos de modulações que podem ser aplicados nessa mudança de ponto de vista, por exemplo, a mudança de perspectiva do abstrato pelo concreto, ou do geral pelo particular, ou da causa pelo efeito, ou do meio pelo resultado, ou da parte pelo todo, ou de uma parte por outra, entre outros (OUSTINOFF, 2003, p. 84), definitivamente são quase que incontáveis as possibilidades de modulações que o tradutor pode lançar mão em sua tradução.

As operações de tradução devem ser observadas como indissociáveis do processo de “reenunciação” (OUSTINOFF, 2003, p.82). O EF é sempre reformulado para se transformar em um novo enunciado (o EA). Esse processo é parte integrante do processo de tradução, posto que a tradução ocorre a partir de contextos, e não com base em palavras desconectadas. Dessa forma, podemos concluir que não é possível realizar traduções mecanicamente sem levar em consideração a análise do texto-fonte com base na língua em ação e não, tão somente, considerando essa língua como uma entidade abstrata.

2. Tradução e cultura

A reflexão sobre o tema cultura na problematização acerca do que é tradução não pode ser esquecida. A tradução ocupa um espaço central na linguagem que orbita no universo da comunicação, em que, por sua vez, precisa lidar com a pluralidade de línguas que o processo de globalização tem promovido.

A globalização ocasiona a aproximação de culturas diferentes, bem como, de certa forma, democratiza o acesso que, anteriormente, seria privilégio de determinados grupos nas sociedades, onde esses conseguiam se comunicar com outras culturas por causa de sua condição social, intelectual ou política.

Nesse cenário, conseguimos visualizar a importância da pluralidade das línguas como princípio da diversidade cultural. Uma língua significa uma maneira de pensar, de argumentar e de agir em uma sociedade, representando sua identidade e diferenciando o modo como funciona uma sociedade em relação à outra. As diferenças entre culturas estão fundadas nessa pluralidade que, segundo Homi Bhabha, é impossível de ser encaixada em “uma moldura universalista” (RUTHERFORD, 1996, p. 36).

A tradução como técnica pode ser entendida como um complexo meio de comunicação, visto que as línguas não são simples instrumentos que podem ser transpostos entre si, pelo fato de estarem situadas na base da identidade de uma sociedade.

Assim, uma discussão sobre uma língua universal é muito complicada e complexa porque esbarra no questionamento fundamental no desenvolvimento de uma língua: a uniformização de uma cultura universal é impossível. Edward Sapir afirma que “a língua é um guia para a realidade social” em que “a experiência é amplamente determinada pelos hábitos linguísticos de uma comunidade e cada estrutura separada representa uma realidade separada” (1956 *apud* BASSNETT, 2002, p. 22).

Como vimos, a relação entre língua, cultura e tradução torna-se indissociável, o que nos leva aos problemas enfrentados pela tradução sobre o que é traduzível e qual o papel do tradutor nesse processo de reformulação. Nesse ponto temos que retomar a questão sobre a natureza do texto a ser traduzido, ou seja, o tipo de texto.

Conforme aponta Oustinoff (2003, p. 115), ao estudarmos a tradução a partir do significante (a forma) é possível visualizar sua natureza polivalente. Uma palavra escrita em uma página apresenta, inicialmente, uma forma visual, mas pode ser

traduzida pela fala e se torna uma forma auditiva, ou pode ser traduzida por linguagem de sinais e se torna um gesto e, mais ainda, pode ser traduzida em alfabeto braile e se torna um elemento tátil.

Essa natureza polivalente impossibilita tratarmos uma tradução como um elemento imutável, isto quer dizer que se mais de um tradutor fizer a tradução do mesmo texto, nenhuma será igual a outra. Ao traduzirmos podemos dizer que nunca é igual, nem ao original e nem a outra tradução. Se fosse igual teríamos que denominar como transcodificação de termos e, no caso de uma outra tradução, como cópia. Esse processo de simples transcodificação ou cópia pode ser feito por uma máquina, porém a tradução de determinados tipos de textos requer mais habilidades do que o processo automático de transposição de códigos. Na verdade, o processo de tradução requer a transposição de signos. O signo linguístico é o elemento traduzível mais tangível de um texto, porém, como já mencionado, precisa ser analisado de acordo com suas relações interdependentes dentro do texto.

A equivalência do significado, no sistema de signos, encontra problemas com as teorias da tradução, em especial, quando tentamos estabelecer uma teoria da equivalência com base na noção de equivalência que temos da matemática, ou seja, a busca por uma equivalência precisa. Não é possível encontrar equivalência precisa e completa numa tradução como em uma operação matemática. Uma tradução é uma imitação do original que sofreu uma transformação durante o processo de transposição de signos, assim como passou pela interpretação do tradutor que trouxe sua identidade ideológica no processo de análise e reestruturação.

Segundo Bassnett (2002, p. 32-38), para Jakobson a confluência entre significante e significado pode supor uma equivalência na diferença entre os elementos durante a transposição dos signos em cada língua, enquanto para Nida o princípio do efeito equivalente é de que o receptor na língua de chegada tenha a mesma sensação do efeito que a mensagem causou na língua de partida. Esse último pode causar controversas e precisa ser discutido para problematizar situações em que o original não acabe resultando, por exemplo, numa descaracterização de estilo, entre outras mudanças, em detrimento de uma adaptação para um momento atual, a fim de alcançar o efeito equivalente.

A questão da equivalência pode ser considerada o problema central da tradução. Os Estudos da Tradução tratam o tema como um dos assuntos centrais nas discussões sobre desenvolvimento de novas teorias, por entender que se trata de uma

questão fundamental nos processos da tradução, porém ainda encontra grandes dificuldades para estabelecer definições.

Outro tema que merece destaque é a questão da intraduzibilidade. A intraduzibilidade pode ser linguística, que se define pela falta de um substituto lexical ou sintático na língua de chegada para um elemento da língua de partida, mas, geralmente, o que causa mais dificuldades para o tradutor é a intraduzibilidade cultural, que como se auto explica está ligada às diferenças culturais entre línguas. Aqui tratamos diferenças culturais num cenário de falta de recursos que correspondam a um contexto cultural na língua de chegada, em que causaria diferença no comportamento entre sociedades, impactando diretamente nas escolhas do tradutor em relação à tradução cultural.

Essas escolhas precisam ser refletidas e discutidas mais profundamente, a simples adaptação de um sistema de valor da cultura da língua de partida na cultura da língua de chegada pode parecer o caminho fácil, especialmente, se se tratar de uma escola ideológica com a qual o tradutor se identifica. No entanto, pode ser um caminho controverso, pois resultaria na determinação das intenções originais do autor em um texto de próprio conteúdo do tradutor.

De acordo com Rutherford (1996, p. 36), Bhabha questiona a tradução cultural do ponto de vista da reflexão em sociedade a respeito de como tratar as transformações culturais que as rodeiam, revisando seus valores, normas e sistemas de referências com o intuito de encontrar o caminho de sua identidade, no sentido de não se fechar totalmente às influências externas, mas, ao mesmo tempo, não deixar tradições e memórias de sua formação como sociedade serem corrompidas. A noção de diversidade cultural, no sentido de unir diferentes culturas para que elas possam coexistir naturalmente, é impossível.

A ideia de tradução cultural como um elemento adaptável não é tão simples como parece. O pensamento de que todas as culturas podem ser compreendidas com base em conceitos universais como humanidade, classe e raça é absolutamente limitador. A partir dessas reflexões de Bhabha (Rutherford, 1996, p. 36) em relação à organização social e à construção de práticas sociais, entendemos que não é apenas o princípio da similaridade, familiaridade com certos conteúdos, que representa práticas culturais, mas o que explica a natureza transformadora de uma cultura é sua capacidade de desenvolver questionamentos e discussões em sociedade.

A noção de tradução cultural passa então a ser questionada, uma vez que negaria a natureza de uma cultura antecedente ou original, surgindo assim o conceito de hibridismo cultural. O hibridismo cultural ocupa um espaço para tratar dos temas contraditórios e conflitantes de elementos linguísticos e culturais, uma vez que se entende o conceito a partir de uma perspectiva da linguagem e da identidade (SOUZA, 2004, p. 8).

Dessa forma, podemos entender a noção de que as culturas estão em constante movimento porque estão constantemente em um processo de hibridização em que algo novo é desenvolvido, novas ideias abrem espaço e novas políticas começam a ser discutidas, ou seja, a movimentação das práticas culturais na verdade é resultado do processo de hibridação cultural.

Agora, com essa percepção em mente, conseguimos justificar que é impossível uma tradução exata e equivalência completa. Assim, é papel do tradutor considerar a diversidade de estruturas disponíveis na língua de chegada levando em conta a diversidade de gênero, mas também a diversidade cultural, considerando o sistema linguístico do texto-fonte, bem como seu sistema cultural.

3. Mecanização da linguagem

A linguagem natural é a linguagem usada pelos seres humanos, o recurso de comunicação nativa do homem, a língua pura; não precisamos dizer como foi criada, ela faz parte da descoberta humana. Logo, a linguagem formal situa-se num campo que tem o objetivo de estudar a linguagem a partir de modelos matemáticos, aqui o entendimento sobre estudo da linguagem tem uma conotação ampla e inclui as linguagens de programação dentro das ciências da computação.

Um dos grandes desafios das ciências da computação é desenvolver sistemas que sejam capazes de entender a linguagem natural de acordo com os parâmetros da linguagem formal. Assim, a língua natural se torna objeto para práticas técnicas e tema de disciplinas relacionadas aos estudos conduzidos pelas ciências da computação.

O interesse pela mecanização da linguagem vem com a necessidade de responder a problemas, como de fato acontece com a maior parte das descobertas mundiais. A busca por respostas impulsionou três principais revoluções, consideradas tecnolinguísticas, conforme sugerido por Sylvain Auroux (1998, p. 289): 1) A invenção da escrita; 2) a gramatização das línguas do mundo e 3) o tratamento eletrônico da informação apresentada em linguagem natural.

Este estudo se debruça na investigação sobre o terceiro item relacionado ao tratamento eletrônico da informação em língua natural, em que a TA ocupa a posição de destaque entre um dos primeiros domínios linguísticos que se tentou mecanizar. Os primeiros protótipos de sistemas que se propunham traduzir automaticamente empregavam operações de decodificação, em que a tradução direta de palavra por palavra associava cada unidade lexical da língua de partida a um correspondente na língua de chegada e tal operação era fundamentalmente baseada em dicionários bilíngues ou glossários, que evidentemente se mostrou bem pouco eficiente em termos de qualidade linguística.

O cenário pós Segunda Guerra Mundial (1939-1945) impulsionou o avanço de pesquisas relacionadas ao Tratamento Automático das Línguas Naturais (TALN), bem como o surgimento de modelos matemáticos com base numa analogia sobre o funcionamento neurológico, descoberto pelos matemáticos W. McCulloch e W. Pitts em 1943. Uma descoberta que significa todo o suceder de pesquisas em torno da concepção das máquinas inteligentes e dos estudos em torno da IA.

O desenvolvimento de estudos com base em RNA (Redes Neurais Artificiais) são fundamentais no desenvolvimento da IA, lidar com a máquina na perspectiva de um cérebro humano e a busca pela reprodução das tarefas humanas é uma das maiores revoluções no século XX. Mas o tratamento da língua natural a partir da abordagem matemática não seria novidade, pois desde o início da formalização das línguas em gramáticas, ainda no século XVII, os sistemas formais matemáticos servem de inspiração, assim como a questão da calculabilidade, com o objetivo de esquematizar a língua natural dentro de sistemas de linguagens formais.

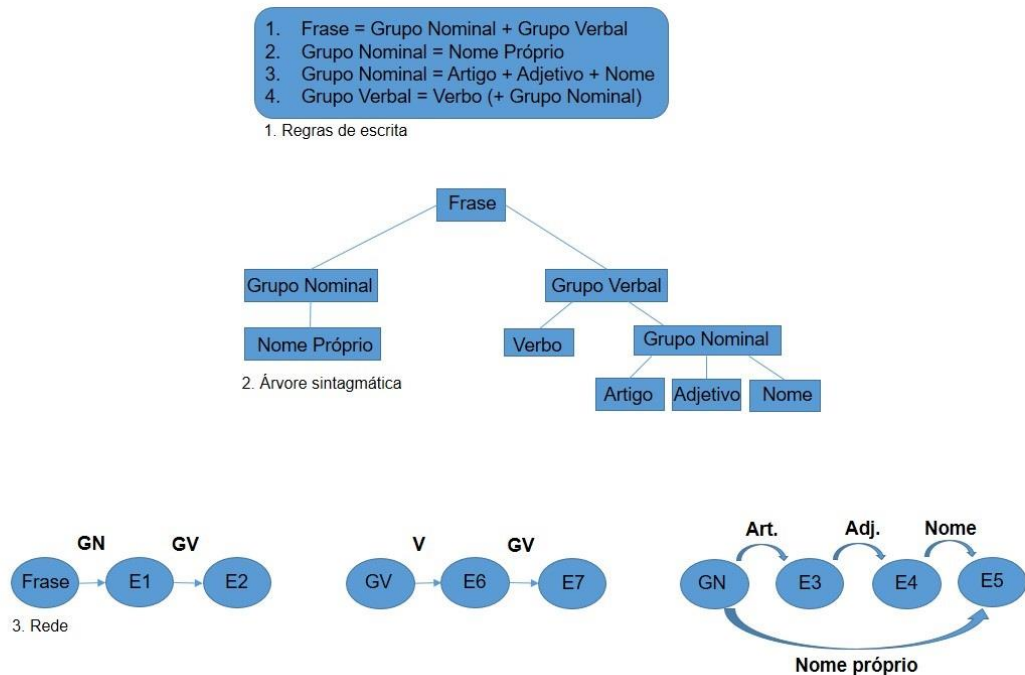
Nesse sentido, Chomsky teve um papel fundamental quando consolidou o conceito de “concepção matemática das linguagens formais para descrição das línguas naturais” (AUROUX, 1998, p. 462), a partir de uma abordagem da linguagem formal com regras baseadas em algoritmos. No entanto, ainda encontravam muitas dificuldades para implementar na máquina a gramática proposta por Chomsky.

No início dos anos 1970, surgiram os primeiros programas desenvolvidos com a função de analisadores sintáticos de línguas naturais, baseados no conceito de ATN (*Augmented Transition Network*, Redes de Transição Ampliadas), que consistiam em sistemas de representação gráfica empregados na codificação das informações gramaticais. O esquema de representação supõe codificar regras de escrita, por exemplo, regras de formação de sintagmas e de frases equivalentes do ponto de vista

da gramática formal, de maneira que possam ser transformados em algoritmos para interpretação da máquina.

O Diagrama 4³ ilustra a representação de uma frase separada em unidades lexicais e esquematizada primeiro em forma de árvore sintagmática e, em seguida, em forma de rede.

Diagrama 4: Representação de ATN



Fonte: AUROUX, S. *A filosofia de linguagem*. Campinas: Ed. Da Unicamp, 1998. P. 466-467

O objetivo no desenvolvimento de um sistema com base na análise da estrutura sintática torna-se a alternativa para buscar a solução dos problemas de qualidade na estratégia da TA palavra por palavra. Porém, as pesquisas evoluíram e as discussões também, assim a concepção que o conhecimento da estrutura sintática seria fundamental para a reconstrução do sentido começou ser questionada e surgiram novos modelos com base na gramática sistêmico-funcional proposta por M. Halliday (1978).

³ A figura pretende ilustrar de maneira simplificada o processo de codificação, mas não corresponde a um exemplo completo de representação de uma ATN, porque não apresenta as condições que precisariam ser definidas no percurso dos arcos, os quais representam as transições das unidades lexicais de um estado para outro.

A ideia nesse momento era a de uma abordagem de análise do enunciado a partir de três módulos (inferencial, semântico e pragmático), no qual funcionariam de forma linear, uma vez que a visão em relação à concepção de que o processo linguístico que passa por uma operação sintática e, em seguida, uma interpretação semântica, como duas etapas separadas, já estava descartada.

A partir de então, trabalhos de pesquisa em torno da IA se intensificaram e um nome de destaque surgiu, Terry Winograd, que se destaca por ser responsável pelo desenvolvimento do sistema SHRDLU, que representa um marco na evolução tecnológica de sistemas de compreensão de língua natural que combina a sintaxe, semântica e lógica, ainda na década de 1970.

O trabalho com o sistema SHRDLU de Winograd nos anos de 1970 abriu caminhos para que alguns pesquisadores deixassem de lado os estudos em torno dos problemas de tratamento da linguagem para se dedicarem ao da representação dos conhecimentos, chegando-se à conclusão de que se a máquina fosse alimentada com um grande volume de conhecimento poderia aprender e melhorar seu desempenho. Com isso, entenderam que o objetivo no desenvolvimento da IA estaria em realizar uma “simulação por computador das atividades inteligentes do homem” (AUROUX, 1998, p. 297). Isso só seria possível com base no conhecimento humano implantado na máquina, resultando em um dos motores ideológicos para o desenvolvimento de sistemas de tratamento automático das línguas.

Esse processo de ensino da máquina é baseado no conceito dos sistemas-perito, em que um banco de dados de conhecimento é implantado na máquina, a fim de que ela possa tratar de maneira automática as questões que lhe forem solicitadas, por exemplo, a TA. A ideia de perito remete à de um perito-humano, ou seja, um especialista com o objetivo de transferir os conhecimentos humanos para a máquina que funcionará como um recurso externo.

A metodologia pressupõe uma etapa de formalização do conhecimento implantado que significa torná-lo explícito e invariável, como no caso das tarefas repetitivas, no qual a máquina deverá sempre interpretar igual, independente das circunstâncias e das variações possíveis, isso porque a ambiguidade para a máquina não é tão facilmente transponível. No entanto, a formalização também resulta na perda ou no distanciamento de situações reais, com isso não é possível falar em automatização de línguas sem considerar as perdas em relação a uma realidade concreta.

Quando tratamos da mecanização da linguagem, temos que pensar num recurso capaz de reproduzir atividades humanas inteligentes, como interagir automaticamente com um indivíduo sem parecer que seja uma máquina. A maneira de operação da máquina está totalmente ligada à forma como ela é ensinada, quando o conhecimento adquirido e formalizado resulta na interação com o humano. Esse por sua vez, fornecerá mais conhecimento para melhoria no desempenho da máquina, tornando assim um processo cíclico de aprendizagem com infinitas possibilidades de desenvolvimento de novas habilidades na máquina.

Com a evolução nos estudos da automatização da linguagem e da TA, evoluiu também a maneira como se enxergam as complexidades das línguas naturais e do processo de tradução. Assim entende-se hoje que não seria possível uma tradução automática realizada sem uma análise linguística do texto, com foco em reproduzir a compreensão do conteúdo, assim como fazer distinção entre os tipos de textos que conseguem ser mais bem traduzidos, automaticamente, devido a sua complexidade.

Segundo Aurox (1998, p. 293-294), as linhas teóricas que se mostram mais viáveis nas pesquisas para o avanço da TA estão voltadas para duas vertentes: a) uma que propõe fazer uma representação da língua de partida no sentido de convencionar uma “linguagem conceitual universal”, que representaria um modelo padrão no processo de compreensão de um texto em qualquer língua; b) outra que apresenta a opção de trabalhar com pares de línguas, em que se reproduz também uma representação no texto-fonte, a partir de um módulo de análise, porém com o objetivo de uma representação de equivalência no texto-alvo.

Nesse segundo processo, as representações de equivalência por serem linguísticas apresentam melhor entendimento em relação à sintaxe, semântica, gramática e léxico, mostrando melhor entendimento do contexto, porém não busca por uma representação conceitual do texto-fonte. Para a reconstrução no texto-alvo, o mesmo processo é realizado por um módulo de análise especializado na língua de chegada.

A similaridade na estrutura das línguas e o tipo de texto escolhido para TA fazem toda a diferença no resultado, assim essas linhas de pesquisa buscam trabalhar com línguas que apresentem estruturas linguísticas parecidas e textos considerados de menor profundidade, como os técnicos ou científicos, com o objetivo de atingir maior fluência no texto traduzido. As linhas de estudo precisam ser diferentes quando se trata de línguas orientais devido às diferenças entre sistemas.

A mecanização da linguagem pode ser entendida como mais um objeto externo resultado das revoluções tecnolinguísticas que busca resolver problemas. No entanto, ela não funciona de maneira autônoma, precisa da intervenção humana para que os resultados sejam alcançados, resultados esses que dependem também da necessidade humana. E nesse cenário da necessidade de resolver problemas, encontramos a TA como mais um recurso da mecanização da linguagem que busca atender a uma demanda que não seria possível resolver apenas com a tradução humana. Porém, a TA permanece num patamar de estudos diferente das operações da tradução, que são tema dos Estudos da Tradução, visto que o processo tradutório com base na reenunciação continua sendo um grande obstáculo para as ciências da computação.

CAPÍTULO III

PROGRAMAS DE INVESTIGAÇÃO NA HISTÓRIA DA TRADUÇÃO AUTOMÁTICA

O historiógrafo escolhe o ponto de vista do qual pretende contar uma história, seja esse a partir das continuidades, descontinuidades ou ambos. A história da TA que contamos neste trabalho é caracterizada por uma história de descontinuidades, visto que observamos um movimento que não é linear desde o surgimento da TA, apontando evidências de rupturas teóricas e metodológicas.

Para a construção deste estudo historiográfico, adotamos como base metodológica o conceito de programas de investigação proposto por Swiggers (1981), conforme tratado no capítulo I. O objetivo é agrupar os progressos da TA, de acordo com características em comum de cada grupo de pesquisa que resultaram no trabalho para alcançar êxito em direção à tecnologia mais recente, a NMT.

Cada um dos quatro programas de investigação da HL, propostos por Swiggers em 1981, agrupam ideias que compartilham do mesmo conceito sobre visão de língua e dos mesmos procedimentos metodológicos de análise de incidência da língua. Com base nesse modelo, propomos um novo quadro para análise da história da TA a partir de três novos programas de investigação que são: **1) programa de mecanização; 2) programa de matematização; e 3) programa de operacionalização.**

Esses três programas de investigação definidos para a TA agruparão estudos de acordo com os parâmetros internos definidos (ver capítulo I) de visão de tradução e nível linguístico privilegiado naquele programa, assim como os agentes da produção, circulação e recepção de ideias linguísticas definidos nos parâmetros externos (ver capítulo I), conforme recorte histórico que abrange desde o desenvolvimento dos primeiros modelos de TA, até a consolidação do mercado de tradução que propiciou o estabelecimento de ferramentas e a evolução tecnológica das mesmas, associada a tecnologias que permitiram o avanço do setor.

Os programas de investigação da TA tem a característica de representarem o fechamento de um ciclo nas pesquisas para alcançarem um novo patamar no programa seguinte. Esses eventos geralmente representavam rompimento com

paradigmas tidos anteriormente como intransponíveis, mas que se mostraram contestáveis a partir de novas perspectivas científicas e tecnológicas.

Cada um dos programas definidos permitirá associar o trabalho desenvolvido pelos grupos de pesquisa de acordo com sua semelhança em relação aos procedimentos metodológicos nos estudos em TA, independentemente do período em que se sucederam, uma vez que alguns processos seguiram em uso ao mesmo tempo que uma nova fase de pesquisa era implementada e um novo modelo de TA desenvolvido. Essa perspectiva rompe com uma visão linear histórica, exatamente como o conceito de programas de investigação de Swiggers (1981) implica.

Para entendermos os critérios aplicados no agrupamento em programas dos estudos da TA, apresentamos os dois principais eventos que determinamos para demarcar os grandes rompimentos na história da TA:

- O primeiro evento é a publicação do relatório da ALPAC em 1966, que marca uma descontinuidade tanto das pesquisas em TA, como da ideia de que a automatização da tradução seria possível, resultando numa década inteira de atraso e engavetamento de projetos oriundos de altos investimentos.
- O segundo evento é a descoberta de um modelo de TA com base em estatística, a SMT. O desenvolvimento desse modelo representa uma quebra de paradigma com a tradição de pesquisas que se baseavam puramente na aplicação mecânica de regras gramaticais nas duas línguas. Sem a descoberta da SMT não teríamos a NMT.

Os três programas englobarão estudos e grupos de pesquisa relevantes que foram marcados pelos dois eventos e seus desdobramentos.

No programa de mecanização, por exemplo, estarão agrupados os estudos com o uso de dicionários e glossários bilíngues, baseados no conceito da tradução direta de palavra por palavra, frase por frase e regra por regra, antes da publicação do relatório da ALPAC. Esse programa incluirá as ferramentas que foram desenvolvidas para responder a uma demanda específica de tipo de texto e usuário.

Já o programa de matematização incluirá os estudos que sucederam a retomada das pesquisas em TA a partir da década de 1980, após o período de paralização das pesquisas e que resultaram no desenvolvimento do modelo de TA de base estatística. Esse modelo revolucionou o desenvolvimento de ferramentas de TA a partir do conceito estatístico do uso de corpus linguístico, com o objetivo de atender

um mercado de tradução em crescimento que exigia maior qualidade do texto traduzido e eficiência dos programas.

Por último, o programa de operacionalização da TA pretende agrupar os estudos em torno da NMT, que tem sua base conceitual fundada em *deep learning* e IA, resultando numa TA com diferentes possibilidades de integração e que serve de pano de fundo para o avanço de novas áreas de pesquisa.

O agrupamento dos estudos no desenvolvimento da TA permite visualizar o percurso histórico, com o objetivo de realizar uma análise argumentativa e desenvolver uma opinião crítica a respeito da TA, bem como explorar novas áreas de pesquisa e abrangência. Na sequência, apresentamos a análise dos programas de investigação da TA.

1. Programa de mecanização

Um estudo historiográfico precisa buscar a demanda que motivou o evento histórico, e o surgimento da TA foi impulsionado pela necessidade de atender a um volume de tradução que não era possível resolver apenas com tradução humana, assim como trazer agilidade e redução de custos com a automatização da tradução. Motivada por essas necessidades imediatas, a TA começou ser idealizada e construída num cenário acadêmico de financiamento público que atravessou diferentes fases de pesquisa, construindo uma base de conhecimento fundamental para a área computacional.

A necessidade por um volume maior de tradução e pela agilidade no trabalho era basicamente uma motivação política. O cenário da Guerra Fria (1947-1991) exigia vigilância constante dos movimentos de um potencial inimigo que impulsionado pelo avanço da inteligência demandava compreensão do russo e do inglês imediatamente. Assim, as pesquisas em TA tanto no território americano como russo tiveram uma disparada a partir da década de 1950 com grupos de pesquisas emergindo das principais universidades nos dois países. Com isso, os primeiros protótipos de sistemas de TA foram executados em artigos científicos que apresentavam um grande volume impresso e demandavam tradução rápida, por essa razão aceitava-se um texto de menor qualidade, apenas para entendimento geral do tema.

O programa de mecanização agrupa as pesquisas que tentaram mecanizar a gramática, isto é, as tentativas nessa fase estavam voltadas para fazer a máquina

aplicar de maneira mecânica regras gramaticais que são aplicadas naturalmente pelo ser humano na língua natural.

As primeiras arquiteturas desenvolvidas ainda entre as décadas de 1940 e 1950 tinham como base o modelo de tradução palavra por palavra, em que o maior esforço estava na geração de glossários e dicionários bilíngues que fizessem a transferência entre línguas, essas arquiteturas conhecidas como “transformadoras” (cf. Arnold e al., 1994) são consideradas extremamente limitantes se comparadas com arquiteturas de base neural. Eram, porém, consideradas elaboradas para a época, posto que a limitação de recursos computacionais e tecnológicos naquele momento era a realidade.

As arquiteturas transformadoras tinham como método básico “transformar” o texto fonte pela substituição de palavras na língua de partida por seu equivalente na língua de chegada com base num dicionário bilíngue. A reordenação do texto de chegada ficava a cargo apenas de um processo simples de *parsing* que não conseguia ordenar completamente na língua de chegada. Em geral, esse tipo de arquitetura não possuía conteúdo gramatical abrangente da língua de partida porque estava focalizada na língua de chegada, o que resultava em traduções incompletas pela falta de referência na análise do texto de partida para que fosse transformado na língua de chegada.

Esse tipo de arquitetura não apresentava conhecimento linguístico independente sobre a língua de chegada porque não possuía uma gramática independente para aquela língua. Isso significa que a máquina não conseguia diferenciar corretamente regras gramaticais de cada uma das línguas (a de partida e a de chegada) transformando partes da língua de partida em partes da língua de chegada.

Apesar do resultado obtido inicialmente ser totalmente insatisfatório, apresentando baixa qualidade do texto traduzido, a justificativa encontrada para o andamento dos estudos estava no fato de que o novo processo atendia a um dos pilares da demanda principal: agilidade. Visto que a máquina conseguia entregar uma suposta tradução rapidamente, em comparação ao trabalho de um tradutor especialista.

Nessa fase, a principal preocupação dos pesquisadores estava em “decodificar” a língua de partida, a necessidade de entender o inimigo mostrava uma grande barreira: a língua. As máquinas executavam apenas em uma direção e não

conseguiam refazer o caminho contrário ainda, realizavam a transformação de um texto russo para o inglês mas não no sentido contrário, para isso sempre era necessário que uma nova máquina fosse construída. Os recursos computacionais eram poucos e as ferramentas de programação não eram adequadas para o PLN, seguindo então até o começo da década de 1960 com a ideia de uma TA puramente mecânica.

O começo do desenvolvimento de arquiteturas de TA não contou com a ajuda de linguistas visto que as pesquisas estavam concentradas em especialistas das exatas e da emergente computação. O envolvimento de linguistas era simplesmente gramatical, com a função de definir regras gramaticais que a máquina pudesse aplicar mecanicamente. Como consequência, os métodos de pesquisa seguiam uma abordagem de tentativa e erro, com métodos estatísticos para detectar regularidades léxicas e gramaticais ou abordagens muito teóricas envolvendo pesquisas linguísticas como fundamento.

Os extremos eram evidentes no processo e a combinação da falta de noção prática na aplicação de teorias linguísticas complexas demonstravam uma mentalidade pela busca de um texto traduzido perfeito sem a necessidade de uma pós-edição, em outras palavras, uma busca impossível naquele momento.

As décadas de 1950 e 1960 foram os anos de maiores expectativas e também desilusões sobre a TA. Nesse período as pesquisas em TA também serviram para impulsionar vários outros trabalhos contemporâneos em áreas como (cf. Hutchins, 2001) a linguística estrutural e formal, semiótica, semântica lógica, linguística matemática, linguística quantitativa e no que hoje chamamos de linguística computacional e engenharia da língua. Seu período de ascensão foi marcado pela grande importância para diferentes campos no que diz respeito à aplicação de computadores para realização de tarefas "intelectuais" humanas. A demanda pela TA nesse período significou uma das grandes motivações para pesquisas em computação, servindo como fundamento para o desenvolvimento de linguagens de programação.

1.1 Grupos e áreas de pesquisa

Os EUA concentravam a maior parte dos grupos de pesquisa que contribuíram nos estudos da TA nas décadas de 1950 e 1960, deixando seu legado para a próxima geração de pesquisadores na retomada a partir da década de 1980. Esses grupos se

formaram dentro de universidades americanas que contavam com investimento público e privado numa era de importante produção acadêmica e avanço computacional. O Quadro 1 relaciona os principais grupos de pesquisa e sua liderança nesse período, bem como a principal área de estudos de cada um deles:

Quadro 1: Grupos de pesquisa americanos nas décadas de 1950 e 1960

<p>Grupo de Erwin Reiffler na Universidade de Washington (Seattle, EUA): otimizou a abordagem da tradução direta com base em dicionário que envolveu a construção de grandes dicionários bilíngues em que a informação lexical era usada também para resolver problemas gramaticais, sem o uso de análise sintática.</p>
<p>Grupo de Leon Dostert na Universidade de Georgetown (Washington, EUA): conduziu análises de textos apenas quando as informações da gramática tradicional eram inadequadas. Desenvolveu o sistema GAT (<i>Georgetown Automatic Translation</i>) que possuía três níveis de análise: morfológica, sintagmática e sintática. O sistema foi instalado na Itália em 1963 pela Euratom e posteriormente também adotado pela Comissão de Energia Atômica Americana em 1964.</p>
<p>Grupo de Anthony Oettinger na Universidade de Harvard (Massachusetts, EUA): concentrou as pesquisas na compilação de um dicionário Russo-Inglês e em 1959 avançou no desenvolvimento de um sistema para identificação de sequências de categorias gramaticais (pronomes, adjetivos, verbos, etc.) e sua predição probabilística.</p>
<p>Grupo de Victor Yngve no MIT (Massachusetts, EUA): em 1953 concentrava os estudos na relação sintática. A língua de partida era analisada como representações de estruturas de frase, em que uma rotina de transferência de estrutura convertia em estruturas de frases equivalentes na língua de chegada. O principal problema encontrado era a barreira semântica do texto traduzido. Chomsky fez parte desse grupo por pouco tempo mas não se encontra evidências de uma abordagem chomskiana nas pesquisas em TA nesta época.</p>
<p>Grupo de Winfried Lehmann na Universidade do Texas (Austin, EUA): grupo atuando no LRC (<i>Linguistic Research Center</i>) em 1958 que também concentrava os estudos na relação sintática. Tentou construir um sistema que pudesse fazer tradução bidirecional com base na abordagem da transferência sintática, deixando fundamentos para o posterior desenvolvimento do sistema METAL.</p>
<p>Grupo de Sydney Lamb na Universidade da Califórnia (Berkeley, USA): concentrou-se no desenvolvimento de rotinas de dicionários eficientes e de uma teoria linguística adequada para TA. Este conceito seguia a ideia de relações paralelas com a arquitetura de computadores (nós e redes), em que a tradução era vista como um processo de codificação e decodificação.</p>

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

A ideia da automatização da tradução a partir da mecanização de regras gramaticais é a principal linha de pesquisa desse período. Buscava-se programar a máquina para realizar as tarefas de tradução como se fosse uma operação matemática, em que apenas a aplicação da regra certa pudesse transferir um texto de partida para outra língua de chegada sem considerar o que acontece no caminho da transferência.

A análise da língua de partida foi testada de maneiras diferentes, tanto do ponto de vista de uma representação gramatical, em que os elementos da frase eram divididos em sequências morfológicas (pronomes, adjetivos, verbos etc.), como em representações sintáticas (sujeito e predicado). Isto é, buscavam diferentes formas de mecanização sempre com base na mesma linha de pesquisa da aplicação de regras.

Enquanto isso, fora dos EUA outros grupos de pesquisa se formavam, a União Soviética seguia a mesma linha de pesquisa americana, mas a Inglaterra e Itália conseguiram avançar em caminhos um pouco diferentes como podemos observar no Quadro 2.

A linha de pesquisa com base na interlíngua mostrou resultados mais satisfatórios na época do que a linha tradicional de transformação direta com base em regras gramaticais. Os grupos americanos não se aprofundaram nessa área, fato que pode ter influenciado a decepção com os resultados da TA naquele momento, mas que manteve uma chama acesa nas pesquisas conduzidas na Europa no período de paralização das pesquisas americanas.

Os registros demonstram um menor avanço nas pesquisas russas, no entanto podemos supor que os resultados obtidos não fossem divulgados de maneira global, uma vez que o cenário do período representava uma tensão internacional que poderia determinar um controle interno dos dados.

Quadro 2: Grupos de pesquisa fora dos EUA nas décadas de 1950 e 1960

Rússia (Antiga União Soviética): concentravam-se em sistemas inglês-russo e na análise morfológica e sintática para a transformação da língua de partida por meio de regras gramaticais. Apesar de o inglês ser o principal idioma estudado devido a questões políticas entre os países, as pesquisas ampliaram para uma diversidade maior de idiomas do que nos EUA. Os principais grupos de pesquisa foram: o grupo de D. Y. Panov no *Institute of Precision Mechanic* (hoje *Lebedev Institute of Precision Mechanics and Computer Engineering*), o grupo de Ljapunov, Kulagina e Mel'cuk no *Steklov Mathematical Institute* e o grupo de Nikolaj Andreev na Universidade Estadual de Leningrado.

Londres, UK: o grupo de Margaret Masterman na *Language Research Unit* de Cambridge adotou a linha de abordagem "interlíngua" que consiste no método de transformação do texto fonte numa representação abstrata de língua independente (a interlíngua) e, em seguida, o texto de chegada é gerado a partir da interlíngua. No caso das pesquisas de Cambridge, o desenvolvimento do método interlíngua foi baseado em palavra por palavra produzindo traduções denominadas "pidgin".

Milão, Itália: o grupo de Silvio Ceccato na Universidade de Milão também concentrou suas pesquisas no método da interlíngua e o desenvolveu com base em processos cognitivos, especialmente no conceito de análise de palavras e suas possíveis correlações com outras palavras no texto.

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

A importância desses grupos de pesquisa é fundamental na história da TA, conforme Murray (1994) são as elites universitárias responsáveis pela difusão de uma teoria e legitimação no ambiente acadêmico pelo clima de opinião estabelecido naquele cenário. A rede de influências para o estabelecimento da teoria é consolidada pela organização desses grupos que contam com um líder intelectual e liderança organizacional, promovendo a circulação da produção acadêmica e sua recepção.

A divulgação do trabalho desses grupos avançou e em 1954 a primeira revista sobre TA intitulada “Tradução Mecânica” foi lançada por William Locke e Victor Yngve, que sucederam Bar-Hillel no Instituto de Tecnologia de Massachusetts (MIT) a partir de 1953. Nesse ano também foi concluída a primeira tese de doutorado em TA, a do pesquisador Anthony G. Oettinger (Harvard), que estudou um dicionário mecânico em russo. No ano seguinte, em 1955, o primeiro livro sobre TA foi lançado e era uma coleção editada por William Locke e Andrew Booth, incluindo artigos acadêmicos publicados por Yehoshua Bar-Hillel (MIT), Leon Dostert (Georgetown), Anthony Oettinger (Harvard), Erwin Reifler (Universidade de Washington) e Victor Yngve (MIT), além do memorando de Warren Weaver de 1947.

Apesar de várias iniciativas para a divulgação da TA, as pesquisas publicadas ainda eram poucas. Não encontramos trabalhos na área das linguagens nessa época que discutam a TA. A produção era fundamentalmente sobre estudos técnicos, que não demonstravam interesse por parte das humanidades na discussão sobre os impactos sociais da revolução tecnológica que iniciativas como a TA provocariam.

A retórica da rede de influências promovida pelos líderes desses grupos contava com o apoio das universidades atuantes para garantir a circulação do que seriam os futuros grandes resultados promovidos pela TA. Essa idealização gerou grandes expectativas que naquele momento não alcançavam os resultados idealizados. A possibilidade de uma tradução rápida a custo reduzido começou a parecer mais difícil do que se pensava e a TA começou ser desacreditada por alguns pesquisadores, entre eles um de seus idealizadores: Bar-Hillel.

Para enterrar definitivamente a continuidade dos investimentos nas pesquisas em TA, ocorreu em 1966 a publicação do relatório da ALPAC que serviria como validação para o corte brutal no financiamento público e privado do setor nos EUA.

O relatório foi cercado por justificativas inconsistentes, e historiadores afirmam que o parecer teria sido fundamentado apenas nas pesquisas de Bar-Hillel, que passou de defensor para grande cético sobre a viabilidade da TA. O resultado foi o

desmantelamento completo dos grupos de pesquisa americanos com os cortes nos investimentos.

Dentre as principais justificativas que convenceram o abandono das pesquisas estava o custo. A promessa de que a automatização da tradução reduziria o custo, em comparação à tradução realizada por tradutores especialistas, caiu em descrédito a partir da justificativa de que o investimento na construção de sistemas para TA seria muito mais alto.

A solução recomendada era a redução nos serviços de tradução financiados pelo governo de maneira generalizada. O relatório alegava que ocorriam excessos tanto na quantidade de tradutores profissionais oficiais contratados, como no volume de documentos que demandavam tradução, recomendando um corte nos serviços e sugerindo que os cientistas estudassem o idioma russo (que era o mais solicitado naquele momento) para entendimento do conteúdo acadêmico de que precisavam. O relatório divulgava um programa de aprendizagem do idioma russo em seis semanas que prometia excelentes resultados.

O relatório foi criticado ainda na época e já apontava inconsistências nas justificativas para o abandono das pesquisas. Aparentemente, os resultados pouco satisfatórios que relatam o documento da ALPAC partem de uma visão simplista de resultados obtidos com base no conceito de FAHQMT (*Fully Automatic High Quality Machine Translation*). A visão de FAHQMT foi concebida por Bar-Hillel que idealizou uma TA perfeita, em que a necessidade de pré/pós-edição no processo não representaria uma TA.

A retórica da desmoralização da TA e do trabalho desenvolvido até aquele momento marcado pelo relatório da ALPAC contou com o apoio institucional da *National Science Foundation* que era a organização americana responsável pelos departamentos de defesa e agências de inteligência. A recepção do documento foi de grande impacto em virtude do poder de decisão desse órgão, além disso, os EUA passavam pela Guerra do Vietnã (1955-1975) e ainda se preparavam para uma corrida espacial, competindo acirradamente com a União Soviética. Com isso concentraram os investimentos científicos no desenvolvimento bélico e nas ciências espaciais, dando menor importância ao desenvolvimento de outras áreas científicas no país. A ALPAC em si não teria tido tamanha relevância se não fosse o apoio institucional que legitimou sua participação nesse processo.

Todos os grupos de pesquisa americanos que estudavam a TA foram desmontados e seus membros se voltaram para outras áreas de estudo. Os efeitos do relatório também foram percebidos fora do território americano, porém alguns grupos ainda continuaram seu trabalho. Os países europeus e asiáticos que seguiram com suas pesquisas tiveram um desenvolvimento tímido e talvez se as pesquisas americanas não tivessem sido interrompidas poderíamos ter visto um avanço mais rápido na construção da TA em todo o mundo.

Um dos grupos que continuaram ativo e conseguiram evoluir consistentemente foi o TAUM na Universidade de Montreal no Canadá. O grupo continuou os estudos no desenvolvimento do sistema METEO que a partir de 1977 passou a ser adotado pelo Centro de Meteorologia canadense para tradução dos boletins meteorológicos, pois o país tinha que lidar com mais de um idioma oficial. O METEO significou um dos primeiros exemplos de sucesso com a TA, em que após a implantação do sistema os tradutores foram reaproveitados para continuarem se dedicando na implementação de melhorias na ferramenta. Em vez de ficarem traduzindo boletins meteorológicos maçantes e repetitivos, puderam se dedicar a tarefas de tradução mais intelectuais e como resultado observaram uma redução na rotatividade da equipe de tradutores. Na década de 1990, o METEO precisava de menos de 4% de correção e mesmo onde o resultado era de baixa qualidade o custo para correção ainda era menor.

Os grupos fora do território americano conseguiram atingir resultados importantes; o que motivou uma retomada gradativa de grupos de pesquisa nos EUA a partir da década de 1980. O Quadro 3 destaca alguns dos principais grupos fora dos EUA que tiveram grande relevância para a retomada americana e posterior aceleração mundial nas pesquisas a partir da década de 1990.

Quadro 3: Grupos de pesquisa fora dos EUA nas décadas de 1970 e 1980

Grenoble, França: o grupo começa a desenvolver o sistema GETA-Ariane ainda na década de 1960 e seus resultados influenciaram outros projetos ao redor do mundo. As principais características do sistema eram: flexibilidade, modularidade, algoritmos para manipulação de representações em árvore e conceito de gramáticas estáticas e dinâmicas. O sistema não chegou se tornar um sistema operacional e as pesquisas terminaram no final da década de 1980.

Kyoto, Japão: o grupo liderado por Makoto Nagao da Universidade de Kyoto desenvolveu o sistema Mu. Tinha fundamentos parecidos com o conceito do sistema de Grenoble GETA-Ariane, como o uso de análise gramatical de casos e representações em árvore de dependência, bem como desenvolvimento de um ambiente de programação para escrever gramática (GRADE). O projeto influenciou vários outros projetos japoneses de TA. Em 1986 tornou-se um sistema operacional usado no Centro de Informações e Tecnologia do Japão para tradução de *abstracts*.

Saarbrücken, Alemanha: o grupo começou desenvolver em meados da década de 1970 o sistema SUSY que possuía como características principais: regras de estrutura de frase, regras transformacional, quadros de gramática de caso e valência, gramática de dependência, uso de dados estatísticos, etc.

Moscú e Leningrado, Rússia (antiga União Soviética na época): Em 1976 o grupo de Moscú (*All-Union Centre for Translation*) desenvolveu os sistemas AMPAR (inglês-russo) e NERPA (alemão-russo) sob coordenação de Yuri Apres'jan que baseava-se no modelo de "significado-texto".

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

1.2 A retomada do interesse pela TA

A Universidade de Georgetown (EUA) tinha um dos trabalhos mais avançados na área no momento do corte dos investimentos e foi praticamente a única universidade americana que seguiu de maneira independente com pesquisas em TA. Em 1968, pesquisadores da universidade fundam a empresa Systran, que na década seguinte começa comercializar seu sistema de TA (o Systran) que seria adotado pela Comunidade Europeia, representando o maior projeto na área.

O programa já vinha sendo usado pela NASA e pelas Forças Aéreas americanas com um único módulo russo-ínglês e passou a ser desenvolvido nas versões francês-ínglês e italiano-ínglês para atender às demandas de tradução da Comunidade Europeia. Na ocasião, o Systran ainda era desenvolvido a partir do conceito de uma arquitetura transformadora e com base em regras.

A década de 1980 também foi marcada pelo estouro nas pesquisas em tecnologia no Japão e, conseqüentemente, aumentou o interesse desse país no desenvolvimento de sistemas de TA. Essas pesquisas também influenciaram o interesse pela TA de outros países daquele continente como Coreia, Taiwan, China e Malásia.

Nesse ponto, os grupos começaram se interessar pelo processo do pensamento humano (o que futuramente seriam os estudos em IA) no PLN e acreditavam que naquele caminho encontrariam as soluções para os problemas da TA. Esses grupos impulsionaram a consolidação da TA como uma área de pesquisa importante. No entanto, ainda se tratava de um campo de pesquisa complexo, tanto do ponto de vista técnico como de gestão organizacional e de infraestrutura, e caro em relação ao esforço de formação de equipes e tempo para obter resultados.

O mercado de TA teve uma expansão e vários sistemas entraram em funcionamento em instituições públicas e grandes corporações privadas.

Multinacionais investiram no desenvolvimento de seus próprios sistemas de TA e, posteriormente, empresas de *softwares* entraram no mercado oferecendo opções personalizadas também. Alguns desses primeiros sistemas serviriam de base para o desenvolvimento de programas futuros.

Quadro 4: Primeiros sistemas comercializados nos anos de 1980

Sistema	Grupo de pesquisa
GAT	Universidade de Georgetown (USA)
METEO	Universidade de Montreal (Canadá)
TITUS	<i>Institut Textile de France</i> (França)
CULT	Universidade de Hong Kong (China)
Systran	Systran (USA)
Logos	Logos Corporation (USA)
METAL	Universidade do Texas em parceria com a empresa Siemens (USA e Alemanha)

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

Alguns projetos deixaram importantes legados, sendo precursores da retomada dos estudos em TA nos anos de 1980, pavimentando as vias para a grande virada na área a partir da década seguinte. Entre eles está o grupo do GETA-Genoble (França) que trabalhou no projeto EUROTRA para traduções demandadas pela Comunidade Europeia e que, apesar de o projeto ter terminado no final dos anos de 1980, conseguiu atingir seu objetivo de estimular a colaboração em pesquisa entre nações na área da linguística computacional.

2. Programa de matematização

A década de 1980 tem destaque para as pesquisas na área de tecnologia no Japão e o interesse pela tradução da língua falada também desponta nesse país, enfrentando o desafio de combinar o reconhecimento de voz e a interpretação linguística de uma conversa. O Japão tinha um interesse particular em abrir seu mercado para o mundo e a comunicação enfrentaria uma barreira linguística devido à grande diferença no sistema da língua japonesa, com isso as pesquisas para automatização da tradução entram na corrida.

O grupo liderado por Makoto Nagao da Universidade de Kyoto (Japão) em 1981 desenvolve o modelo conhecido como EBMT (*Example-based Machine Translation*, Tradução Automática com base em Exemplo) ou “*memory-based*” (com base em memória) que partia do pressuposto que uma tradução geralmente envolve a busca

por exemplos análogos, ou seja, procurar como uma expressão ou frase foi traduzida anteriormente. Com isso o método EBMT estaria fundado no processo de extração e seleção de frases ou grupos de palavras equivalentes a partir de um corpus bilíngue. O cálculo das correspondências poderia ser feito com base em métodos semânticos ou informações estatísticas sobre frequências lexicais na língua de chegada. Nesse evento observamos os primeiros passos em direção ao desenvolvimento de sistemas de TA com base estatística.

No Canadá, tentativas para o tratamento probabilístico com base em corpus também progredia. Neste caso o processo consistia no alinhamento de frases, grupos de palavras e palavras individuais a partir de textos extraídos de relatórios dos debates de parlamentares canadenses que eram alinhados paralelamente entre dois idiomas. O objetivo era calcular a probabilidade de que qualquer palavra ou quaisquer palavras na sentença traduzida estaria(m) alinhada(s) com a língua de partida.

O primeiro sistema com base estatística é desenvolvido por um grupo da empresa IBM de Nova Iorque nos EUA no final da década de 1980. O grupo tinha desenvolvido anteriormente técnicas para automatizar a transcrição de voz e com base nesse conceito iniciou experimentos de modelos para sistemas de TA. Os pesquisadores partiram da ideia de que há várias possibilidades de tradução para uma mesma sentença, independente da língua em questão, e essas possibilidades passam por uma decisão pessoal do tradutor. Com base em tal hipótese, desenvolveram modelos a partir de teorias probabilísticas baseadas no teorema de Bayes, no qual o sistema seria capaz de escolher entre as sentenças na língua de chegada a que teria maior probabilidade de representar uma tradução da língua de partida.

Para isso o grupo desenvolveu uma fórmula matemática que calculasse de maneira estatística qual combinação de sequência de palavras na língua de chegada era mais provável corresponder à combinação de sequência de palavras na língua de partida. Essa fórmula matemática ficou conhecida como a “equação da TA” (POIBEAU, 2017, p. 84) porque desde então serve como base para o desenvolvimento de sistemas SMT.

O modelo baseia-se numa estratégia de alinhamento lexical que resultaria na escolha das palavras que compõem a tradução na língua de chegada. Esse processo consiste em duas etapas: a) na primeira, trabalha-se com um grande corpus bilíngue no qual é aplicado um algoritmo que tem o objetivo de formar um dicionário bilíngue com o mais provável alinhamento de palavras num nível de sentenças; essa etapa

conhecida como “treinamento” ou “aprendizagem” é o momento que a máquina “aprende” a partir do corpus que lhe é fornecido. b) na segunda etapa, conhecida como “processamento” ou “teste” é a fase de tradução em si, onde o sistema usa as novas sentenças geradas na fase de aprendizagem para traduzir as novas sentenças solicitadas.

As operações de TA no modelo palavra por palavra na década de 1990 eram vistas como simplista. O processo de tradução precisava envolver conhecimento de vocabulário (nas duas línguas), de gramática (o sistema de regras) e do elemento relacionado ao conhecimento de mundo real, do conhecimento sobre a natureza das coisas e da maneira como funcionam. O reconhecimento de que o texto-fonte precisava ser adaptado ou preparado para TA também entra nas diretrizes, além da aceitação de que o texto traduzido precisava de pós-edição feita por um tradutor especialista.

A SMT é fundamentalmente baseada no treinamento da máquina em um grande corpus bilíngue. O sistema primeiro usa esse corpus que está alinhado no nível de sentença para aprender sobre as traduções de palavras e seus possíveis alinhamentos no nível de sentença. Esse conhecimento adquirido pela máquina no processo de extração do corpus é empregado para a tradução das novas frases, conhecido como processo de “decodificação”.

O processo de decodificação requer um módulo de aplicação de regras da língua de chegada, isto é necessário para que o sistema consiga considerar também traduções que não são necessariamente baseadas nas equivalências mais prováveis no nível da palavra, visto que a tradução de uma sentença feita com base somente no modelo probabilístico pode resultar num baixo resultado de probabilidades como um todo.

O grupo da IBM desenvolveu alguns modelos de alinhamento que são conhecidos como os “modelos da IBM” e representaram uma grande virada no desenvolvimento de sistemas SMT, servindo de fundamento até hoje. O grupo teve grande relevância na consolidação de modelos de desenvolvimento de sistemas com base estatística onde esses modelos apresentam fundamento nos estudos do professor Weaver na década de 1940, em relação ao conceito de “decodificação” da língua-fonte para traduzi-la na língua-alvo.

O fato relevante para o progresso das pesquisas em TA, em especial, nos sistemas SMT é o surgimento de métodos com base em corpus. Essa descoberta abre

caminhos para inúmeras outras áreas dentro dos Estudos da Tradução, da Linguística Computacional, entre outras, possibilitando o desenvolvimento de novas disciplinas, como a Linguística de Corpus.

O estudo de métodos com base em corpus propiciou também o desenvolvimento da TM (*translation memory*, memória de tradução) que deu origem às *CAT tools* transformando o trabalho do tradutor a partir dos anos de 1990. Abriu-se então acesso à um mercado para tradução comercial que foi posicionada como um negócio em desenvolvimento.

As *CAT tools* são ferramentas que fazem a gestão do texto com base na TM que alinha o texto-fonte ao lado do texto-alvo no nível de sentença com o objetivo de facilitar o trabalho de tradução. Essas ferramentas também podem ser integradas ou trazer no seu módulo nativo recursos como dicionários bilíngues, glossários, corretor ortográfico, tradução automática, entre outros recursos.

Quadro 5: Principais *CAT tools* comercializadas na década de 1990

Sistema	Empresa
Trados	TRADOS (Alemanha), adquirida pela SDL (USA) em 2005 e hoje SDL Trados
Transit	Star AG (Suíça) hoje Star Group
TranslationManager	IBM (EUA)
LANT	Eurolang Optmizer (Bélgica)
Déjà Vu	Atril (França)
EURAMIS	Desenvolvido pela própria União Européia para integrar com o Systran

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

O desenvolvimento da SMT foi um pontapé para o avanço dos recursos disponíveis para facilitar o trabalho do tradutor nessa época. As *CAT Tools* surgem na mesma fase que os sistemas SMT começam dar bons resultados e o recurso de memória de tradução tem suas bases fundadas no conceito de desenvolvimento de TA. Suas técnicas ainda continuam andando juntas.

Tudo isso impulsionado pela melhoria nas tecnologias computacionais que possibilitaram termos computadores mais eficientes e a custo acessível, motivam o desenvolvimento de programas e recursos para facilitar mais o trabalho de tradução, estimulando um mercado em ascensão e o surgimento de cursos profissionalizantes e de nível superior no Brasil.

O mercado começava aceitar a TA e os recursos para automatização da tradução, diversos sistemas passam ser comercializados e a consolidação da adoção das ferramentas se dá por serviços do governo, empresas multinacionais, entre outros, onde traduções são realizadas em larga escala, principalmente de material técnico. Com o aquecimento do mercado grandes corporações começam adotar os sistemas de TA e implantá-los em seus *mainframes*. Com uma demanda crescente as empresas desenvolvedoras de sistemas de TA também incluem o serviço de tradução como unidade de negócio em seu modelo e começam oferecer os serviços para a fatia do mercado que não queira adquirir a ferramenta. Grandes projetos de localização de *softwares* começam ser desenvolvidos e o mercado de tradução técnica consolida-se definitivamente.

Nos anos de 1990, a demanda pelos serviços de tradução continuou crescendo mas num mercado ainda dominado por alguns poucos *players*, assim como geralmente acontece no mercado computacional. A adoção da TA torna-se cada vez mais indispensável em grandes corporações, despertando o interesse também no domínio da internet. A primeira ferramenta a ser disponibilizada de maneira gratuita para traduzir páginas da internet é o Babelfish oferecido pelo primeiro navegador da internet, Altavista (hoje Yahoo).

Essa ferramenta possibilitava que conversas nas salas de bate-papo pudessem ser traduzidas instantaneamente, bem como o conteúdo que passava ser produzido pelos usuários da internet naquele momento. Nesta fase as soluções para TA passam ser pensadas para atender um tipo de texto não mais técnico e sim de linguagem coloquial.

A partir desse momento os recursos de TA se tornam populares e acessíveis para usuários além do ambiente acadêmico e corporativo, neste ponto os usuários da internet também se beneficiam da acessibilidade que a TA poderia oferecer. O novo usuário agora está em busca da facilidade que tais recursos podem proporcionar num cenário de globalização da comunicação, em que a informação poderia navegar o mundo em poucos minutos.

A primeira ferramenta SMT comercializada foi o *Language Weaver* da empresa SDL no ano de 2002. A empresa ainda hoje é reconhecida e domina o setor de *CAT Tools* e TA. Como observamos, apesar dos estudos com base estatística terem seu auge na década de 1990, é necessário tempo para consolidação de uma tecnologia e lançamento no mercado. Apesar da primeira ferramenta SMT entrar no mercado no

começo dos anos 2000, não observamos uma grande corrida da concorrência para posicionar suas ferramentas nessa mesma tecnologia.

O mercado se posiciona diante de um consumidor de internet que consome TA, e surge uma primeira onda de ferramentas de TA para consumidores usuários. As empresas desenvolvedoras começam investir não apenas numa interface corporativa, mas também em produtos para o usuário final. O mercado não supera e o espaço vem ser ocupado por uma empresa concorrente em motor de busca na internet que despontava naquele momento. Assim, a Google em 2006 lança o Google Tradutor de forma gratuita, com poucos pares de idioma no começo, mas que evoluiu rapidamente para vários idiomas, concorrendo diretamente com Babelfish. Este evento consolida definitivamente a TA como um recurso acessível e democrático.

2.1 Grupos e áreas de pesquisa

Os grupos de pesquisa que motivaram o interesse pela retomada dos Estados Unidos nos estudos em TA estavam concentrados no Japão e em alguns países da Europa que tinham interesses particulares no avanço de sistemas de TA.

No Japão, o interesse era em melhorar a comunicação com países europeus e os Estados Unidos devido suas negociações comerciais, isso motivou o desenvolvimento de tradutores portáteis com o objetivo de auxiliar a tradução de reuniões presenciais. Visionavam também um recurso que pudesse ser usado em conferências telefônicas internacionais para facilitar a comunicação em virtude dessa demanda de comercialização com países no exterior. Enfim, outra demanda iminente era pela automatização do serviço de reservas de hotéis que na maioria ainda precisava do contato telefônico ou correio eletrônico para efetivação de uma reserva.

O Quadro 6 relaciona os grupos de pesquisa japoneses que criaram esses recursos e avançaram rumo ao desenvolvimento de sistemas de TA, posicionando o país de maneira competitiva nos estudos dessa área.

Quadro 6: Grupos de pesquisa no Japão na década de 1990

<p>Nara (próximo à Osaka), Japão: o grupo da <i>ATR Interpreting Telecommunications Research Laboratories</i> concentrou-se nos estudos para desenvolver um sistema para registro telefônico que atendessem à demanda para automatização de conferências telefônicas, com o desafio de encontrar uma solução para combinar o reconhecimento de voz e a interpretação linguística de uma conversa.</p>
--

<p>Kyoto, Japão: o grupo liderado por Makoto Nagao da Universidade de Kyoto desenvolveu o conceito “<i>example-based</i>” ainda na década de 1980 que fundamentou sistemas com base estatística e manteve-se ativo na década seguinte.</p>

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

A Comunidade Europeia representou um dos principais incentivos nas pesquisas em TA na Europa, em virtude da necessidade de tradução de todas as línguas dos países pertencentes à comunidade, motivando também áreas de estudos diferentes que buscavam atender à necessidade de sistemas que suportassem pares de idiomas diversos.

O projeto Eurotra foi o maior programa de pesquisa em TA na região entre as décadas de 1980 e 1990 e envolvia, inicialmente, sete idiomas (dinamarquês, holandês, inglês, francês, alemão, grego e italiano) e, posteriormente, incluindo outros dois (português e espanhol) no decorrer de seu desenvolvimento, resultando numa combinação de 72 pares de idiomas. O objetivo da comissão era desenvolver um programa multilíngue para tradução de documentos da Comunidade Europeia, em especial documentos técnicos.

O Eurotra possibilitou que outros sistemas fossem desenvolvidos com base em seu escopo e conceito. O projeto enfrentava um desafio em relação à descentralização dos centros de pesquisa, ao mesmo tempo que trazia a inovação por ser um dos primeiros sistemas de TA multilíngue.

Apesar do programa envolver nove países, nem todos investiram nas pesquisas de maneira significativa e no Quadro 7 temos destacado os principais grupos que influenciaram o projeto e ainda se mantiveram ativos mesmo depois do fim do programa em meados dos anos 1990.

Quadro 7: Grupos de pesquisa de destaque no Eurotra até a década de 1990

<p>Grenoble, França: o grupo GETA e o grupo IAI (Alemanha) serviram de base para a criação do projeto Eurotra, a partir de seus estudos no desenvolvimento dos sistemas GETA-Ariane e Susy, respectivamente.</p>

<p>Saarbrücken, Alemanha: o grupo IAI da Universidade de Saarlandes desenvolveu o sistema SUSY na década de 1970, mas a partir do projeto Eurotra criou também o sistema CAT2.</p>

<p>Copenhague, Dinamarca: o grupo CST da Universidade de Copenhague participou do programa até início dos anos 1990 e manteve-se ativo ainda depois, desenvolvendo o sistema Patrans para tradução do dinamarquês/inglês com base nos fundamentos do Eurotra.</p>
--

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

Com os trabalhos desses grupos servindo de base para estudos americanos, os Estados Unidos começa uma retomada gradual pelo interesse no desenvolvimento de estudos em TA em seus centros de pesquisas na década de 1980, resultando no lançamento de alguns sistemas logo na década seguinte. Essa retomada faz surgir nos anos de 1990 novos grupos de pesquisa, como podemos observar no Quadro 8 que em parceria com empresas privadas conseguiram avançar com novos projetos de TA.

Quadro 8: Novos grupos de pesquisa nos EUA na década de 1990

<p>Pittsburgh, Pensilvânia: o grupo do Instituto de Tecnologias da Linguagem da Universidade de Carnegie Mellon (CMU) desenvolveu em colaboração com a empresa Caterpillar o sistema CATALYST em 1992. O sistema com base no modelo de conhecimento (<i>knowledge-based</i>) e seguindo o conceito de linguagem controlada foi adotado pela Caterpillar na tradução de manuais técnicos em vários idiomas. O grupo no final da década de 1990 também desenvolveu o sistema chamado DIPLOMAT para traduzir do idioma sérvio para o croata, especificamente voltado para operações militares em que o país estivesse envolvido.</p>
<p>Las Cruces, Novo México: o grupo do laboratório de pesquisas em computação da Universidade do Estado do Novo México (NMSU) em meados dos anos de 1990 lança o sistema ULTRA com a proposta de ser um programa multilíngue traduzindo cinco idiomas (chinês, inglês, alemão, japonês e espanhol).</p>
<p>Advanced Research Project Agency, ARPA: o grupo financiado pela ARPA e formado por pesquisadores dos centros de pesquisa da Universidade do Estado do Novo México, da Universidade de Carnegie Mellon e da Universidade da Califórnia (USC) desenvolve em colaboração o sistema Pangloss no modelo com base em conhecimento e proposta de ser um programa multilíngue.</p>

Fonte: elaboração própria a partir de dados de HUTCHINS, W. J. *Machine Translation over fifty years. Histoire Épistémologie Langage*, v. 23, n. 1, p. 7-31, 2001. (2020)

Observamos na década de 1990 a abertura da indústria da tradução que irá promover a comercialização de vários novos recursos para automatização da tradução, resultando numa movimentação dos centros de pesquisas, que até então se concentrava predominantemente nos centros universitários, para dentro de grandes corporações que passam a desenvolver sistemas com fins comerciais impulsionando o crescimento da indústria de *software*.

2.2 As limitações do modelo de base estatística

Apesar do notório avanço que observamos desde a idealização dos primeiros protótipos para TA, os sistemas com base estatística também enfrentam suas limitações como percebemos em todo percurso das pesquisas no desenvolvimento de sistemas de TA. É uma área de pesquisa difícil, que demanda bastante investimento

e tempo para enxergar resultados, por isso os sistemas SMT demoraram para serem incorporados pelos desenvolvedores. A mudança no modelo de desenvolvimento precisaria de tempo, era necessário adquirir conhecimento investindo em pesquisas para obter melhores resultados do que com os sistemas em funcionamento.

O processo de transição de uma tecnologia para outra precisa de tempo e investimento que geralmente é considerado por grandes corporações com interesse mercadológico num produto. Ainda na virada no milênio a maior parte dos sistemas contava com uma arquitetura transformadora com base em regras, ou seja, com fundamento na tradução direta, não apontando muito investimento e entusiasmo na SMT.

A arquitetura com base em regras oferece várias possibilidades de experimentos, paralelamente às pesquisas sobre a SMT estavam os testes de arquiteturas que incluíssem o conhecimento linguístico nos dois idiomas. Apesar de tratar-se ainda de um conceito totalmente fundamentado em regras, os resultados obtidos com esse tipo de técnica mostrou-se de mais qualidade do que as arquiteturas transformadoras. Neste caso as regras gramaticais eram aplicadas para ambos os idiomas; o processo compreendia primeiro numa análise do texto-fonte com base na gramática da língua de partida que resultaria numa tradução com regras aplicadas segundo a língua de chegada. O resultado eram sentenças perfeitamente construídas, mas nem sempre produzindo traduções corretas, isto porque faltava “fluência” da máquina que aplicava as regras de maneira mecânica.

Este problema da fluência de certa forma é resolvido pela SMT. A técnica de treinamento da máquina em um corpus traduzido por um humano possibilita o tratamento estatístico de um material humano que é o motivo da diferença na qualidade da tradução com base estatística. A máquina não estaria criando suas próprias traduções como acontece na arquitetura com base em regras, na verdade, o que ocorre é a réplica de um conteúdo produzido por humano, que pensou linguisticamente. A SMT não tem o objetivo de robotizar a tradução, ela reproduz um material construído pela mente humana.

Assim, se a máquina trabalhar com um corpus que tenha muitos problemas linguísticos, conseqüentemente, encontramos problemas na tradução. Além do fato que outro fator limitante é o tamanho do corpus. Existe um certo “mantra” dentro das técnicas de automatização da tradução com base em corpus que diz “quanto maior, melhores os resultados”, esse é um tema que gera discordância entre pesquisadores,

no entanto, é consenso que mais que o tamanho do corpus, a qualidade do texto é sobretudo importante. A questão do grande volume de palavras e sentenças de um corpus faz diferença na quantidade de possibilidades que a máquina pode analisar para uma mesma tradução, gerando mais opções de combinações no momento da tradução.

Os anos de 1990 possibilitaram o desenvolvimento de técnicas com base em corpus devido ao grande volume de textos que tornam-se disponíveis em formato que poderia ser entendido pelo computador, o avanço na tecnologia computacional permitiu a produção desse grande volume digital.

Como o corpus é a base para a SMT algumas limitações precisam ser consideradas: a) apesar das discussões em torno do tamanho do corpus, a quantidade de informações de um corpus faz muita diferença na fase de treinamento da máquina. É necessário um grande volume de texto para atingir um desempenho razoável de aprendizagem da máquina, portanto, certos idiomas não conseguem usar SMT caso tenham um volume de dados pequeno para ser treinado; b) como já mencionado, o volume de dados do corpus é fundamental no processo de aprendizagem da máquina e textos com problemas de alinhamento, baixa qualidade linguística, entre outros problemas, resultarão em traduções de baixa qualidade; c) não é possível solucionar todos os problemas da TA apenas com o uso de corpus robustos e de boa qualidade linguística, existem fenômenos raros e complexos de uma língua que apesar do treinamento em corpus de qualidade a máquina não consegue aprender.

O alinhamento do texto-fonte com o texto-alvo em um corpus bilíngue é um assunto que precisa de esclarecimentos. Um alinhamento de textos bilíngues com base em frases é mais fácil e eficiente de se fazer, do que um alinhamento com base em palavras, isto porque em certos idiomas uma palavra pode ter não apenas uma correspondência na outra língua, mas seu significado corresponder a duas ou mais palavras no outro idioma. No alinhamento com base em frases não importa quantas palavras existem no texto-alvo e nem quantas corresponderão no texto-fonte.

Outra característica do alinhamento está na proximidade entre os idiomas, o alinhamento com base em frase funciona melhor quando a língua de partida e de chegada tem uma certa proximidade na estrutura linguística, além de facilitar o alinhamento possibilita melhores resultados probabilísticos na fase de treinamento da

máquina. Enquanto que sistemas linguísticos muito diferentes podem resultar em traduções inconsistentes do ponto de vista linguístico.

Este ponto determina outra ruptura em relação ao padrão de pesquisa aplicado aos sistemas com base em regras. Naquela arquitetura o conceito de tradução direta com base em dicionários partia da abordagem de tradução palavra por palavra, ou seja, a máquina era programada para fazer a transferência de uma língua para outra no nível da palavra e não da frase.

A SMT representa a descontinuidade e negação de teorias anteriores aplicadas ao desenvolvimento de sistemas de TA que eram vistas como pioneiras. No entanto, introduz problemas novos que não eram observados no modelo anterior e o principal deles é não sabermos a origem de alguns problemas de tradução da SMT. Isto quer dizer que não é possível rastreamos e pontuar a origem de um problema na tradução, visto que o processo depende puramente do desempenho estatístico da máquina. Obviamente podemos tentar corrigir o corpus ou aplicar certas regras gramaticais, no entanto, não temos garantia que solucionamos o problema porque ainda assim dependemos do desempenho da máquina. Os modelos SMT são sistemas de tradução direta, eles buscam equivalências entre a língua de partida e a língua de chegada, esse processo não necessariamente é coerente linguisticamente porque o método é matemático e não semântico. Com isso precisamos de soluções híbridas que tragam módulos com base em regras linguísticas, estatística e neural.

Como já comentamos, uma tecnologia nova precisa de tempo para ser implantada e seus resultados demoram para serem vistos. Porém, essa é uma realidade desse milênio onde temos rapidez na descoberta de novas tecnologias, isso por conta do avanço computacional que presenciamos desde as décadas de 1930-1940, quando são criadas as primeiras máquinas que só depois viriam a ser algo mais próximo dos computadores que temos hoje. Lidamos com essa troca de tecnologias que nos parece imediata há poucos anos, a estratégia da maioria das empresas desenvolvedoras é integrar módulos novos aos sistemas em funcionamento, como método para migração para o novo modelo.

Nas ferramentas de TA ocorre esse processo e a estratégia de desenvolvimento de soluções híbridas torna-se a principal estratégia para empresas desenvolvedoras desse tipo de ferramenta. Os sistemas com base em regras foram combinados com sistemas SMT e seguem nesse modelo. A popularização da TA

exigiu que fossem desenvolvidos produtos de acordo com públicos-alvo, surgindo ferramentas para usuários, tradutores, grandes corporações, pequenos projetos, etc.

Apesar da base fundamental que cada ferramenta tem os setores que podem implantá-la são muitos e precisam de personalização adequada, portanto não temos mais um modelo engessado que atenda apenas um determinado tipo de demanda, mas observamos modelos que podem ser adaptados para atender diferentes demandas.

A maioria dos sistemas de TA hoje incorporaram a SMT. A maneira como as informações estatísticas podem ser integradas é de ordem infinita, as empresas podem adaptar seus sistemas das mais diferentes maneiras possíveis. A maior parte das ferramentas híbridas possuem um módulo com base em regras e outro em SMT, porém com o avanço da NMT algumas agora também trazem esse módulo. O Google Tradutor é um exemplo de ferramenta híbrida e que já aplica o módulo NMT.

O modelo com base em regras ainda precisa ser utilizado para idiomas em que o volume disponível de corpus é baixo ou situações em que a SMT não funciona bem, por isso a busca por ferramentas híbridas é a solução encontrada pelas empresas e estratégia que certamente continuará como solução. O que poderemos esperar a partir de agora são novos módulos ainda mais eficientes que os atuais.

As ferramentas híbridas torna-se essencial na tradução de pares de idiomas raros ou de sistemas diferentes. Nesses casos é necessário integrar o módulo estatístico a módulos linguísticos avançados que permitam o treinamento da máquina e a aplicação de regras para correção de possíveis desvios. Isso mostra o avanço alcançado na flexibilização das ferramentas em comparação à limitação que observávamos há alguns anos.

O programa de matematização contempla os eventos relacionados a um período que rompeu com um pensamento tradicional de tratar o processo tradutório como uma tarefa mecânica para aceitar que a máquina não faz o mesmo que um humano faria sozinha. A máquina precisa da inteligência humana para realizar tarefas. Resumidamente, no campo da tradução a máquina precisa de conteúdo humano para realizar uma tarefa automática.

3. Programa de operacionalização

Este programa discutirá as pesquisas que buscam soluções de TA com base em RNA, IA e *deep learning*. Enquanto os outros programas lidavam com a TA do

ponto de vista puramente sistematizado, neste programa observamos uma resposta personalizada às demandas de tradução.

A descoberta e avanço da computação muda completamente a sistematização do trabalho, de uma maneira que as demandas tinham de ser adaptadas ao novo modelo sistêmico. Hoje, as tecnologias com base no conceito de IA buscam adaptar os recursos às necessidades desse modelo já estabelecido. Assim, observamos a personalização de recursos com o objetivo de atender demandas e isso só é possível porque as tecnologias existentes permitem uma flexibilidade de adaptação antes inalcançável.

O conceito de IA está fundamentado no processo de automatização de tarefas intelectuais realizadas pelo ser humano, isto quer dizer que a pura mecanização de atividades humanas não atende mais as necessidades da indústria. O TALN é o primeiro passo nessa jornada de progresso tecnológico e dentre todas as técnicas já experimentadas (algumas discutidas neste trabalho) chegamos ao *deep learning*. O termo pode ser traduzido como “aprendizagem profunda”, mas adotamos o inglês nesta dissertação por se tratar da forma ainda mais divulgada no campo das Ciências da Computação.

Deep learning é um campo de estudos dentro das Ciências da Computação que busca definir métodos de aprendizagem da máquina com base em modelos de RNA. Essa área abre infinitas possibilidades de estudos em uma grande diversidade de aplicações, se tornando uma das linhas de pesquisa na TA atualmente. Esse método demonstra bastante sucesso em áreas como jogos eletrônicos e na interação automática com o ser humano. No entanto, devido às complexidades das línguas naturais e seu processamento, a maneira como é adotada a *deep learning* na TA precisa ser um pouco diferente e os resultados podem demorar mais para serem notados.

As RNAs são modelos computacionais inspirados no funcionamento do cérebro humano, o conceito surgiu na década de 1940 e vem servindo de base em pesquisas no desenvolvimento de técnicas de IA ao longo das décadas, onde exerce um papel fundamental no avanço da área. Esses modelos, assim como o cérebro, são supostamente capazes de construir conceitos complexos a partir de pedaços diferentes de informações que são montados de maneira hierárquica.

Os métodos de *deep learning* aplicados a técnicas de automatização das linguagens humanas enfrentam dificuldades em relação às ambiguidades de

significados de uma palavra, uma vez que não há fronteiras claras entre o sentido e suas relações. Como o sentido muda de uma língua para outra e a noção de significado de uma palavra pode ser expressada por uma única palavra ou um grupo de palavras na outra língua, é impossível especificar manualmente todas as informações que seriam necessárias para um sistema de TA ter um desempenho igual a um tradutor especializado. Isso posiciona a área como extremamente desafiadora exigindo grandes investimentos e tempo para assistir pequenos resultados.

Os primeiros resultados bem sucedidos na adoção de métodos de *deep learning* foram observados nos anos 2000 no reconhecimento de imagens. Geralmente, o método consiste numa operação a partir de um grande número de exemplos que busca extrair de maneira automatizada as características mais relevantes de uma imagem. O aprendizado da máquina ocorre de maneira hierárquica em camadas, primeiro extraíndo os elementos básicos de uma imagem, depois evoluindo para estruturas mais complexas até obter uma análise geral do objeto analisado. Esse processo é similar ao cérebro humano que analisa primeiro um grupo de itens simples, de maneira rápida, para identificar características gerais de uma imagem e, em seguida, reconhece as formas complexas que podem chegar numa representação complexa partindo apenas de informações parciais.

O conceito aplicado a TA busca construir sistemas em que poucos elementos sejam especificados manualmente, a ideia é permitir que o sistema faça deduções por ele mesmo sobre a melhor representação a partir dos dados extraídos de corpus. Essa ideia é fundamentada no modelo de desenvolvimento estatístico para TA em que a máquina passa pela fase de aprendizagem. Com a *deep learning* torna-se possível, ao menos na teoria, que a máquina aprenda características complexas de maneira totalmente autônoma e gradual a partir desses dados extraídos, sem nenhum prévio esforço humano.

A NMT, hoje também chamada de *deep learning machine translation*, segue basicamente a mesma metodologia dos sistemas de SMT. O processo inclui um “codificador” que faz parte do sistema e analisa o corpus na fase de aprendizagem e o “decodificador” processa a tradução automaticamente com base nos dados analisados pelo codificador. A principal diferença da NMT em relação à SMT é que tanto o codificador como o decodificador são fundamentados unicamente em uma RNA. No caso da SMT esses mecanismos tem base em módulos mais tradicionais, como um módulo de língua e um módulo de tradução, usando diferentes estratégias

para otimizar o processo e obter melhores resultados. Assim, podemos definir a NMT como uma abordagem estatística da tradução, porém com base puramente em redes neurais.

É possível dizer que a NMT segue uma arquitetura mais tradicional do que a SMT, pois o codificador pode ser entendido como um analisador da língua de partida e o decodificador um gerador de língua de chegada, o que caracteriza o modelo como dois módulos que desempenham de maneira independente.

O desenvolvimento das arquiteturas de TA com base neural demonstram um avanço em direção à solução de problemas que não conseguiam ser ultrapassados com a RBMT nem com a SMT. A flexibilidade dos sistemas de base neural possibilita novos experimentos na busca por uma tradução da máquina com resultados mais parecidos com uma tradução humana. O que não significa necessariamente uma tradução de maior qualidade linguística, mas sim de maior coerência cognitiva.

Os modelos de base neural possibilitam o gerenciamento simultâneo de vários tipos de informações, permitindo que a máquina faça escolhas mais precisas. Eles possuem a característica de análise de cada elemento numa perspectiva multidimensional, onde uma palavra ou sentença é selecionada dentro de um cenário mais detalhado porque são agrupadas de acordo com o contexto que estão inseridas.

O sistema NMT identifica as palavras que aparecem em contextos interpretados como similares e as agrupa em conjuntos de “palavras integradas”. Esse processo soluciona o problema de tradução de palavras raras, uma vez que o sistema consegue fazer a associação com outros contextos de mesma similaridade e que podem indicar uma potencial tradução. Além disso, também melhora o desempenho do sistema em relação a palavras com vários significados visto que uma palavra pode pertencer a distintos conjuntos de palavras integradas, refletindo os diferentes contextos em uso que a máquina pode escolher.

Esse processo possibilita que unidades linguísticas maiores como frases, sentenças ou grupos de palavras possam ser comparadas num cenário contínuo o que resulta numa abordagem flexível e capaz de identificar inclusive paráfrases. O método de *deep learning* possibilita tratar esse processo de maneira generalizada em comparação a tradicional abordagem com base em dicionário, pois busca o agrupamento de acordo com o contexto, classificando similaridades numa perspectiva cognitiva de maneira mais fiel.

As palavras que possuem uma relação próxima dentro de uma sentença também são agrupadas durante a fase de aprendizagem da máquina. Isso denota a característica hierárquica de sistemas NMT que tem a capacidade de descobrir estruturas, ou seja, identificar a relação de palavras ou grupos de palavras dentro de uma sentença com base nas regularidades constatadas nos milhares de exemplos oferecidos ao sistema na fase de aprendizagem. Embora o processo de *deep learning* não contemple um módulo para codificação da sintaxe, é possível observar sua capacidade em identificar automaticamente relações sintáticas.

3.1 Grupos e áreas de pesquisa

A preparação para a virada do milênio fez com que grandes corporações investissem massivamente em pesquisas para adaptação de seus sistemas às possíveis limitações técnicas que poderiam enfrentar. Esse evento impulsionou a contratação de pesquisadores no campo da computação que provocou uma descentralização dos centros de pesquisa, antes majoritariamente dentro do ambiente acadêmico.

Uma das grandes corporações privadas - a IBM – ocupa posição de destaque no investimento em pesquisas e sistemas de TA, concentrando importantes núcleos de pesquisa nessa área. Além disso, a empresa Microsoft impulsiona a indústria da tradução técnica nos anos de 1990 com grandes projetos de tradução em virtude da localização (adaptação de partes de um *software* para atender limitações técnicas de um país) de seu sistema operacional adotado nos computadores do mundo inteiro.

O resultado da grande abertura do mercado de tradução foi uma fragmentação de tipos de serviços e soluções para automatização da tradução. A demanda por tradução abrange uma infinidade de setores que podem ou não se beneficiar da TA, mas que certamente encontram vantagens em outros recursos de automatização da tradução. Com isso, a fragmentação dos estudos e grupos de pesquisa em TA também é evidente.

Esse novo mercado abriu a possibilidade de muitas frentes de pesquisa diferentes em TA, motivando o desenvolvimento de pesquisas tanto no ambiente universitário como em organizações privadas e de fins não-governamentais que conseguem criar sistemas de maneira infinitamente mais rápida do que seus antecessores, isso devido à flexibilidade dos modelos atuais. O Quadro 9 destaca as principais empresas americanas que se consolidaram no mercado de TA como

solução, investindo na indústria da tradução para atender diferentes demandas desse novo mercado.

Quadro 9: Empresas norte americanas de destaque no mercado de TA nos anos 2000

Systran: o grupo formado inicialmente por pesquisadores na Universidade de Georgetown fundou a empresa Systran no final da década de 1960 e consolidou-se no mercado de fornecedores de ferramentas para TA. O sistema Systran acompanhou a transformação das tecnologias aplicadas a TA e se integrou à modelos de base estatística, híbrida e neural.
Google Tradutor: em 2006 a empresa Google lança sua ferramenta de TA no seu motor de busca de forma gratuita. O sistema seguia o modelo com base estatística e uma década depois, em 2016, atualiza seu modelo para base neural.
Microsoft Translator: em 2009 a empresa Microsoft lança sua ferramenta de TA também em seu motor de busca e com acesso gratuito. Inicialmente de modelo estatístico mas atualizada para o modelo neural a partir de 2016.

Fonte: elaboração própria a partir de dados obtidos no website das empresas mencionadas (2020).

O movimento por “software livre” a partir dos anos 2000 motiva grupos de pesquisa desenvolverem programas *open-source* também no campo da TA, incentivando que empresas desenvolvam suas próprias ferramentas para TA nos casos em que avaliarem ser mais produtivo o desenvolvimento interno do que a aquisição de uma solução no mercado. Com isso, alguns programas *open-source* surgem tanto com base estatística como neural, conforme podemos observar no Quadro 10.

Quadro 10: Programas de TA *open-source* a partir dos anos 2000

Moses: o programa é resultado dos estudos de Hieu Hoang e seu orientador Philipp Koehn no grupo de pesquisa da Universidade de Edimburgo (Escócia) em 2005. A ferramenta é desenvolvida com base estatística e sua característica <i>open-source</i> propicia que outras ferramentas sejam desenvolvidas e/ou integradas de maneira simplificada.
OpenLogos: é a versão <i>open-source</i> da ferramenta Logos que foi desenvolvida pela empresa Logos Corporation de Saarbrücken na Alemanha entre as década de 1980-90. Em 2005 a empresa DFKI passa a converter a ferramenta em <i>open-source</i> para em 2010 ser disponibilizada como OpenLogos.
Apertium: o programa tem como objetivo a tradução de pares de idiomas divergentes, como inglês e catalão, e conta com estudos de grupos de pesquisa da Universidade de Alicante na Espanha.
OpenNMT: uma das primeiras ferramentas <i>open-souce</i> de base neural, com projeto desenvolvido por Yoon Kim em 2016, membro do grupo de pesquisas em processamento de linguagem natural da Universidade de Harvard.

Fonte: elaboração própria a partir de dados obtidos no website dos grupos responsáveis pelas ferramentas (2020).

As pesquisas em NMT atualmente contam com diferentes grupos espalhados por universidades, instituições e corporações numa abrangência global. A história dos

grupos de pesquisas anteriores ficaram registradas por suas publicações acadêmicas e legitimação de seus sistemas de TA na indústria de tradução, a partir de agora continuamos observando o trabalho desses grupos ativos ao redor do mundo para avaliarmos futuramente o impacto de seus trabalhos no mercado e na sociedade.

3.2 As limitações do modelo de base neural

Assim como os sistemas desenvolvidos no modelo de base estatística, os modelos com base neural enfrentam obstáculos que ainda precisam de estudos para buscar soluções de contorno. A fase de treinamento da máquina ainda é complexa, especialmente nos casos em que uma quantidade maior de parâmetros precisam ser ajustados automaticamente, pois o processo pode ser obscuro do ponto de vista dos ajustes que a máquina aplicará, deixando poucas pistas para corrigir erros potenciais.

As palavras desconhecidas, que não foram incluídas na fase de treinamento da máquina, geralmente não são processadas corretamente e isso se torna um dos grandes problemas que o modelo neural precisa ultrapassar. No modelo de base estatística é mais fácil incluir um módulo para resolver o problema com palavras desconhecidas, no entanto no processo de *deep learning* o processamento não é simples, o modelo trabalha numa abordagem de aprendizado de maneira holística, o que limita o aprendizado específico.

Em virtude do modelo de aprendizagem mais amplo do corpus, o sistema traz melhores resultados em sentenças simples, com linguagem controlada, em comparação a sentenças complexas, questão melhor resolvida com o modelo com base estatística. Com isso, alguns grupos de palavras podem não ser traduzidos ou oferecer uma tradução errada.

O processo de aprendizagem da máquina ainda é longo, podendo durar dias dependendo do tamanho do corpus. Os desenvolvedores tem usado processadores específicos para as máquinas e buscam novas técnicas de programação para cada vez mais agilizar o processo. O objetivo com essas técnicas é fazer o processo de aprendizagem ser menos complexo para a máquina e para isso tentar integrar módulos que possam criar conexões entre o codificador e o decodificador, uma vez que ambos executam de maneira independente.

Assim como no modelo estatístico, é difícil mapear e corrigir problemas pontuais de tradução. No caso do modelo neural a tradução é realizada com base no centro da sentença e há casos em que a sentença original pode ser interpretada de

maneira separada, o que pode resultar em traduções desconectadas com a sentença inteira. Essa dificuldade de entender e analisar como funciona um sistema neural se dá pelo fato da representação interna do sistema ser puramente numérica, grande e complexa, resultando em uma representação que não é inteligível pelo ser humano. As pesquisas atuais buscam técnicas para tentarem tornar essas representações internas das redes neurais mais inteligíveis, de maneira que o ser humano consiga “entender” melhor o funcionamento da máquina que ele próprio criou.

O conceito de *deep learning* trouxe para a TA possibilidades de avanço que não se conseguia ultrapassar até a década de 1990, por isso a adoção da metodologia foi bem mais rápida em comparação ao métodos antes implementados. Observamos na discussão anterior que o modelo de base estatística demorou para ser adotado, no entanto, o modelo de base neural parece estar convencendo desenvolvedores e empresas investirem nessa tecnologia de uma maneira mais rápida.

As pesquisas sobre redes neurais datam dos anos de 1980, porém podemos afirmar que a falta de recursos computacionais robustos não possibilitaram um avanço significativo do seu desenvolvimento, fato que torna-se diferente com as pesquisas em *deep learning*, que encontram um ambiente mais desenvolvido para seu progresso.

O desempenho do modelo de base neural associado ao método de *deep learning* certamente poderá superar o desempenho do modelo de base puramente estatística, mas as pesquisas ainda tem muito para evoluir nessa área. Podemos alcançar melhores resultados em menor tempo do que levamos até agora em virtude do conhecimento computacional que temos hoje, porém as pesquisas em PLN ainda enfrentam desafios e não deixaram seu posto de uma das áreas mais complexas no campo das Ciências da Computação, por isso não podemos inferiorizar os estudos e nem duvidar da capacidade científica nessa área.

CAPÍTULO IV

ANÁLISE DOS PROGRAMAS DE INVESTIGAÇÃO DA TRADUÇÃO AUTOMÁTICA

Este capítulo contempla a segunda e terceira etapas de análise conforme especificado no capítulo I, de acordo com as diretrizes propostas por Swiggers (1981, 2013). O material de análise selecionado para compor a narrativa histórica interpretativa da evolução da TA será destacado nesta fase, com o objetivo de apoiar a análise dos parâmetros internos e externos estabelecidos para cada programa de investigação da TA.

Dos parâmetros internos, analisaremos a incidência no nível descritivo de língua em cada programa, quais unidades linguísticas foram privilegiadas naquele modelo de TA e a visão de tradução em relação à unidade linguística privilegiada (palavra, sentença, significado), confirmado pelas técnicas e procedimentos adotados em cada modelo de TA.

Dos parâmetros externos, buscamos discutir o contexto histórico e social em que circulavam as ideias linguísticas que impactaram cada programa de TA. O objetivo é demonstrar a influência do clima de opinião e das circunstâncias institucionais no trabalho dos grupos de pesquisa, no recorte temporal de cada programa, que resultaram em mudanças no desenvolvimento do campo e seu possível impacto na sociedade.

Observamos que os três programas demonstram uma visão geral de língua num modelo de sistema formal, combinando com as características dos estudos do “programa de projeção” (Swiggers, 1981) que trata a língua num modelo abstrato formal, em que se busca por fórmulas matemáticas descrever um padrão linguístico de uma língua natural. No entanto, cada programa da TA trata sua incidência em um nível descritivo linguístico (léxico, sintaxe e semântica) numa perspectiva diferente, bem como as técnicas que demonstram essa visão de língua no desenvolvimento dos modelos de TA, conforme serão apresentados na sequência.

1. Programa de mecanização

1.1 Parâmetros internos

Neste programa a incidência no léxico é a principal característica da linha de estudos dos grupos de pesquisa. O modelo com base em regras pressupõe a transformação da língua de partida na língua de chegada por um processo de transformação com base em regras gramaticais formais, apoiado no uso de dicionários.

A visão da língua como um sistema codificado é observada desde as primeiras correspondências de Warren Weaver (considerado o inventor da TA) nos anos de 1947, 1948 e 1949. A perspectiva de língua como código é percebida quando Weaver propõe a construção de uma máquina de tradução automática a partir de técnicas de criptografia, conforme carta de 1947 ao professor Norbert Wiener do Instituto de Tecnologia de Massachusetts (MIT).

<p>Powerful new mechanized methods in cryptography – methods which I believe succeed even when one does not know what language has been coded – one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say “This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode”. (WEAVER, 1947, p. 1)</p>	<p>Novos métodos mecanizados e poderosos de criptografia - métodos que acredito ter sucesso mesmo quando não se sabe qual linguagem foi codificada -, naturalmente se pergunta se o problema da tradução poderia ser considerado um problema de criptografia. Quando leio um artigo em russo, digo: “Na verdade, está escrito em inglês, mas foi codificado em alguns símbolos estranhos. Eu irei agora decodificá-lo”. (Tradução nossa)</p>
---	--

No ano seguinte (1948), outra correspondência de Weaver, agora encaminhada a W. Prager da Universidade Brown (nos EUA), busca a confirmação de um experimento de tradução com base em criptografia, em que Prager teria sido envolvido. O método consistiu no desafio de Prager em codificar uma mensagem em turco com números de cinco dígitos e passar a mensagem a um colega matemático da instituição para tentar decodificá-la. O experimento não foi de grande sucesso, pois o colega não conseguiu decodificar a mensagem corretamente, porém depois de correções aplicadas a mensagem ficou próxima à original.

<p>You wrote out a message in Turkish and then coded it by some rather complicated code into a column of five digit numbers: and that he brought this back to you within a short time (actually, it seems to me that you said</p>	<p>Você escreveu uma mensagem em turco e depois a codificou com algum código bastante complicado numa coluna de números de cinco dígitos: em que ele trouxe de volta para você em pouco tempo (na</p>
---	---

<p>the next day) with the remark that they had apparently not had much success: but that what he handed you was, when properly broken up into words, and when corrected mildly, your original message in Turkish. (WEAVER, 1948, p. 1)</p>	<p>verdade, parece-me que você disse no dia seguinte) com a observação de que ele aparentemente não teve muito sucesso: mas o que ele te entregou foi, depois de devidamente dividido em palavras e ligeiramente corrigido, sua mensagem original em turco. (Tradução nossa)</p>
--	--

Em 1949, a correspondência de Weaver, na verdade um memorando mais extenso, descreve os procedimentos que precisariam ser seguidos para efetivar a construção da máquina de tradução automática com base no conceito de codificação por ele provado.

O argumento de Weaver tem fundamento em sua experiência de guerra, quando recebiam mensagens criptografadas de maneira secreta sem ter ideia a respeito do idioma original, fato que sustentava sua teoria de que a máquina poderia ler os códigos também. Além disso, a experiência em turco também servia de justificativa, pois o colega que participou do experimento não sabia que se tratava de uma mensagem naquele idioma e tampouco conhecia a língua. Sua teoria assim provaria que a máquina, mesmo sem conhecer um idioma, poderia decodificá-lo.

<p>During the war, when the whole field of cryptography was so secret, it did not seem discreet to inquiry concerning details of this story; but one could hardly avoid guessing that this process made use of frequencies of letters, letter combinations, intervals between letters and letter combinations, letter patterns, etc., which are to some significant degree independent of the language used. This at once leads one to suppose that, in the manifold instances in which man has invented and developed languages, there are certain invariant properties which are again not precisely but to some statistically useful degree, common to all languages. (WEAVER, 1949, p. 3)</p>	<p>Durante a guerra, quando todo o campo da criptografia era tão secreto, não parecia discreto indagar sobre os detalhes dessa história; mas dificilmente se poderia evitar supor que esse processo fazia uso de frequências de letras, combinações de letras, intervalos entre letras e combinações de letras, padrões de letras, etc., que são, em algum grau significativo, independente da linguagem usada. Isso leva imediatamente a supor que, nas múltiplas instâncias em que o homem inventou e desenvolveu as línguas, existem certas propriedades invariáveis que, novamente, não são precisamente, mas em algum grau estatisticamente útil, comuns a todas as línguas. (Tradução nossa)</p>
---	--

A abordagem de língua como código característica do início das pesquisas em TA permaneceu pela década seguinte. Naquele momento (entre os anos de 1950 e 1960) os pesquisadores não se interessavam pela língua em uso, mas pelas relações estabelecidas no sistema numa perspectiva saussuriana da língua. Podemos observar a relação dicotômica do conceito *langue* e *parole*, onde os experimentos em TA se concentravam na *langue*, ou seja, no sistema de unidades e relações de uma língua.

A visão de língua como sistema é confirmada ao observarmos a técnica empregada na organização das estruturas linguísticas de maneira formalizada. O objetivo é transportar um modelo de regras de um sistema de língua formal para um formato em que a máquina pudesse interpretar também de maneira formal. A técnica com base na tradução de regras no nível da palavra visa uma classificação da unidade lexical dentro de uma categoria, que analisa unidades sintáticas da sentença com base na categoria lexical à qual o item pertence.

O método de análises sintática e lexical torna-se a linha de pesquisa considerada mais viável neste programa. A ideia é desenvolver a máquina para transpor regras formais de uma língua para outra. No artigo de Bar-Hillel (1951) podemos observar os detalhes do processo de análises que a máquina precisaria seguir proposto por ele.

<p>1. Mechanical analysis of each word in the FL into the stem (lexical unit) and morphological category. Since linguistic terminology is not unique, let us exemplify the required type of analysis, taking German as FL. The German "ging" will have to be analysed into "gehen" (past, singular, 1st or 3rd person), the German "lieben" into 1) "lieben" (infinitive), 2) "lieben" (present, plural, 1st or 3rd person), 3) "lieb" (adjective, plural), 4) "lieb" (adjective, singular, genitive, dative or accusative-maculine). The analysis of "lieben" is not complete and is even partly incorrect, but it is not necessary here to go into all the linguistic complexities. 2. Mechanical identification of small syntactical units within the given sentence on the basis of morphological categories to which its words belong and, for most languages, their order. 3. Transformation of the given sentence into another that is logically equivalent to it, and rearrangement of the parts of the transformed sentence in accordance with some standard order of the TL. The performance of these three operations will encounter not only great practical difficulties, especially for operation 1, but will also require certain theoretical preparations, on a linguistic and logic level, of a kind that so far has been dealt with only more or less incidentally. (BAR-HILLEL, 1951, p. 232)</p>	<p>1. Análise mecânica de cada palavra na LF dentro de sua raiz (unidade lexical) e na categoria morfológica. Visto que a terminologia linguística não é única, exemplifiquemos o tipo de análise necessária, tomando o alemão como LF. No alemão "ging" terá de ser analisado em "gehen" (passado, singular, 1^a ou 3^a pessoa), em alemão "lieben" em 1) "lieben" (infinitivo), 2) "lieben" (presente, plural, 1^a ou 3^a pessoa), 3) "lieb" (adjetivo, plural), 4) "lieb" (adjetivo, singular, genitivo, dativo ou acusativo-maculino). A análise de "lieben" não é completa e está até parcialmente incorreta, mas aqui não é necessário entrar em todas suas complexidades linguísticas. 2. Identificação mecânica de pequenas unidades sintáticas dentro de uma determinada frase com base nas categorias morfológicas às quais suas palavras pertencem e, para a maioria das línguas, sua ordem. 3. Transformação da frase dada em outra que seja logicamente equivalente a ela, e rearranjo das partes da frase transformada de acordo com a ordem padrão da LA. A execução dessas três operações encontrará não apenas grandes dificuldades práticas, especialmente para a operação 1, mas também exigirá certos preparativos teóricos, em um nível linguístico e lógico, de uma maneira que até agora tem sido tratado apenas mais ou menos a propósito. (Tradução nossa)</p>
--	---

A visão de tradução em relação à unidade linguística neste programa está no nível da palavra. A técnica de transformação da língua de partida na língua de chegada se dá a partir da busca em dicionário. Os estudos na década de 1950 concentravam-se no aprimoramento de técnicas de busca em dicionário, nas quais os desafios estavam em tornar a busca mais rápida e ultrapassar a barreira dos vários significados que uma mesma palavra pode ter (OSWALD, 1952, p. 2; 1953, p. 2 e YNGVE, 1953, p.1). O desenvolvimento de glossários auxiliava na questão dos diferentes significados para uma mesma palavra, uma vez que era possível criar terminologias específicas de acordo com o tipo de texto a ser traduzido.

<p>Syntactic connection will almost infallibly identify word-function, and we now know that a recognition of syntactic connection can be built into the “memory” of machines of the high speed computer type. Word-meaning, on the other hand, is not a factor capable of being solved mechanically except by an elaborate reduction of the possibilities of multiple significance: that is by the production of a large – possibly very large – number of glossaries that pertain to one radically limited field of discourse. (OSWALD, 1952, p. 2)</p>	<p>A conexão sintática identificará quase infalivelmente a função da palavra, e agora sabemos que um reconhecimento da conexão sintática pode ser embutido na “memória” de máquinas do tipo de computadores de alta velocidade. O significado das palavras, por outro lado, não é um fator capaz de ser resolvido mecanicamente, exceto por uma redução elaborada das possibilidades de múltiplos significados: isto é, pela produção de um grande - possivelmente muito grande - número de glossários que pertençam a um campo de discurso radicalmente limitado. (Tradução nossa)</p>
<p>If translation of technical articles from one language to another is ever to be done by machine, it is obvious that one operation such a machine translation would have to perform is that of consulting a dictionary. Such an operation would be rather easy to mechanize in a special purpose machine, or to program in existing digital computers. It is also clear that other operations will have to be mechanized in order to obtain a good translation. The nature of these other operations is, at present, not completely clear. (YNGVE, 1953, p. 35)</p>	<p>Se a tradução de artigos técnicos de um idioma para outro tiver que ser feita por uma máquina, é óbvio que uma das operações que essa tradução automática teria de realizar é consultar um dicionário. Tal operação seria bastante fácil de mecanizar em uma máquina de uso especial ou de se programar em computadores digitais existentes. É claro também que outras operações terão de ser mecanizadas para se obter uma boa tradução. A natureza dessas outras operações não é, no momento, totalmente clara. (Tradução nossa)</p>

<p>The most effective translator of technical material has the bulk of the necessary vocabulary at his fingers tips, and needs to have recourse of dictionaries, or to seek consultation, only for a relatively small number of forms. Let us admit that any foreseeable apparatus for mechanical translation can only hope to emulate such a translator's facility. If we can devise special glossaries to be scanned by electronic devices, then a mechanical operation might compete with, or even surpass, a translator's more or less instantaneous translation of terms with which he is familiar; but no mechanical device can possibly cope with the problem of unpredicted forms. (OSWALD, 1953, p. 2)</p>	<p>O tradutor mais eficaz de material técnico tem a maior parte do vocabulário necessário nas pontas dos dedos e precisa recorrer a dicionários, ou buscar ajuda, apenas para um número relativamente pequeno de formas. Admitamos que qualquer aparato previsível de tradução mecânica possa apenas esperar que emule a facilidade de tal tradutor. Se pudermos conceber glossários especiais para serem digitalizados por dispositivos eletrônicos, então uma operação mecânica pode competir com, ou mesmo superar, a tradução mais ou menos instantânea de um tradutor em relação a termos com os quais ele esteja familiarizado; mas nenhum dispositivo mecânico pode lidar com o problema de formas imprevistas. (Tradução nossa)</p>
---	---

O modelo de TA com base em palavra dominou o programa de mecanização logo no início, no entanto, não tardou para que os pesquisadores percebessem que seria necessário ampliar a análise da língua de partida para grupos de palavras com ênfase na função sintática de cada palavra na frase.

Os estudos com base no significado da frase começaram a ser desenvolvidos fora dos EUA no período em que as pesquisas foram abandonadas nesse país, logo após o corte nos investimentos em virtude da publicação do relatório da ALPAC em 1966. A partir dessa linha de pesquisa, surgem experimentos apoiados no conceito de interlíngua, que consiste num tipo de arquitetura em que a sentença é analisada na língua de partida pelas relações interlinguais existentes na sentença.

No entanto, a dificuldade na distinção entre os vários significados de uma palavra continuava uma barreira para o método com base no nível da palavra, como mostra os resultados do experimento conduzido na Universidade de Cambridge:

<p>It has been tacitly presupposed, in our work, that the only form of Mechanical Translation which is worth trying for is what we, indirectly following Bar-Hillel, call FARIMT (Fully Automatised Reasonably Idiomatic Mechanical Translation). It might be further inferred, from what we have said, that, believing as we do that the basic problem of determining the nature of semantic structure, - technically called the Problem of Multiple Meaning or of polysemy – has got to be faced and solved before FARIMT becomes even in</p>	<p>Foi tacitamente pressuposto, em nosso trabalho, que a única forma de tradução mecânica que vale a pena tentar é aquela que nós, indiretamente seguindo Bar-Hillel, chamamos de FARIMT (Fully Automatised Reasonably Idiomatic Mechanical Translation, Tradução Mecânica Totalmente Automatizada e Razoavelmente Idiomatica). Pode-se ainda inferir, a partir do que dissemos, que, acreditando que o problema básico de determinar a natureza da estrutura semântica, - tecnicamente chamado de</p>
---	--

<p>principle attainable, we do not think it worth while even to try Experimental Machine Translation, - that as Oettinger has said about us (1), we are "quixotic" about taking the step of trying it. (MASTERMAN; KAY, 1960, p. 1)</p>	<p>Problema de Significado Múltiplo ou de polissemia - deve ser enfrentado e resolvido antes de a FARIMT tornar-se até em princípio alcançável, não achamos que valha a pena tentar a Tradução Automática Experimental - que, como Oettinger disse sobre nós (1), somos "quixotescos" em dar o passo de experimentá-la. (Tradução nossa)</p>
---	--

A técnica de tradução com base na aplicação de regras formais da língua permaneceu como principal linha de estudos dos grupos de pesquisa neste programa, a máquina era programada para analisar a língua de partida com base em regras gramaticais formais daquela língua, e durante o processo de transformação ela aplicaria regras gramaticais pertinentes à língua de chegada. O resultado era uma tradução adequada do ponto de vista de estrutura de língua formal, porém com baixa fluência e coerência semântica.

Ao observarem o padrão em que a máquina produzia melhores resultados, uma solução de contorno poderia ser não somente a pós-edição do texto traduzido, mas também a pré-edição, na qual o texto-fonte seria preparado para a transformação da máquina em texto-alvo. O texto-fonte seria editado num formato simplificado de frases curtas, bem como na padronização de estilo e abreviaturas:

<p>JOSS: And then again, concerning the original author, when editing is spoken of here we generally mean post-editing, eliminating pre-editing, but I do not think that technical translation by machine on a large scale can afford to do without a certain modest kind of pre-editing which can be done by a monolingual editor – an editor who knows nothing about problems of translation, but operates quite mechanically. Such person would have a certain specific jobs to do. One, he must reduce the original text to sentences no more than twenty words long each. By hook or crook, they've got to be broken up in sentences each one which is not more than twenty words long. Then this pre-editor is to introduce few stylistic simplifications. (That's not essential, but I think such an editor can be taught to introduce enough stylist simplifications so that the machine can handle it more easily, and then this pre-editor will have to standardize abbreviations, expand some, expand all abbreviations that are not on the standard list. The three things can be done</p>	<p>JOSS: E, novamente, no que diz respeito ao autor original, quando se fala em edição aqui, geralmente queremos dizer pós-edição, eliminando a pré-edição, mas eu não acho que a tradução técnica por máquina em grande escala possa prescindir de um certo tipo modesto de pré-edição, que pode ser feita por um editor monolíngue - um editor que nada sabe sobre problemas de tradução, mas opera de forma bastante mecânica. Essa pessoa teria certas tarefas específicas a fazer. Primeiro, ele deve reduzir o texto original a frases com não mais do que vinte palavras cada. Por bem ou por mal, eles precisam ser divididos em frases cada uma com não mais do que vinte palavras. Então, este pré-editor irá apresentar algumas simplificações estilísticas. (Isso não é essencial), mas acho que esse editor pode ser ensinado a introduzir simplificações estilísticas suficientes para que a máquina possa lidar com isso mais facilmente, e então este pré-editor terá que padronizar abreviações, expandir algumas e expandir todas as</p>
--	--

<p>by a monolingual, and then the machine that I envisage for technical translation might be complete, I think, in a next generation. (DOSTERT, 1957, p. 47)</p>	<p>abreviações que não estão na lista padrão. As três coisas podem ser feitas por um monolíngue, e então a máquina que imagino para a tradução técnica pode estar completa, eu acho, em uma próxima geração. (Tradução nossa)</p>
--	---

A limitação tecnológica impediu um avanço significativo no desenvolvimento de ferramentas de TA com qualidade, os pesquisadores acreditavam que ainda seria preciso ao menos uma década para obterem resultados razoáveis, uma vez que as pesquisas já começavam ser conduzidas a partir do ponto de vista de que a máquina tinha que aprender com a própria experiência, conforme Dostert (1957, p. 45-47).

<p>JOSS: As I remarked in a publication which will be out in a couple of weeks now, the <i>Readings in Linguistics</i> volume, the development of linguistics in this country from 1925 to 1955, 30 years – one generation – is about the same in extent as the development of mathematics from Newton to about 1850, or in chemistry from Dalton’s atomic hypothesis of 1810 approximately down to nuclear fission, about two generations. So we have skipped from 200 years down to 60 years, and now down to 30. Thus the time scale keeps changing. The culture is developing more rapidly all the time. I do regard machine translation – which I would consider as being adequate in some sort of competition for what I can do myself in some of the languages I know – I do regard that as millennial. But considering the very swift progress in technology, possibly my grandchildren may witness it. (DOSTERT, 1957, p. 45)</p>	<p>JOSS: Como analisei em uma publicação que sairá daqui a algumas semanas, o volume <i>Readings in Linguistics</i> (Leituras na Linguística), o desenvolvimento da linguística neste país de 1925 a 1955, 30 anos - uma geração - é quase a mesma extensão do desenvolvimento da matemática de Newton até perto de 1850, ou da química na hipótese atômica de Dalton em 1810, aproximadamente, até a fissão nuclear cerca de duas gerações. Portanto, saltamos de 200 anos para 60 anos e agora para 30. Assim, a escala de tempo continua mudando. A cultura está se desenvolvendo mais rapidamente o tempo todo. Eu considero a tradução automática - que consideraria adequada em algum tipo de competição pelo que eu mesmo posso fazer em algumas das línguas que conheço – acredito que em um milênio. Mas, considerando o rápido progresso da tecnologia, possivelmente meus netos poderão testemunhar isso. (Tradução nossa)</p>
<p>JOSS: I think that the programming of the kind of machine-computer that they are working on now – a computer that learns by experience – has to include a pleasure principle, or a reward-and-punishment in the machine. (DOSTERT, 1957, p. 47)</p>	<p>JOSS: Acho que a programação do tipo de máquina-computador em que estão trabalhando agora - um computador que aprende por experiência - deve incluir um princípio de agrado ou uma recompensa e punição na máquina. (Tradução nossa)</p>

No programa de mecanização a visão de língua era no nível da palavra, no qual empregavam-se técnicas de estruturação de regras gramaticais de uma língua, com o objetivo de processamento da tradução numa abordagem de formalização da

linguagem. Assim, a incidência da língua no nível descritivo linguístico estava no léxico, em que o método empregado era primeiro a categorização lexical de cada palavra na sentença da língua de partida, seguido de uma análise sintática dessa mesma sentença com base nas categorias identificadas e, finalmente, no processo de transformação, as regras gramaticais da língua de chegada eram aplicadas à essa representação.

1.2 Parâmetros externos

A TA é idealizada por pesquisadores americanos para resolver o problema da crescente demanda por tradução de conteúdo científico estrangeiro, em especial em russo. Um dos principais objetivos era aumentar a rapidez para traduzir a fim de acelerar o entendimento a respeito das estratégias de desenvolvimento de outros países no campo científico e político, motivados por um cenário pós-Segunda Guerra Mundial (1939-1945).

O programa de mecanização teve o papel de estabelecer um padrão de pesquisa científica na área. Os grupos de pesquisa desse programa buscavam uma legitimação científica dos estudos conduzidos no ambiente acadêmico, e para isso contavam com financiamento do governo, bem como com parceria de empresas privadas, como a IBM.

O início das pesquisas em TA não contou com uma teoria linguística específica que servisse de fundamentação para seu desenvolvimento e nem com estudos em tradução, uma vez que os Estudos da Tradução não existia como uma disciplina até final da década de 1970. Nesse período começaram as primeiras discussões em nível acadêmico da teoria da comunicação e da linguística estrutural que seriam aplicadas aos estudos da tradução posteriormente. Assim, a visão de tradução na concepção da TA não contou com suporte em teorias da tradução e a ideia de TA foi concebida como um processo de transformação de estruturas lógicas.

<p>When research on machine translation started in the late forties, the linguistic theory which would have been needed as a foundation for such an ambitious undertaking simply did not exist. Since then progress in linguistics and in Natural Language Processing has been immense, even though we still only partially understand the problem. When Michael</p>	<p>Quando as pesquisas sobre tradução automática começaram no final dos anos 40, a teoria linguística que teria sido necessária como base para um empreendimento tão ambicioso simplesmente não existia. Desde então, o progresso em linguística e em Processamento de Linguagem Natural tem sido imenso, embora ainda entendamos apenas parcialmente o problema. Quando</p>
--	--

<p>McCord of IBM Research started the project Logic-programming based Machine Translation (LMT) in 1985, he was able to build on the latest insights from linguistics, logic, and logic programming. As a result we now have a machine translation system which exhibits the most modern technology commercially available today. (LEHMANN, 1995, p. 1)</p>	<p>Michael McCord, da IBM Research, iniciou o projeto de Tradução Automática com base em Programação Lógica (LMT) em 1985, ele foi capaz de desenvolvê-lo a partir de conceitos mais recentes em linguística, lógica e programação lógica. Como resultado, agora temos um sistema de tradução automática que exhibe a tecnologia mais moderna disponível comercialmente hoje. (Tradução nossa)</p>
---	--

Os grupos de pesquisa do programa de mecanização (ver capítulo III) eram formados por professores, pesquisadores e alunos de universidades americanas e europeias (a maioria americana), que contavam com financiamento do governo para as pesquisas em TA, mas também algumas parcerias privadas como a da empresa IBM, conforme já mencionado. Os grupos coordenados por professores (ver capítulo III) das instituições, faziam a gestão administrativa e financeira de cada grupo, bem como a divulgação de sua produção acadêmica.

De acordo com Murray (1994), um grupo de especialidade teórica precisa de liderança intelectual e organizacional para trabalhar em uma comunidade científica na solução de problemas, e a importância da formação desses grupos no programa de mecanização foi a motivação científica que os levou a busca para solucionar o problema da demanda de tradução que seria impossível de ser atendida apenas com trabalho humano, sem o auxílio da máquina. Essa demanda por tradução significava aperfeiçoar o conhecimento científico de outras áreas, o que resultaria na evolução científica e intelectual de uma comunidade como um todo.

O objetivo por parte dos grupos era alcançar uma TA que pudesse oferecer uma ideia geral de um texto numa língua estrangeira, conheciam as limitações do sistema e não esperavam uma tradução comparável à humana. No entanto, a expectativa por parte do governo que financiava os estudos era diferente, esses esperavam um processo de tradução totalmente automatizado, sem nenhuma intervenção humana, conforme apontado no relatório da ALPAC (1966).

<p>The contention that there has been no machine translation of general scientific text is supported by the fact that when, after 8 years of work, the Georgetown University MT project tried to produce useful output in 1962, they had to resort to postediting. The postedited translation took slightly longer to do and was more expensive than</p>	<p>A alegação de que não houve tradução automática de texto científico geral é apoiada pelo fato de que quando, após 8 anos de trabalho, o projeto de TA da Universidade de Georgetown tentou produzir conteúdo útil em 1962, eles tiveram que recorrer à pós-edição. A pós-edição da tradução demorou um pouco mais para ser feita e foi mais cara do</p>
--	--

<p>conventional human translation. [...] Unedited machine output from scientific text is decipherable for the most part, but it is sometimes misleading and sometimes wrong (as is postedited output to a lesser extent), and it makes slow and painful reading. (ALPAC, 1966, p. 19)</p>	<p>que a tradução humana convencional. [...] O resultado da tradução de máquina de texto científico sem pós-edição é decifrável na maior parte, mas às vezes é enganosa e às vezes é errada (assim como a pós-edição numa proporção menor) e torna a leitura lenta e penosa. (Tradução nossa)</p>
---	---

O relatório da ALPAC publicado com o título “*Language and Machines: computers in translation and linguistics*” (Linguagem e Máquinas: computadores na tradução e linguística) foi usado para justificar os cortes nos investimentos das pesquisas em TA. Ele propunha diferentes medidas que deveriam ser adotadas para redução de gastos com tradução, como a diminuição dos serviços de tradução em geral demandados pelo governo e a redução do quadro de tradutores contratados pelas agências governamentais.

A argumentação desenvolvida no documento estava embasada em custos e incluía relatórios financeiros declarando que teriam sido gastos 10 milhões de dólares naquela última década em pesquisas que não conseguiram produzir traduções automaticamente, nomeando os estudos da Universidade de Georgetown como exemplo de projeto que não gerou resultados satisfatórios. Os dados financeiros buscavam também ressaltar o grande volume de investimentos que ainda seria necessário para dar seguimento aos estudos.

As limitações tecnológicas e computacionais que impediam o avanço da TA começaram desencorajar pesquisadores, como Bar-Hillel que questionaria a viabilidade dos estudos. Ele acreditava no conceito de TA totalmente automatizada - FARIMT (*Fully Automatised Reasonably Idiomatic Mechanical Translation*, Tradução Mecânica Totalmente Automatizada e Razoavelmente Idiomática) - e propunha um processo que incluísse as etapas de pré-edição e pós-edição do texto, porém entendendo que isso não se tratava de TA pura e que outros problemas, como a escassez de tradutores especialistas, seriam enfrentados.

<p>This fact, that high-accuracy, fully automatic MT is not achievable in foreseeable future, has discouraged many thinkers from whom an interest in MT as a socially important, noncomputational, application of digital computer-like machines might have been expected; it has discouraged them to such a degree that they have failed to see that, with a lowering of the target, there appear less ambitious aims to achievement of which is</p>	<p>Este fato, de que a TA de alta precisão e totalmente automática não é alcançável em um futuro previsível, desencorajou muitos pensadores de quem poderia ser esperado um interesse na TA como uma aplicação socialmente importante e não computacional de máquinas digitais semelhantes a computadores; desanimou-os a tal ponto que não conseguiram ver que, com a redução da meta, parecem objetivos menos</p>
---	---

<p>still theoretically and practically valuable. [...] For those targets in which high accuracy is a condition <i>sine que non</i>, pure MT has to be given up in favor of a mixed MT, i. e., a translation process in which a human brain intervenes. There the question arises: Which parts of the process should be given to the human partner? In principle, there are a large number of ways in which a machine and a human brain may collaborate. The following considerations are pertinent: First, the more the human partner does, the less complex the machine will have to be; second, the human partner will have to be placed either at the beginning of the translation process or the end, perhaps at both, but preferably not someone in the midst of it, according to a well-known principle of electronic computer handling; third, and perhaps most important, since the major bottleneck in translation lies in the scarcity of expert translators, the human partner should be required to know only one of the languages, concerned either the FL or the target language (TL). (BAR-HILLEL, 1951, p. 230)</p>	<p>ambiciosos a serem alcançados, os quais ainda podem ser valiosos em termos teóricos e práticos. [...] Para aqueles objetivos em que a alta precisão é uma condição <i>sine que non</i>, a ideia de TA pura deve ser abandonada em favor de uma TA mista, por exemplo, um processo de tradução no qual um cérebro humano intervém. Aí surge a pergunta: Quais partes do processo devem ser dadas ao parceiro humano? Em princípio, há um grande número de maneiras pelas quais uma máquina e um cérebro humano podem colaborar. As seguintes considerações são pertinentes: primeiro, quanto mais o parceiro humano faz, menos complexa a máquina terá de ser; segundo, o parceiro humano terá de ser colocado no início ou no final do processo de tradução, talvez em ambos, mas de preferência não alguém no meio dele, de acordo com um princípio bem conhecido de manipulação de computador eletrônico; terceiro, e talvez o mais importante, uma vez que o maior gargalo na tradução reside na escassez de tradutores especializados, o parceiro humano poderá conhecer apenas um dos idiomas, seja a língua de partida ou a língua de chegada. (Tradução nossa)</p>
---	--

Os argumentos de Bar-Hillel acabaram servindo de justificativa para o relatório da ALPAC, desencadeando o rompimento com o apoio do governo aos grupos de pesquisa em TA nos EUA, em virtude do corte dos investimentos na área.

A repercussão no meio acadêmico foi impactante e as revistas *Science* e *Communications of the ACM* de Washington publicaram artigos com títulos, *Machine Translation: Committee Skeptical over Research Support* (Tradução Automática: comitê cético em relação ao apoio às pesquisas) e *The Nebulous Future of Machine Translation* (O futuro nebuloso da tradução automática), que questionavam as afirmações do relatório da ALPAC. Os artigos traziam a resposta do diretor do centro de pesquisas em TA da Universidade de Georgetown a respeito da visão equivocada em relação aos resultados obtidos nas pesquisas em TA até aquele momento, e a respeito dos valores superestimados para investimentos futuros, conforme a seguir:

<p>In an interview with <i>Science</i> R. Ross Macdonald, director of the Georgetown University Machine Translation Research Project, predicted that “freely usable</p>	<p>Em uma entrevista à <i>Science</i>, R. Ross Macdonald, diretor do Projeto de Pesquisa de Tradução Automática da Universidade de Georgetown, previu que “a tradução</p>
---	---

<p>machine translation will be available within 4 or 5 years, and perhaps earlier than that". Macdonald readily admits that exaggerated claims for machine translation in the past have had the effect of souring many people about the possibility of ever achieving such translation. (NELSON, 1967, p. 2)</p>	<p>automática gratuita estará disponível dentro de 4 ou 5 anos, e talvez antes disso". Macdonald admite prontamente que alegações exageradas sobre a tradução automática no passado tiveram o efeito de desmotivar muitas pessoas sobre a possibilidade de algum dia alcançar esse tipo de tradução. (Tradução nossa)</p>
<p>Even though he is highly critical of the Alpac report, Dr. R. Ross Macdonald, director of the Georgetown University Machine Translation Research Project, is optimistic. In fact, he believes freely usable machine translation could be available in four or five years if a group as far along as his were to receive \$150,000 worth to support each year, one-fourth the amount Alpac recommended for computational linguistics. (TITUS, 1967, p. 190)</p>	<p>Embora seja muito crítico em relação ao relatório da Alpac, o Dr. R. Ross Macdonald, diretor do Projeto de Pesquisa de Tradução Automática da Universidade de Georgetown, está otimista. Na verdade, ele acredita que a tradução automática de uso livre poderia estar disponível em quatro ou cinco anos se um grupo tão antigo quanto o dele recebesse 150 mil dólares de apoio por ano, um quarto da quantia recomendada pela Alpac para a linguística computacional. (Tradução nossa)</p>

A decisão política do corte nos investimentos americanos nas pesquisas em TA provocou uma estagnação dos estudos por mais de uma década e o prejuízo para a área foi de alguma forma global, uma vez que outros países se desencorajaram em apoiar o campo depois da decisão americana.

O programa de mecanização não conseguiu avançar no aprimoramento dos resultados da TA por falta de recursos computacionais que ainda não existiam naquela época. O corte nos investimentos impediu que os estudos evoluíssem naquele momento junto com os progressos que as Ciências Computacionais alcançavam. Conhecer esse programa, nos permite entender os motivos do surgimento da TA, de sua ascensão e queda num espaço temporal de duas décadas, bem demarcados por seu início e fim.

2. Programa de matematização

2.1 Parâmetros internos

O programa de matematização engloba as pesquisas em TA a partir da retomada dos estudos americanos na década de 1980, em que seus primeiros anos caracterizam-se pela incidência da língua no nível da sintaxe, influenciada pelo conceito de linguística gerativa transformacional de Noam Chomsky, conforme Witkam (2005, p.8).

<p>But when MT research developed in the United States, Chomsky's transformational-gerative grammar was very influential among linguists there. That model with its abstractions and constituency syntax became a real fashion that spread among MT researchers in Europe and Japan as well. [...] That was the situation till the end of the 1980's. (WITKAM, 2005, p. 8)</p>	<p>Mas quando a pesquisa em TA se desenvolveu nos Estados Unidos, a gramática gerativa-transformacional de Chomsky foi muito influente entre os linguistas de lá. Esse modelo com suas abstrações e sintaxe constituinte tornou-se uma moda real que se espalhou entre os pesquisadores de TA na Europa e no Japão também. [...] Essa era a situação até o final da década de 1980. (Tradução nossa)</p>
--	--

Devido à influência chomskiana no conceito de linguística, o paradigma do formalismo linguístico está presente nesse programa que adotava o conceito de língua como estrutura formal, numa perspectiva de sistema de unidades linguísticas, priorizando assim o sistema e não o processo comunicativo em si.

O dicionário como referência canônica era um recurso executado na máquina para aplicar a língua no sistema formalizado, e o aprimoramento de verbetes com informações de contexto, expressões idiomáticas, relações semânticas melhoraram o desempenho da máquina que contava com mais possibilidades de escolha para resolver ambiguidades de significados, de acordo com Coughlin (1988, p. 3).

<p>Nonetheless, semantic dictionaries, without being universal encyclopedias, incorporate increasing amounts of knowledge data which contributes to successful homograph resolution. Not only do these semantics dictionaries handle idiomatic expressions such as "hold one's tongue or make away with", they also include the use of contextual information which is common or essential in a sentence or a text. They include what are called semantic "universals" or "primitive elements" such as "human," "animate," "liquid," etc. but also semantic relations such as "agent-action" or "cause-effect." While earlier direct systems would print a list of two or three possible translations for the term "bank" meaning "geological feature" or "financial institution", present systems making use of contextual information will make a selection using semantic categorization which include subject-fields able to distinguish between the world of finance and geology. (COUGHLIN, 1988, p. 3)</p>	<p>No entanto, os dicionários semânticos, sem serem enciclopédias universais, incorporam quantidades crescentes de dados de conhecimento que contribuem para a resolução bem-sucedida de homógrafos. Esses dicionários de semântica não apenas lidam com expressões idiomáticas como "segurar a língua ou fugir", mas também incluem o uso de informações contextuais que são comuns ou essenciais em uma frase ou texto. Eles incluem o que é chamado de "universais" semânticos ou "elementos primitivos", como "humano", "animado", "líquido" etc., bem como, relações semânticas, como "agente-ação" ou "causa-efeito". Enquanto os sistemas diretos anteriores imprimiam uma lista de duas ou três traduções possíveis para o termo "banco" que significa "característica geológica" ou "instituição financeira", os sistemas atuais que fazem uso de informações contextuais fazem uma seleção usando categorização semântica que inclui campos de assunto, permitindo distinguir entre o mundo das finanças e da geologia. (Tradução nossa)</p>
---	--

A criação da técnica de TA com base em corpus propiciou o desenvolvimento do modelo de TA com base estatística, linha de pesquisa que rompeu com padrões anteriores em virtude da visão sobre as relações entre as estruturas sintáticas e o contexto no nível da sentença (e não apenas da palavra como no programa de mecanização) necessárias para a construção de sentido na linguagem.

Para isso foi necessário o desenvolvimento de técnicas de análises das sentenças por meio de analisadores (*parsers*) programados para identificar automaticamente as estruturas sintáticas e semânticas na língua de partida durante o processo de aprendizagem do corpus pela máquina, conforme Coughlin (1988, p.1).

<p>Originally language analysis in AI was not concerned with the translation of texts from one language to another but rather with more general problems of information processing such as the development of question and answer programs or the machine's ability to extract information in order to summarize it or organize it in a given manner. To accomplish these tasks AI research has developed several techniques including the use of semantic parsing, the consultation of expert systems, and knowledge databases as well as the creation of high level of programming languages designed for symbolic rather than numeric computing such as LISP (Winston 84) and PROLOG (Kowalski 1985). [...] More realistic than their predecessors, these scientists are not aiming for FAHQMT, but hope to reduce textual ambiguities with more sophisticated syntactic and semantic parsing. (COUGHLIN, 1988, p. 1-2)</p>	<p>Originalmente, a análise de língua em IA não se preocupava com a tradução de textos de uma língua para outra, mas com problemas mais gerais de processamento de informações, como o desenvolvimento de programas de perguntas e respostas ou a capacidade da máquina de extrair informações para resumi-las, ou organizá-las, de uma determinada maneira. Para realizar essas tarefas, a pesquisa em IA desenvolveu várias técnicas, incluindo o uso de análise semântica, a consulta de sistemas especialistas e bancos de dados de conhecimento, bem como a criação de linguagens de programação de alto nível projetadas para computação simbólica em vez de numérica, como LISP (Winston 84) e PROLOG (Kowalski 1985). [...] Mais realistas do que seus predecessores, esses cientistas não buscam a FAHQMT, mas esperam reduzir as ambiguidades textuais com uma análise sintática e semântica mais sofisticada. (Tradução nossa)</p>
--	---

Ainda nos anos 80, as técnicas de análise começaram priorizar a análise semântica que passa ser mais adotada, demonstrando a partir dessa abordagem maior incidência da língua no nível da semântica, conforme Goughlin (1988, p.4).

<p>As their names indicates these programs apply semantic parsing and semantic analysis approaches as a first step to natural language understanding followed, when necessary, by syntactic analysis. [...] As is pointed out by researchers, semantic based systems do not translate but interpret or paraphrase the text since general meaning is</p>	<p>Como seus nomes indicam, esses programas aplicam abordagens com analisadores semânticos e análise semântica como um primeiro passo para o entendimento da linguagem natural seguido, quando necessário, pela análise sintática. [...] Como é apontado por pesquisadores, os sistemas com base em semântica não traduzem, mas interpretam ou parafraseiam</p>
---	---

retained but textual structure often is not. (COUGHLIN, 1988, p. 4)	o texto, uma vez que o significado geral é mantido, mas a estrutura textual muitas vezes não é. (Tradução nossa)
---	--

As primeiras ferramentas de TA desenvolvidas a partir da década de 1980 combinavam o método de aplicação de regras gramaticais formais e a abordagem de aprendizagem da máquina com base em corpus (Coughlin, 1988, p. 5), isso porque o volume de corpus bilíngue ainda estava em construção.

Most of the recent efforts in MT in eighties have combined both linguistics-based and AI-based approaches and many systems defy any neat classification. Such are Susy at University at Saar, Geta of the University of Grenoble, Metal of University of Texas and Eurotra of the European Economic Communities. AI strategies including semantic parsers, knowledge databases, expert systems and inference routines have all been incorporated more or less extensively into disambiguation routines. (COUGHLIN, 1988, p. 5)	A maioria dos esforços recentes em TA nos anos 80 combinou abordagens com base em linguística e em IA e muitos sistemas desafiam qualquer classificação simples. Como, por exemplo, Susy da Universidade em Saar, Geta da Universidade de Grenoble, Metal da Universidade do Texas e Eurotra das Comunidades Econômicas Europeias. Estratégias de IA, incluindo analisadores semânticos, bancos de dados de conhecimento, sistemas especialistas e rotinas de inferência, foram incorporadas, mais ou menos extensivamente, às rotinas de desambiguação. (Tradução nossa)
--	---

O modelo com base estatística foi desenvolvido a partir da visão de tradução no nível da sentença, que buscava as relações de sentido entre palavras num contexto. Para alcançar esse nível de tradução, a geração de ferramentas entre as décadas de 80 e 90 tinham a característica de serem híbridas, ou seja, combinavam módulos de análise de conhecimento e estatístico.

O projeto do sistema Pangloss, por exemplo, desenvolvido na Universidade de Carnegie Mellon, em Pittsburgh, consistia numa arquitetura híbrida, para tradução do espanhol e japonês para o inglês, que combinava módulos de transferência com base no léxico (tradução de palavra por palavra e sentença por sentença), análise estatística com base em exemplos e análise semântica do texto-fonte, de acordo com Carbonell et al. (1994, p. 240-241).

Some central aspects of Pangloss are: lexical transfer (word-by-word and phrase-by-phrase substitution using a large bilingual glossary); example-based MT (identification of phrases or even passages from the source text that appear in the large pre-constructed bilingual corpus for examples, and subsequent production of the translated phrases or passages as the	Alguns aspectos centrais do Pangloss são: transferência lexical (substituição de palavra por palavra e frase por frase usando um grande glossário bilíngue); TA com base em exemplos (identificação de frases ou mesmo passagens do texto-fonte que aparecem no grande corpus bilíngue pré-construído em exemplos e produção subsequente das frases ou passagens traduzidas como texto-alvo);
--	---

<p>target text); knowledge-based MT (the more traditional route involving source text parsing, analysis and generation, using an internal text representation that is gradually being upgraded to an interlingua). [...] The system involves a mixture of knowledge-based and statistical modules, as well as knowledge resources built up in various ways, by hand, from online dictionaries and knowledge bases, and through statistical extraction of knowledge from text. (CARBONELL et al., 1994, p. 240-241)</p>	<p>TA com base em conhecimento (a rota mais tradicional envolvendo análise do texto-fonte, análise e geração do texto-fonte, usando uma representação de texto interna que tem sido gradualmente atualizada para uma interlíngua). [...] O sistema consiste de uma mistura de módulos com base em conhecimento e estatística, bem como recursos de conhecimento construídos de várias maneiras, à mão, a partir de dicionários online e bases de conhecimento, assim como através da extração estatística a partir do conhecimento do texto. (Tradução nossa)</p>
--	---

Seguindo a mesma abordagem híbrida, outro exemplo é o projeto da Universidade de Regina, no Canadá, para tradução do inglês para tailandês, que empregava três fases de análise da máquina (com base em análise estatística), bem como analisadores sintáticos e semânticos no nível da sentença, conforme Naruedomkul e Cercone (1997, p. 63).

<p>The system is divided into three phases: quick and dirty machine translation (QDMT), conceptual comparison and repair and iterate. QDMT generates the appropriate translation candidates (TCs) in the target language for the input sentence in the source language. Next, the system compares the meaning of the TC with that of the input sentence. If there is a dissimilarity in meaning between the TC and the input sentence, the most appropriate TC will then be "repaired". [...] In the second phase, conceptual comparison, we will employ sophisticated head driven phrase structure grammar (HPSG) [Pollard and Sag, 1987] parsers for the source and target languages in order to conceptually compare the parser's outputs. (NARUEDOMKUL; CERCONE, 1997, p. 63)</p>	<p>O sistema é dividido em três fases: tradução automática rápida e suja (QDMT), comparação conceitual, e reparo e iteração. O QDMT gera os candidatos de tradução (CT) apropriados na língua de chegada para a sentença de entrada na língua de partida. Em seguida, o sistema compara o significado do CT com o da sentença de entrada. Se houver diferença de significado entre o CT e a sentença de entrada, o CT mais apropriado será "reparado". [...] Na segunda fase, a comparação conceitual, empregaremos analisadores sofisticados da gramática de estrutura de frases dirigida por cabeçalhos (HPSG) [Pollard e Sag, 1987] nas línguas de partida e de chegada para comparar conceitualmente os resultados do analisador. (Tradução nossa)</p>
---	--

No programa de matematização a visão de língua era na sentença e seguia o conceito de língua como estrutura formal, priorizando a análise no nível de unidades linguísticas. Por influência da gramática gerativa transformacional de Chomsky (a partir da década de 1960), os primeiros anos do programa são caracterizados pela incidência da língua no nível da sintaxe que empregava técnicas de análise sintática para análise de corpus.

A partir do surgimento do método de TA com base em corpus, o modelo estatístico foi adotado passando a ser empregadas análises não apenas no nível sintático, mas também semântico tornando-se a principal técnica desde então. O resultado foi o desenvolvimento de ferramentas de TA híbridas, que combinavam módulos de análises da língua no nível da sentença com o objetivo de analisar as relações entre palavras no contexto, a partir do estabelecimento de correlações entre a análise de conhecimento e a análise estatística.

2.2 Parâmetros externos

O programa de matematização demonstra um rompimento com padrões de pesquisa anteriores, significando a descontinuidade tanto numa abordagem linguística como nos processos de automatização da tradução e computacionais.

Do ponto de vista dos processos computacionais, o programa estruturou-se num cenário em que já se podia visualizar avanços nessa área, em comparação ao ambiente que se encontrava o programa de mecanização. Os novos recursos tecnológicos disponíveis ofereceram mais opções de linhas de pesquisa em TA, depois do período de suspensão dos estudos devido ao corte brutal nos investimentos, em virtude das exigências do relatório da ALPAC em 1966.

Do ponto de vista linguístico, os estudos em TA retomaram neste programa com uma característica científica tradicional, em outras palavras, com fundamentos científicos clássicos (Stein, 2013, p. 6). Por isso a abordagem em relação ao conhecimento linguístico tinha que ser mais abrangente e a visão de língua passa para o nível da sentença, uma vez que seria preciso ampliar a análise linguística para um patamar sintático e semântico.

<p>In order to react to the results of the ALPAC report and the reduction of resources, the discourse became more classically scientific and tried to integrate linguistic knowledge on a broader basis, above all, semantic analysis. The results that have been achieved by these approaches were promising and so, in the middle of the 1970s and in the course of the rapid development of technology and the introduction of the first personal computers, MT research was revitalized and headed to a continuously increasing popularity from the beginning of the 1980s. (STEIN, 2013, p. 6)</p>	<p>Para reagir aos resultados do relatório da ALPAC e à redução dos recursos, o discurso tornou-se mais classicamente científico e procurou integrar o conhecimento linguístico de forma mais ampla, sobretudo a análise semântica. Os resultados alcançados por essas abordagens foram promissores e, portanto, em meados da década de 1970 e no curso do rápido desenvolvimento da tecnologia e da introdução dos primeiros computadores pessoais, a pesquisa em MT foi revitalizada e se dirigiu a uma popularidade cada vez maior desde o início da década de 1980. (Tradução nossa)</p>
---	--

Em relação ao processo de tradução, a visão de língua como código, defendida por Weaver em 1947, mostrou-se simplista na abordagem do uso de técnicas de criptografia que funcionava bem para solução de outros problemas matemáticos, porém era inadequada para a tradução, conforme Stein (2013, p. 6).

Although the mathematical methods prove useful for cryptology, they turned out to be inadequate for more challenging and complex translation tasks. Accordingly, the subsequent systems that were developed were based on dictionaries and selectively used syntactic operations (this was the time, when J.J. Bechers article on the universal language was republished with the subtitle “A programming approach from the year 1661”). From today’s point of view, these approaches were remarkably naïve. (STEIN, 2013, p. 6)

Embora os métodos matemáticos sejam úteis para a criptologia, eles se mostraram inadequados para tarefas de tradução mais desafiadoras e complexas. Assim, os sistemas subsequentes que foram desenvolvidos foram baseados em dicionários e seletivamente usava operações sintáticas (esta foi a época em que o artigo de J.J. Bechers sobre a linguagem universal foi republicado com o subtítulo “Uma abordagem de programação do ano 1661”). Do ponto de vista de hoje, essas abordagens eram incrivelmente ingênuas. (Tradução nossa)

O programa de matematização desenvolveu-se no período em que os Estudos da Tradução consolidaram-se como disciplina (na década de 1980), constituindo quatro áreas de estudos (ver capítulo 1, p. 28), entre elas o estudo da tradução e linguística que estudava a organização dos elementos linguísticos do texto-fonte e do texto-alvo nos níveis da fonética e morfologia, do léxico e da sintaxe.

Em relação aos estudos da linguagem, o programa de matematização contou com apoio de teorias linguísticas, visto que era um campo em crescimento como disciplina científica. Assim, conflitos teóricos entre pesquisadores da TA e linguistas surgiram, uma vez que a visão de linguagem dentro das áreas dos estudos da linguagem e das ciências da computação tinha perspectivas diferentes e ainda em construção, em termos de interdisciplinaridade.

Em Wilks (1993, p. 3-4) esse conflito de perspectivas linguísticas entre as áreas pode ser observado no artigo em que ele compara o sistema Candide, desenvolvido pela IBM, com o sistema Systran. O autor defende a superioridade dos modelos da IBM em comparação a outras técnicas, porém questiona o conceito de linguística chomskiana adotado, uma vez que Chomsky era contra métodos estatísticos para TA.

I personally never believed Chomsky’s arguments in 1957 against other theories than his own any more than I did what he was for: his attacks on statistical and behaviorist methods (as on every thing else, like phrase structure grammars) were

Eu, pessoalmente, nunca acreditei nos argumentos de Chomsky em 1957 contra outras teorias que não as dele, mais do que o que ele defendia: seus ataques a métodos estatísticos e behavioristas (como em tudo mais, como gramáticas de estrutura de frase)

<p>always in terms of their failure to give explanations, and I will make no use of such arguments here, noting as I say that how much I resent IBMs use of “linguist” to describe everyone and anyone they are against. There is a great difference between linguistic theory in Chomsky’s sense, as motivated entirely by the need to explain, and theories, whether linguistic/AI or whatever, as the basis of procedural, application-engineering-orientated accounts of language. [...] Indeed, the IBM work to be described here has something in common with Chomsky’s views, which formed the post-1957 definition of “linguist”. It is clear from Chomsky’s description of statistical and Skinnerian methods that he was not of all opposed to relevance/pragmatics/semantics-free methods – he advocated them in fact – it was only that, for Chomsky, the statistical methods advocated at the time were too simple a method to do what he wanted to do with transformational grammars. More recent developments in finite state (as in Phrase Structure) grammars have shown that Chomsky was simply wrong about the empirical coverage of simple mechanisms. (WILKS, 1993, p. 3-4)</p>	<p>sempre foram relacionadas ao seu fracasso para dar explicações, e não farei uso de tais argumentos aqui, observando, ao dizer, o quanto me ressinto ao uso de “linguistas” pela IBM para descrever todos e qualquer pessoa contra quem eles são. Há uma grande diferença entre a teoria linguística no sentido de Chomsky, motivada inteiramente pela necessidade de explicar, e as teorias, sejam linguísticas de IA ou o que quer que seja, como a base de relatos de linguagem orientados para a engenharia de aplicação procedimental. [...] Na verdade, o trabalho da IBM a ser descrito aqui tem algo em comum com as visões de Chomsky, que formaram a definição pós-1957 de "linguista". Fica claro pela descrição de Chomsky dos métodos estatísticos e skinnerianos que ele não se opunha de todo aos métodos de relevância/pragmática/sem semântica - ele os defendia de fato - era apenas que, para Chomsky, os métodos estatísticos defendidos na época eram métodos muito simples para fazer o que ele queria fazer com as gramáticas transformacionais. Desenvolvimentos mais recentes em gramáticas de estado finito (como na Estrutura de Frase) mostraram que Chomsky estava simplesmente errado sobre a cobertura empírica de mecanismos simples. (Tradução nossa)</p>
--	--

Outra descontinuidade de padrão observada é a respeito da visão do processo de tradução em relação à combinação da automatização da tradução com o trabalho humano, questão criticada no programa de mecanização pela ALPAC, especialmente, por causa dos argumentos de Bar-Hillel sobre uma tradução totalmente automatizada (conceito de FAHQMT) (Coughlin, 1988, p. 2).

No programa de matematização a possibilidade de combinar o trabalho humano com a máquina passa ser aceita e investigada como uma linha de pesquisa, resultando no desenvolvimento de outros tipos de ferramentas para automatização da tradução que são as *CAT tools*.

<p>When reviewing MT in relation to AI strategies we need to distinguish several categories of computerized translation programs. These programs can be fully automatic, requiring no human intervention. They can also be interactive, meaning that human intervention takes place before,</p>	<p>Ao revisar a TA em relação às estratégias de IA, precisamos distinguir várias categorias de programas de tradução computadorizados. Esses programas podem ser totalmente automáticos, não exigindo intervenção humana. Eles também podem ser interativos, o que significa que a intervenção humana</p>
---	---

<p>during or after machine translation has occurred. The degree of interaction between the computer and the human translator varies greatly depending on the system. At the lower level of human intervention are systems called human-aided MT, while at the other end of the scale are computer-aided translation systems. If in the late fifties MT researchers were hoping to develop Fully Automatic High Quality Machine Translation (FAHQMT), it is generally admitted today that MT requires human intervention to be brought up to the quality level of work produced by a professional translator. (COUGHLIN, 1988, p. 2)</p>	<p>ocorre antes, durante ou depois da tradução automática. O grau de interação entre o computador e o tradutor humano varia muito, dependendo do sistema. No nível mais baixo de intervenção humana estão os sistemas chamados de TA auxiliada por humanos (HAT), enquanto na outra extremidade da escala estão os sistemas de tradução auxiliados por computador (CAT). Se no final dos anos 50 os pesquisadores de TA esperavam desenvolver a Tradução Automática de Alta Qualidade Totalmente Automática (FAHQMT), é geralmente admitido hoje que a TA requer intervenção humana para atingir o nível de qualidade do trabalho produzido por um tradutor profissional. (Tradução nossa)</p>
---	--

Essa mudança de perspectiva possibilita o desenvolvimento de ferramentas com base em corpus, visto que os pesquisadores perceberam que precisavam levar em consideração o trabalho do tradutor profissional, e para isso o material produzido por tradução humana poderia ser usado para melhorar o desempenho da máquina, conforme Witkam (2005, p. 2-3).

<p>Giving up on any knowledge of linguistics at all, the new turned to the model of a professional human translator! That was almost a revolution, certainly a paradigm change. Should we now interview experienced translators, accurately observe them at work, and guess how their brains are working? Even more simply: should we study the results of their work, their translated texts, of course with the originals beside them? Or even more convenient, let computers do that? An excellent idea! ((WITKAM, 2005, p. 2-3)</p>	<p>Desistindo de qualquer conhecimento de linguística, o novo voltou-se para o modelo de um tradutor humano profissional! Isso foi quase uma revolução, certamente uma mudança de paradigma. Devemos agora entrevistar tradutores experientes, observá-los com precisão no trabalho e adivinhar como seus cérebros estão funcionando? De forma ainda mais simples: deveríamos estudar os resultados de seu trabalho, seus textos traduzidos, é claro com os originais ao lado? Ou ainda mais conveniente, deixar os computadores fazerem isso? Excelente ideia! (Tradução nossa)</p>
---	--

O destaque para o papel do tradutor neste programa ocorre a partir da abertura do mercado de tradução, do crescimento pela demanda das grandes corporações por projetos nas áreas técnica e computacional, impulsionando a contratação de serviços oferecidos por tradutores *freelancers* que adotaram as *CAT tools* para melhorar seu desempenho. Com mais recursos computacionais disponíveis, o trabalho poderia ser realizado em menos tempo e com mais precisão (Beliaeva, 2001, p. 119). Assim, o mercado de soluções para automatização da tradução ampliou, alcançando também usuários finais, o que motivou empresas de *softwares*, em expansão na época,

investirem no desenvolvimento de soluções para um novo consumidor e não apenas para grandes corporações.

<p>Critical aspect of society development in modern world is its scientific and cultural interaction, in which case both information technologies as a whole and machine translation become the most important facilities. Nowadays, in the age of global communications, the need for a fast, accurate and cheap translation from one language into another has become very pronounced. [...] Furthermore, both fast and accurate translation means and translator opportunity to choose the proper means and to acquire special competence in their usage is the crucial point in implementing modern tools in translator work. (BELIAEVA, 2001, p. 119)</p>	<p>O aspecto crítico do desenvolvimento da sociedade no mundo moderno é a sua interação científica e cultural, caso em que tanto as tecnologias da informação como um todo e a tradução automática torna-se um dos recursos mais importantes. Hoje em dia, na era das comunicações globais, a necessidade de uma tradução rápida, precisa e barata de um idioma para outro tornou-se muito evidente. [...] Além disso, os meios de tradução rápidos e precisos, e a oportunidade do tradutor de escolher os meios adequados e de adquirir competência especial em seu uso é o ponto crucial na implementação de ferramentas modernas no trabalho do tradutor. (Tradução nossa)</p>
--	--

Os grupos de pesquisa deste programa (ver capítulo III) eram formados em centros de pesquisa de língua e tecnologia, em sua maior parte universidades, ao redor do mundo que contavam com apoio financeiro público e parcerias com empresas privadas que buscavam desenvolver suas próprias ferramentas *in-house* ou tinham interesse de comercialização. Alguns membros desses grupos fundaram, posteriormente, empresas independentes, como é o caso da empresa Systran, entre outras.

Witkam (2005, p. 11) aponta que o maior desafio para os grupos de pesquisa passou ser organizacional e não tecnológico, visto que era necessário investimento para as pesquisas evoluírem e alguns governos não investiriam nos grupos das universidades que passam a depender de organizações privadas com interesse comercial nas ferramentas.

<p>But the most critical factor, on which depends the breakthrough to high-quality translation machines, is organizational, not technological! The researchers, scattered in their universities, naturally fond of creating ever newer variations, rarely commit themselves to common and on-going work on one sole system. For the commercial world, general and high-quality translation systems are not sufficiently attractive, and an international government like that in Brussels is afraid to risk (again) big expenses for it. Lobbying and excellent</p>	<p>Mas o fator mais crítico, do qual depende o avanço para máquinas de tradução de alta qualidade, é organizacional, não tecnológico! Os pesquisadores, espalhados em suas universidades, gostam naturalmente de criar variações cada vez mais novas, raramente se comprometem com o trabalho comum e contínuo em um único sistema. Para o mundo comercial, sistemas de tradução gerais e de alta qualidade não são suficientemente atraentes, e um governo internacional como o de Bruxelas tem medo de arriscar (de novo) grandes investimentos</p>
---	---

<p>organization are necessary so that competent zealots can effectively join forces and carry out a difficult collaboration lasting many years. As a senior MT-researcher once stated [Carbonell 1992]: "in Machine Translation, what matters is persistence". (WITKAM, 2005, p. 11)</p>	<p>por isso. Lobby e excelente organização são necessários para que fanáticos competentes possam efetivamente unir forças e realizar uma difícil colaboração que durará muitos anos. Como um pesquisador sênior de TA uma vez afirmou [Carbonell 1992]: "na tradução automática, o que importa é a persistência". (Tradução nossa)</p>
--	--

Nishigaki (1999, p. 15) acreditava que o século XXI seria o período de abertura para o mundo do "multilinguismo" e do "multiculturalismo" em virtude do acesso à internet, e atribuía um papel fundamental da TA à essa função. A abertura do mercado de tradução promoveu maior visibilidade dos recursos de TA, que certamente tem contribuído para democratização do acesso à informação de maneira global a partir deste programa.

O Japão, por exemplo, explica Nishigaki (1999, p. 15), esteve fechado por anos sem ter acesso à cultura mundial e a TA trouxe a possibilidade de tal abertura com a tradução de romances, artigos acadêmicos, entre outros conteúdos, aproximando o país a um conhecimento que antes estaria apenas disponível para a elite que falasse inglês, uma vez que o inglês era o idioma dominante da internet.

<p>The recent rapid spread of the Internet suggests that soon everybody on earth will be able to communicate freely with each other across national borders. The 21st century will be the age of multilingualism and multiculturalism, when various languages and cultures are dynamically exchanged on a global scale. One may call it the New Great Age of Translation. In such an age, Machine Translation (MT) is obviously expected to play an essential role. [...] Japan is typical of those nation-states, where even English texts have few chances to be published. So people in Japan tend to live in a small closed world of Japanese communication. Accordingly, the level of cross-border communication remains unsatisfactory, even though it is eagerly encouraged by the current trend of globalization. [...] Our chances to read and/or write foreign texts increase drastically, which brings about high motivation among people to study foreign languages. Multilingualism on the Internet aims to encourage different peoples on earth to study foreign languages with each</p>	<p>O rápido crescimento recente da Internet sugere que em breve todos na terra serão capazes de se comunicar livremente entre si além das fronteiras nacionais. O século XXI será a era do multilinguismo e do multiculturalismo, quando várias línguas e culturas são trocadas dinamicamente em escala global. Podem chamá-la de Nova Grande Era da Tradução. Em tal época, obviamente se espera que a tradução automática (TA) desempenhe um papel essencial. [...] O Japão é típico desses estados-nação, onde até os textos em inglês têm poucas chances de serem publicados. Então, as pessoas no Japão tendem a viver em um pequeno mundo fechado da comunicação japonesa. Consequentemente, o nível de comunicação além fronteiras continua insatisfatório, embora seja fortemente encorajado pela tendência atual de globalização. [...] Nossas chances de ler e / ou escrever textos estrangeiros aumentam drasticamente, o que traz grande motivação para o estudo de línguas estrangeiras. O multilinguismo na Internet visa encorajar diferentes povos da terra a estudar línguas estrangeiras entre si, promovendo assim uma</p>
---	---

other, thus promoting true cross-cultural communication. (NISHIGAKI, 1999, p. 15)	verdadeira comunicação intercultural. (Tradução nossa)
---	--

Para Nishigaki (1999, p. 20) o desenvolvimento da TA no Japão fez parte de uma “metamorfose cultural” no país que se comparado, nas devidas proporções, ao que foi a tradução na era renascentista para a história ocidental ou a tradução de documentos ocidentais para países asiáticos no desenvolvimento das civilizações modernas, o avanço da TA nesse país proporcionou uma abertura cultural de pouco contato aos japoneses até aquele momento.

In short, the New Great Age of Translation is just around the corner. There have already been several such ages – in Renaissance era for example, a lot of Greek and Arabic documents were translated into Latin in Italy. This translation is considered to have set the base of the modern civilization. Then the modernization of Asian countries was brought about by another great age of translation - a lot of Western documents were translated into Asian languages in the 20 th Century. The development of MT technology in Japan is also a part of this cultural metamorphosis. (NISHIGAKI, 1999, p. 20)	Resumindo, a Nova Grande Era da Tradução está chegando. Já houve várias dessas idades - na era da Renascença, por exemplo, muitos documentos gregos e árabes foram traduzidos para o latim na Itália. Esta tradução é considerada como tendo estabelecido a base da civilização moderna. Em seguida, a modernização dos países asiáticos foi provocada por outra grande era da tradução - muitos documentos ocidentais foram traduzidos para as línguas asiáticas no século 20. O desenvolvimento da tecnologia de TA no Japão também faz parte dessa metamorfose cultural. (Tradução nossa)
---	--

O acesso a ferramentas de TA gratuitas, como Google Tradutor, Microsoft Translator e Babelfish, promoveram a popularização do recurso, no entanto, trouxe uma expectativa em relação a resultados espetaculares. Nishigaki (1999) defendia que a TA deveria ser aceita como um recurso com limitações, que enfrentava dificuldades muito complexas na tradução entre línguas, diferente da perspectiva da tarefa de um robô que pudesse entender a linguagem humana.

A perspectiva da TA como ferramenta social também pode ser observada na argumentação de Nishigaki (1999) que menciona uma “elite manipuladora” com conhecimento do inglês que dominava a internet.

Unless this Babel-situation is improved, most people will be locked up in the world of their mother tongue, with only the elites being able to manipulate English freely and monopolize Internet resources. The economical and cultural gap between the elites and ordinary people will cause social instability. Therefore everybody expects of MT to solve this tough problem. In order for	A menos que essa situação de Babel seja melhorada, a maioria das pessoas ficará presa no mundo de sua língua materna, com apenas as elites sendo capazes de manipular o inglês livremente e monopolizar os recursos da Internet. A lacuna econômica e cultural entre as elites e as pessoas comuns causará instabilidade social. Portanto, todos esperam que a TA resolva esse problema difícil. Para
---	---

<p>MT to meet this expectation, all of us - not only researchers and engineers but also general users - must change our attitudes. We must have a clear realistic vision, rather than seeking an omnipotent robot who understands human language. (NISHIGAKI, 1999, p. 19)</p>	<p>que a TA atenda a essa expectativa, todos nós - não apenas pesquisadores e engenheiros, mas também usuários em geral - devemos mudar nossas atitudes. Devemos ter uma visão clara e realista, ao invés de buscar um robô onipotente que entende a linguagem humana. (Tradução nossa)</p>
--	---

O programa de matematização trouxe uma visão contemporânea aos estudos da TA, que conseguiram evoluir consideravelmente em virtude dos recursos tecnológicos disponíveis e da atualização das expectativas em relação aos resultados que pretendia-se alcançar.

A evolução das ferramentas de TA e automatização da tradução possibilitou um crescimento da indústria da tradução, promovendo aquecimento econômico no setor que desenvolveu novos projetos de tradução e contratou mais serviços, tanto de tradutores *freelancers* como contratados nas empresas, assim como motivou a contratação de mão de obra para o desenvolvimento dessas ferramentas.

Com um mercado aquecido, a popularização da TA aumentou, especialmente, em virtude da disponibilidade de opções gratuitas que proporcionaram acesso à informação numa perspectiva global, impulsionada pela presença crescente da internet.

3. Programa de operacionalização

3.1 Parâmetros internos

O programa de operacionalização situa-se a partir da virada do milênio, nos anos 2000. As pesquisas de agora em diante concentravam-se na visão de língua a partir do significado, ocupando a posição de unidade linguística mais importante na análise semântica de corpus.

Uma das principais vantagens para este programa era a crescente disponibilidade de corpus que passa ser o principal instrumento no desenvolvimento das pesquisas em TA com base em IA. Os métodos de IA aplicados ao PLN priorizavam a abordagem da língua no nível da semântica, visto que a máquina realizava as análises com base em técnicas de RNAs e serviram de fundamento para o desenvolvimento de ferramentas com base neural.

O método de TA com base em corpus funcionava de maneira probabilística, combinando “modelos” aproximados de linguagem e tradução a partir da

aprendizagem no corpus bilíngue. Um algoritmo executava a tradução e formava a sentença, que era o resultado do valor mais alto de probabilidade (com base no modelo de linguagem), combinado com a tradução das palavras alinhadas e sua ordem (de acordo com o modelo de tradução), conforme explica Stein (2013, p. 9).

This ideal situation is confronted with the impossibility to access all sentences of a language. Therefore, SMT works with approximations, so-called models. A bilingual aligned corpus defines the translation model that represents all possible translations between two languages, i.e. the larger the translation model, the better the expected results. Generally, every word is considered a potential translation of all the others, but the probability is the highest for those that they are aligned to. An additional monolingual corpus of the target language is defined as the language model. It represents all valid sentences (or better words or word sequences which is a more operable abstraction) of a language. A search algorithm then determines the sentence, by finding the highest product of the values sentence validity (language model), word translation and word order (translation model). The result is the most probable translation. (STEIN, 2013, p. 9)

Essa situação ideal é confrontada com a impossibilidade de acessar todas as frases de uma língua. Portanto, a SMT trabalha com aproximações, os chamados modelos. Um corpus bilíngue alinhado define o modelo de tradução que representa todas as traduções possíveis entre duas línguas, ou seja, quanto maior o modelo de tradução, melhores são os resultados esperados. Geralmente, cada palavra é considerada uma tradução potencial de todas as outras, mas a probabilidade é maior para aquelas às quais estão alinhadas. Um corpus monolíngue adicional do idioma de destino é definido como o modelo de idioma. Ele representa todas as sentenças válidas (ou melhores palavras ou sequências de palavras, o que é uma abstração mais operável) de um idioma. Um algoritmo de pesquisa então determina a sentença, a partir do produto mais alto de validação da sentença em valores (modelo de linguagem), tradução de palavras e ordem de palavras (modelo de tradução). O resultado será a tradução mais provável. (Tradução nossa)

As técnicas com base neural associadas ao método de *deep learning* oferecem infinitas possibilidades de desenvolvimento de estratégias para aprendizagem da máquina. A ideia é que as estruturas de RNAs funcionem numa perspectiva equivalente ao cérebro humano, com capacidade para resolver problemas complexos com base em partes de informações que são montadas de maneira hierárquica.

Quando o conceito é aplicado à linguagem encontra complexidades em relação aos significados, uma vez que não são claros os limites de sentido nas relações estabelecidas entre palavras. Assim, as ambiguidades de significados continuavam um desafio a ser superado, motivando os experimentos a partir do conceito de redes neurais

O projeto da Universidade de Pohang (Coréia do Sul), por exemplo, estudou um método para melhorar o problema de ambiguidade na tradução automática do coreano para o japonês. O estudo mostrou possíveis motivos que causavam a dificuldade de identificação dessas ambiguidades pela máquina, como por exemplo,

as palavras ao redor da palavra central da sentença (no processo de *deep learning* são agrupadas palavras que aparecem em contextos similares a partir da definição de uma palavra com sentido central na sentença), entre outros fatores, de acordo com Chung et al. (2002, p. 1).

<p>The execution of our neural network model is based on the concept codes of a thesaurus. Most previous word sense disambiguation approaches based on neural networks have limitations due to their huge feature set size. By contrast, we reduce the number of features of the network to a practical size by using concept codes as features rather than the lexical words themselves. [...] Much research on word sense disambiguation has revealed that several different types of information can contribute to the resolution of lexical ambiguity. These include surrounding words (an unordered set of words surrounding a target word), local collocations (a short sequence of words near a target word, taking word order into account), syntactic relations (selectional restrictions), parts of speech, morphological forms, etc (McRoy, 1992, Ng and Zelle, 1997). (CHUNG et al., 2002, p. 1).</p>	<p>A execução do nosso modelo de rede neural baseia-se nos códigos de conceito de um tesouro. A maioria das abordagens anteriores sobre desambiguação de sentido de palavras está fundada em redes neurais que tem limitações devido ao seu enorme conjunto de recursos. No entanto, reduzimos o número de recursos da rede a um tamanho prático, usando códigos de conceito como recursos, em vez das próprias palavras lexicais. [...] Depois de muito estudo sobre a desambiguação do sentido das palavras, revelou-se que vários tipos diferentes de informações podem contribuir para a solução da ambiguidade lexical. Entre elas estão palavras circundantes (um conjunto não ordenado de palavras em torno de uma palavra-alvo), colocações locais (uma sequência curta de palavras perto de uma palavra-alvo, levando em conta a ordem das palavras), relações sintáticas (restrições seletivas), partes do discurso, formas morfológicas, etc (McRoy, 1992, Ng e Zelle, 1997). (Tradução nossa)</p>
---	---

Outra dificuldade seria o reconhecimento de palavras desconhecidas, isso porque um corpus linguístico nem sempre incluirá todas as possíveis palavras de uma língua, como por exemplo, vocabulário de áreas de conhecimento específico ou línguas raras, em que o volume de corpus disponível era menor.

Um estudo conduzido por pesquisadores de duas universidades (a Universidade de São Paulo, USP e a Universidade de Wolverhampton, na Inglaterra) em colaboração com o centro de pesquisas da empresa Xerox na Europa, investigou um método para solucionar a questão a partir do uso de paráfrases. O objetivo era incorporar um modelo de especialidade ao sistema, a fim de melhorar o desempenho de uma área de conhecimento específico.

A técnica com base no uso de paráfrases consistia na substituição de uma palavra desconhecida por uma de sentido equivalente, não necessariamente um sinônimo, visto que o processo baseava-se no treinamento da máquina a partir de um

corpus monolíngue, o WordNet, desenvolvido pela Universidade de Princeton nos EUA, conforme Aziz et al. (2010, p. 1).

<p>“Out-of-Vocabulary” (OOV) words, that is, words for which no translation is provided in the system phrase table. The problem is particularly severe when bilingual data are scarce or the text to be translated is not from the same domain as the data used to train the system. One approach consists in replacing the OOV word by a paraphrase, i.e. a word that is equivalent and known to the phrase-table. For instance, in the sentence “The police hit the protester”, if the source word “hit” is OOV, it could be replaced by its paraphrase “struck”. In previous work such paraphrases are learnt by “pivoting” through parallel texts involving multiple languages (Callison-Burch et al., 2006) or on the basis of monolingual data and distributional similarity metrics (Marton et al., 2009). Mirkin et al. (2009) go beyond the use of paraphrase to incorporate the notion of an entailed phrase, that is, a word which is implied by the OOV word, but is not necessarily equivalent to it — for example, this could result in “hit” being replaced by the entailed phrase “attacked”. Both paraphrases and entailed phrases are obtained using monolingual resources such as WordNet (Fellbaum, 1998). (AZIZ et al., 2010, p. 1).</p>	<p>Palavras “fora do vocabulário” (OOV), ou seja, palavras para as quais nenhuma tradução é fornecida na tabela de frases do sistema. O problema é particularmente grave quando os dados bilíngues são escassos ou o texto a ser traduzido não é do mesmo domínio que os dados usados para treinar o sistema. Uma abordagem consiste em substituir a palavra OOV por uma paráfrase, ou seja, uma palavra que é equivalente e conhecida na tabela de frases. Por exemplo, na frase “A polícia agrediu o manifestante”, se a palavra-fonte “agrediu” for OOV, ela poderia ser substituída por sua paráfrase “golpeou”. Em trabalhos anteriores, tais paráfrases são aprendidas por meio de “pivôs” a partir de textos paralelos envolvendo várias línguas (Callison-Burch et al., 2006) ou com base em dados monolíngues e métricas de similaridade distributiva (Marton et al., 2009). Mirkin et al. (2009) vão além do uso de paráfrase para incorporar a noção de uma frase implicada, ou seja, uma palavra que está implícita na palavra OOV, mas não é necessariamente equivalente a ela - por exemplo, isso poderia resultar em “agrediu” sendo substituído pela frase associada “atacar”. Tanto as paráfrases quanto as frases implicadas são obtidas usando recursos monolíngues como o WordNet (Fellbaum, 1998). (Tradução nossa)</p>
---	--

A técnica do uso de paráfrases foi empregada em outros estudos, como é o caso do projeto desenvolvido na Universidade de Dublin, que empregava o modelo probabilístico de língua com base neural (NPLM). Tratava-se de um módulo que agrupava palavras com funções semânticas e sintáticas similares, de maneira que uma palavra dependesse do contexto em questão. O resultado desse agrupamento eram paráfrases criadas com base no modelo de tradução, que incorporava as novas sentenças como novas entradas no módulo de combinação do sistema, conforme Okita (2012, p. 71).

Em Okita (2012, p. 66) podemos ver o processamento desse modelo probabilístico que aplicava a técnica de análise semântica e sintática para formar sequências de palavras menores como resultado das probabilidades. Essa técnica

ajudava diminuir a complexidade para a máquina decidir entre os resultados de maior probabilidade.

<p>A NPLM considers the representation of data in order to make the probability distribution of word sequences more compact where we focus on the similar semantical and syntactical roles of words. For example, when we have two sentences "The cat is walking in the bedroom" and "A dog was running in a room", these sentences can be more compactly stored than the n-gram language model if we focus on the similarity between (the, a), (bed room, room), (is, was), and (running, walking). Thus, a NPLM provides the semantical and syntactical roles of words as a language model. NPLM of Bengio et al. (2000) implemented this using the multi-layer neural network and yielded 20% to 35% better perplexity than the language model with the modified Kneser-Ney methods (Chen and Goodman, 1998). (OKITA, 2012, p. 66).</p>	<p>Um NPLM considera a representação de dados a fim de tornar a distribuição de probabilidades das sequências de palavras mais compacta, onde focalizamos nas funções semânticas e sintáticas semelhantes dessas palavras. Por exemplo, quando temos duas frases "O gato está andando no quarto" e "Um cachorro estava correndo em uma sala", essas frases podem ser armazenadas de forma mais compacta do que o modelo de linguagem de n-gramas se nos concentrarmos na semelhança entre (o, a), (quarto, sala), (é, era) e (correndo, caminhando). Assim, um NPLM fornece as funções semânticas e sintáticas das palavras como um modelo de linguagem. O NPLM de Bengio et al. (2000) implementado para isso, usou a rede neural multi-camada e representou de 20% a 35% em melhoria de perplexidade do que o modelo de linguagem a partir dos métodos modificados de Kneser-Ney (Chen e Goodman, 1998). (Tradução nossa)</p>
--	---

O programa de operacionalização é um programa em andamento, em que a visão de língua a partir do significado continua na abordagem das pesquisas. A análise semântica mostrou-se o grande interesse do programa que procurava pela integração do conhecimento linguístico à máquina, considerada ainda um desafio para os pesquisadores.

Em gerações anteriores de ferramentas, essa integração demonstrou pouca melhoria no significado, porém acredita-se que a integração com grandes bancos de dados semânticos possa solucionar alguns problemas de significado encontrados mesmo no modelo neural. A ferramenta Google Tradutor, por exemplo, usa o banco de dados semântico Wordnet, entre outras opções de recursos semânticos, dependendo do idioma. (POIBEAU, 2017, p. 104)

3.2 Parâmetros externos

As pesquisas no programa de operacionalização demonstravam uma maturidade em termos científico e tecnológico, em comparação aos programas anteriores. A disponibilidade de recursos tecnológicos proporcionou a implementação

de novas técnicas e métodos de TA que anteriormente esbarravam na limitação computacional.

A velocidade no desenvolvimento de novos métodos e na construção de novas ferramentas, bem como o processo de aprendizagem da máquina é consideravelmente menor neste programa (cf. Stein, 2013, p. 5), comparado ao programa de mecanização e até mesmo ao começo do programa de matematização.

<p>Until today, machine translation (MT) is known as the crowning discipline of natural language processing. Due to current MT approaches, the time needed to develop new systems with similar power to the older ones, has decreased enormously. (STEIN, 2013, p. 5)</p>	<p>Até hoje, a tradução automática (TA) é conhecida como a disciplina culminante do processamento de linguagem natural. Devido às abordagens atuais da TA, o tempo necessário para desenvolver novos sistemas com potência semelhante aos mais antigos diminuiu enormemente. (Tradução nossa)</p>
---	---

Os grupos de pesquisa neste programa encontram-se descentralizados, uma vez que os centros de pesquisa em TA hoje podem estar localizados em universidades ao redor do mundo mas também em organizações privadas tanto de empresas de grande porte, como pequenas; e em segmentos de desenvolvimento de *softwares*, fornecedores de serviços de tradução, etc.

Outra característica dos grupos de pesquisa a partir do programa de operacionalização é o patrocínio, que parece ser de maior interesse por parte das empresas privadas, com objetivos comerciais no desenvolvimento de ferramentas, do que de investimento público com fins científicos (cf. Witkam, 2005, p. 11).

Em relação aos Estudos da Tradução, a disciplina neste programa encontrava-se institucionalizada e promoveu a formação de tradutores tanto em cursos técnicos como universitários. No entanto, nos anos 2000 muitos cursos universitários ainda não ensinavam sobre o novo mercado de tradução emergente. Nos anos seguintes, em virtude da necessidade de preparação do tradutor para esse novo mercado de trabalho, algumas disciplinas relacionadas ao ensino de recursos de automatização da tradução, como as *CAT tools*, começaram timidamente serem incorporadas à formação de tradutores, mas ainda não ocupando espaço significativo em sua formação acadêmica, provocando impacto com suas expectativas ao se depararem com a realidade do mercado de trabalho, conforme Gough (2011, p. 1).

<p>Imagine you are a recent graduate of one of these courses. You even achieved a cum laude. You reflect on the knowledge you have gained, and the skills you have honed</p>	<p>Imagine que você acabou de se formar em um desses cursos. Você até conseguiu um certificado. Você reflete no conhecimento que adquiriu e nas habilidades que aperfeiçoou e</p>
--	---

<p>and mastered. You are now a 'seasoned' linguist, an expert in intercultural communication and a competent information miner. Exciting theories, approaches and strategies still reverberate in your head. As do lectures about text types, equivalence, cohesion and coherence. You can see lexical gaps everywhere and you know that there is no such thing as absolute synonymy. You may even have learned about CAT technology and tried out some of the software applications. You might even have some understanding of machine translation. You are about to enter a giant, complex and dynamic organism – the translation industry. But how much do you really know about its internal structure, its individual parts and the way they interact? How much do you really know about its evolution and current trends? How much do you know about translation as a business process? Have you learnt about the financial mechanisms, technology management or technical standards? A few months later you look back and think that perhaps what you learnt about navigating your boat on a still lake of theory on a sunny day of academic cloud hopping may not be sufficient to sail a super fast, technology ridden vessel through the rough, global waters of the modern translation industry. (GOUGH, 2011, p. 1).</p>	<p>domina. Você agora é um linguista "experiente", um especialista em comunicação intercultural e um competente minerador de informações. Teorias, abordagens e estratégias empolgantes ainda reverberam em sua cabeça. Assim como palestras sobre tipos de texto, equivalência, coesão e coerência. Você pode ver lacunas lexicais em todos os lugares e sabe que não existe sinonímia absoluta. Você pode até ter aprendido sobre a tecnologia das <i>CAT tools</i> e experimentado alguns dos aplicativos de <i>software</i>. Você pode até ter algum conhecimento de tradução automática. Você está prestes a entrar em um organismo gigante, complexo e dinâmico - a indústria da tradução. Mas quanto você realmente sabe sobre sua estrutura interna, suas partes individuais e a maneira como elas interagem? Quanto você realmente sabe sobre sua evolução e tendências atuais? Quanto você sabe sobre tradução como um processo de negócios? Você aprendeu sobre os mecanismos financeiros, gestão de tecnologia ou padrões técnicos? Alguns meses depois, você olha para trás e pensa que talvez o que você aprendeu sobre como navegar seu barco em um lago calmo da teoria em um dia ensolarado de saltos nas nuvens acadêmicas pode não ser suficiente para navegar em uma embarcação super rápida e movida por tecnologia através das águas inóspitas e globais da moderna indústria da tradução. (Tradução nossa).</p>
--	--

O programa de operacionalização proporcionou a ampliação do uso da TA para além das tradicionais áreas de tradução técnica, como localização de *softwares* e conteúdo especializado. Outras novas áreas começaram testar os recursos personalizáveis da TA, como por exemplo, os projetos de legendagem de filmes. Esse segmento que trabalha com uma linguagem informal, diferencia-se bastante do material técnico tipicamente traduzido automaticamente, para isso a personalização dos recursos e o direcionamento de corpus especializado mostrou-se fundamental para o desempenho da máquina.

O projeto desenvolvido na Universidade de Estocolmo descreve as estratégias no desenvolvimento de uma ferramenta de TA para legendagem de filmes do idioma sueco para o dinamarquês, a partir de uma arquitetura com base estatística que contou com o treinamento da máquina em um amplo corpus específico no gênero.

Uma das dificuldades encontradas foi exatamente o gênero, que conforme Volk e Harder (2007, p. 1-2) poderia ser considerado um conteúdo próximo ao literário, uma vez que lidava com a criatividade e espontaneidade da língua falada, bem como com diversos temas que poderiam aparecer nos filmes e, conseqüentemente, exigiriam vocabulários diferentes.

<p>We focus on the evaluation of the system output against independent and post-edited translations. The task of translating subtitles also differs from most other machine translation applications in that we are dealing with creative language and thus we are closer to literary translation than technical translation. This is obvious in cases where rhyming song-lyrics or puns are involved, but also when the subtitler applies his individual linguistic feeling to achieve a natural and appropriate wording which blends into the video without disturbing. Finally the language of subtitling covers many domains from educational programs on any conceivable topic to exaggerated modern youth language. (VOLK; HARDER, 2007, p. 1-2).</p>	<p>Nós nos concentramos na avaliação da produção do sistema em comparação com traduções independentes e pós-editadas. A tarefa de traduzir legendas também difere da maioria dos outros aplicativos de tradução automática, pois lidamos com uma linguagem criativa e, portanto, estamos mais próximos da tradução literária do que da tradução técnica. Isso é óbvio nos casos em que rimas de letras de músicas ou trocadilhos estão envolvidos, mas também quando o legendador aplica seu sentimento linguístico individual para alcançar um texto natural e apropriado que se mistura ao vídeo sem perturbar. Por fim, a linguagem da legendagem abrange muitos domínios, desde programas educacionais sobre qualquer tópico concebível até a exagerada linguagem juvenil moderna. (Tradução nossa)</p>
---	---

Outra área que despertou interesse da TA neste programa foi a tradução de conteúdo produzido na internet. Um estudo conduzido no Reino Unido (Jiang et al., 2012, p. 1) mostrou que o domínio do inglês na internet começou diminuir depois da Web 2.0, significando que os usuários passaram a desenvolver seus próprios conteúdos nos mais variados idiomas.

O dado apontado no estudo indica que em 2010 o inglês representava cerca de 27% de todas as publicações na web, justificando o motivo do investimento por parte dos maiores provedores de ferramentas gratuitas na internet (Google e Microsoft) incorporarem mais pares de idiomas às suas ferramentas, a fim de atender a demanda de um novo usuário da internet.

<p>Increasingly so, the influence of English as an Internet language is declining, so much so that recent data from June 2010 suggests that English-language users comprise just 27% of the overall Web population. Nonetheless, while Carrera et al. (2009) acknowledge that user-generated content is suitable for machine</p>	<p>Cada vez mais, a influência do inglês como idioma da Internet está diminuindo, tanto que dados recentes de junho de 2010 sugerem que os usuários do idioma inglês representam apenas 27% da população geral da web. No entanto, enquanto Carrera et al. (2009) reconhecem que o conteúdo gerado pelo usuário é adequado para tradução automática</p>
--	---

translation (MT), they also state that most such content usually remains untranslated. (JIANG et al., 2012, p. 1)	(TA), eles também afirmam que a maior parte desse conteúdo geralmente permanece sem tradução. (Tradução nossa)
---	--

O interesse pela tradução de conteúdo da internet parte não é apenas do usuário, mas também de empresas que buscam analisar o comportamento desses usuários da web, em especial, nas redes sociais. Esse conteúdo, apesar de apresentar linguagem informal e variados temas, demonstrou ser apropriado para ser traduzido automaticamente em virtude de sua finalidade de uso, na mineração de dados, para mapeamento comportamental. A vantagem de ferramentas de TA que podem ser personalizadas de acordo com a necessidade do material a ser traduzido, também ampliou a diversidade de conteúdos que podem ser processados na TA.

Além do interesse mercadológico, a TA mostrou benefícios no chamado conceito de “localização social” que significa a tradução de conteúdo estrangeiro por comunidades que dominam as habilidades de tradução, numa abordagem não mercadológica, ou seja, fora do domínio de corporações.

Um projeto da Universidade de Dublin procurou adaptar uma plataforma de tradução de uso livre para melhorar o desempenho da TA, com o objetivo de aprimorar esse recurso para projetos de tradução de empresas e organizações não-governamentais que contavam com tradutores voluntários (Du et al., 2015, p.57).

Social localisation is a kind of community action, which matches communities and the content they need, and supports their localisation efforts. The goal of social localisation-based statistical machine translation (SL-SMT) is to support and bridge global communities exchanging any type of digital content across different languages and cultures. Trommons is an open platform maintained by The Rosetta Foundation to connect non-profit translation projects and organisations with the skills and interests of volunteer translators, where they can translate, post-edit or proofread different types of documents. The concept of social localisation was proposed In Schäler (2011), which is based on the fact that currently large communities with language skills are ready to support good causes, and large amounts of content are accessible to communities that should be translated, but is not being done currently. Accordingly, social localisation aims to	A localização social é um tipo de ação da comunidade, que combina comunidades e o conteúdo de que precisam e apoia seus esforços de localização. O objetivo da tradução automática estatística com base em localização social (SL-SMT) é apoiar e estabelecer uma ponte entre comunidades globais que trocam qualquer tipo de conteúdo digital em diferentes idiomas e culturas. <i>Trommons</i> é uma plataforma aberta mantida pela <i>The Rosetta Foundation</i> para conectar projetos e organizações de tradução sem fins lucrativos com as habilidades e interesses de tradutores voluntários, onde eles podem traduzir, pós-editar ou revisar diferentes tipos de documentos. O conceito de localização social foi proposto em Schäler (2011), que se baseia no fato de que atualmente grandes comunidades com habilidades linguísticas estão prontas para apoiar boas causas, e grandes quantidades de conteúdo estão acessíveis às comunidades que deveriam ser traduzidas, mas não está sendo feito atualmente. Consequentemente, a localização
---	---

<p>match communities and slice content, and support the localization efforts. The main objective of social localisation is the promotion of a demand– rather than a supply–driven approach to localisation. It is based on the recognition that it is no longer exclusively the fact that the corporations control the global conversation, but rather communities. Social localisation supports user-driven and needs-based localisation scenarios, in contrast to mainstream localisation, driven primarily by short-term financial return-on-investment considerations. (DU et al., 2015, p.57)</p>	<p>social visa combinar comunidades e dividir conteúdo, bem como apoiar os esforços de localização. O principal objetivo da localização social é a promoção de uma demanda - ao invés de uma abordagem voltada para a oferta de localização. Baseia-se no reconhecimento de que não é mais exclusivo o fato de que as corporações controlam a conversa global, mas sim as comunidades. A localização social oferece suporte a cenários de localização orientados pelo usuário e com base nas necessidades, em contraste com a localização convencional, orientada principalmente por retorno financeiro de curto prazo sobre o investimento. (Tradução nossa)</p>
--	---

No programa de operacionalização a indústria da tradução mostrava-se consolidada. Com os serviços de tradução abrangendo diversas áreas, a profissionalização do trabalho do tradutor também foi alcançando novos patamares de exigência. A carreira tornou-se mais atrativa para jovens recém formados nos cursos de Tradução, Letras, Comunicação e Jornalismo.

A demanda pela contratação de profissionais de tradução originava-se tanto de organizações privadas (agências de tradução, multinacionais, empresas de *softwares*, etc.) como de órgãos públicos (agências do governo, organizações de relações exteriores e diplomáticas, etc.). No entanto, uma grande parcela desses profissionais instalaram-se como trabalhadores autônomos na categoria de *freelancers*, característica atrativa para profissionais com interesse numa carreira que proporcionasse maior autonomia.

De acordo com Gough (2011, p. 1), o relatório "*US News and World Report*" de 2011 apontou a carreira de tradutor como uma das 50 melhores daquele ano e a autora definia a indústria da tradução como um "organismo gigante, complexo e dinâmico".

<p>The translation industry is growing rapidly and, according to U.S. News and World Report, translation is one of the 50 best careers to pursue in 2011. Every year, thousands of students join translation courses attracted by such encouraging predictions, hoping they are placing their time, effort and money on a safe bet, hoping the chosen course will lead to a promising career. You are about to enter a giant, complex and dynamic organism – the translation industry. (GOUGH, 2011, p. 1)</p>	<p>A indústria da tradução está crescendo rapidamente e, de acordo com o <i>US News and World Report</i>, a tradução é uma das 50 melhores carreiras a seguir em 2011. Todos os anos, milhares de estudantes ingressam em cursos de tradução atraídos por tais previsões encorajadoras, na esperança de estarem dedicando seu tempo, esforço e dinheiro em uma aposta segura, na esperança de que o curso escolhido leve a uma carreira promissora. Você está prestes a entrar em um organismo gigante,</p>
--	---

	complexo e dinâmico - a indústria da tradução. (Tradução nossa)
--	---

A formação do tradutor enfrentava a questão tecnológica que exigia maior capacitação desse profissional que precisa constantemente se atualizar nos processos avançados de automatização da tradução. Aprender sobre as *CAT tools* e a TA tornou-se fundamental para o aprimoramento do trabalho do tradutor, visto que o profissional com entendimento nesses mecanismos poderá tanto desfrutar dos benefícios da tecnologia, como auxiliar na sua melhoria, a fim de aprimorar os resultados do seu próprio trabalho, segundo Gough (2011).

With the general awareness and ability to see a bigger picture comes the recognition that the changing nature of the translation industry creates demand for new translator skills and competencies. Perhaps the most significant factor affecting and shaping the portfolio of skills currently in demand is the galloping technological advancements and the automation processes this brings about. Machine translation, imperfect as it is, is here to stay, but will require human assistance to move forward. Are we translators being properly educated about this technology, about its enormous potential and its inherent limitations? How much do we know about the different types of machine translation, their usability for different text types and language pairs? Are new translators being prepared to handle this technology and deal with its aftermath? (GOUGH, 2011, p. 1)	Com a consciência geral e a capacidade de ver uma imagem maior, vem o reconhecimento de que a natureza mutante da indústria da tradução cria demanda por novas habilidades e competências do tradutor. Talvez o fator mais significativo que afeta e molda o portfólio de habilidades atualmente em demanda sejam os avanços tecnológicos galopantes e os processos de automação que isso acarreta. A tradução automática, por mais imperfeita que seja, veio para ficar, mas exigirá ajuda humana para avançar. Nós, tradutores, estamos sendo devidamente educados sobre essa tecnologia, sobre seu enorme potencial e suas limitações inerentes? Quanto sabemos sobre os diferentes tipos de tradução automática, sua usabilidade para diferentes tipos de texto e pares de idiomas? Os novos tradutores estão sendo preparados para lidar com essa tecnologia e lidar com suas consequências? (Tradução nossa)
---	--

Algumas características recomendadas para o novo perfil do tradutor, em uma publicação da revista TAUS de 13 de outubro de 2010, incluía sugestões sobre como usar ferramentas gratuitas, colaborar com outros tradutores no sentido de compartilhamento de recursos (como memórias de tradução) e abrir-se para aprender sobre novas tecnologias.

First, think 'open'. Governments are opening up their data, software companies are opening up source codes (even SDL Trados recently), big organizations like the EU are opening up their language resources. As the tenets of open innovation permeate industries,	Primeiro, pense em "abrir". Os governos estão abrindo seus dados, as empresas de softwares estão abrindo códigos-fonte (até mesmo o SDL Trados recentemente), grandes organizações como a UE estão abrindo seus recursos linguísticos. Como os princípios da inovação aberta permeiam as indústrias, o 'aberto' se
---	--

<p>'open' has become a key word. But more importantly, think 'open' in terms of trends and technology. Research shows that bias against MT or crowdsourcing is often lined with insufficient knowledge about the processes involved, accompanied by the fear of the unknown. Only an open, positive attitude and unbiased perspective can reveal the potential benefits these new technologies and processes can offer you. Second, think 'open tools'. Think that technology is your ally, not a foe. Imagine your one-man orchestra could become a band of skilled musicians playing a big piece for a large audience – without a director. There are freely available tools out there which will allow you to do just that – and because they are free to use, you are freed from expensive licensing and initial investment. Yes, they are not rich in features as some proprietary tools, but they are getting better and by their very nature they are being constantly developed by user communities. Of course there is some tinkering required with file conversions, but there is a lot of community support available and, with a little effort, this can be overcome. (TAUS, 2010, p. 1)</p>	<p>tornou uma palavra-chave. Mas o mais importante, pense "aberto" em termos de tendências e tecnologia. Pesquisas mostram que o preconceito contra a TA ou <i>crowdsourcing</i> está muitas vezes ligado com o conhecimento insuficiente sobre os processos envolvidos, acompanhado pelo medo do desconhecido. Somente uma atitude aberta e positiva e uma perspectiva imparcial podem revelar os benefícios potenciais que essas novas tecnologias e processos podem oferecer a você. Em segundo lugar, pense em "ferramentas gratuitas". Pense que a tecnologia é sua aliada, não uma inimiga. Imagine que sua orquestra de um só homem pudesse se tornar uma banda de músicos qualificados tocando uma grande peça para um grande público - sem um diretor. Existem ferramentas disponíveis gratuitamente que permitirão que você faça exatamente isso - e como elas são de uso gratuito, você está livre de licenciamento caro e investimento inicial. Sim, elas não são ricas em recursos como algumas ferramentas proprietárias, mas estão ficando cada vez melhores e, por sua própria natureza, estão constantemente em desenvolvimento por comunidades de usuários. É claro que é necessário fazer alguns ajustes nas conversões de arquivos, mas há muito suporte da comunidade disponível e, com um pouco de esforço, isso pode ser superado. (Tradução nossa)</p>
--	--

O futuro da profissão é considerado promissor, conforme Cough (2011). A demanda por tradução no mundo cresceu e continuará crescendo de maneira vertiginosa. As empresas que no passado traduziam seus produtos em 10 ou 20 idiomas, hoje precisam se preparar para traduzir em uma centena de idiomas ou poderão perder posição diante da concorrência.

Essa demanda por tradução nunca vista demonstra ser favorável para a profissão de tradutor, porém pode trazer frustração do ponto de vista do conceito de "ser tradutor" de décadas anteriores. No entanto, para o tradutor que estiver aberto à inovação, o avanço que presenciamos no programa de operacionalização pode representar um dos melhores momentos para a profissão.

<p>Change is the name of the game. And reinventing the profession is extremely hard if your days are spent just getting the jobs done and trying to make a modest living.</p>	<p>Mudança é o nome do jogo. E reinventar a profissão é extremamente difícil se você passa os dias apenas fazendo o trabalho e tentando ter uma vida modesta. No entanto,</p>
---	---

<p>Yet, for the first time in the history of the planet, translation is a really strategic activity. Thanks to Google Translate, Yahoo! Babelfish and Microsoft Bing, every soul on our planet now knows what translation means. Hundreds of millions people press the translate button every day which makes them realize how difficult it is to get a good, accurate translation. As professionals we must realize that our community is far too small (just 250,000 or so professional translators in a world of 6,000 languages?) to serve the needs of seven billion citizens. (COUGH, 2011, p.1)</p>	<p>pela primeira vez na história do planeta, a tradução é uma atividade realmente estratégica. Graças ao Google Translate, Yahoo! Babelfish e Microsoft Bing, cada alma em nosso planeta agora sabe o que significa tradução. Centenas de milhões de pessoas pressionam o botão traduzir todos os dias, o que as faz perceber como é difícil obter uma tradução boa e precisa. Como profissionais, devemos perceber que nossa comunidade é muito pequena (apenas 250.000 ou mais tradutores profissionais em um mundo de 6.000 idiomas?) para atender às necessidades de sete bilhões de cidadãos. (Tradução nossa)</p>
--	---

O interesse pelo uso de ferramentas de TA cresceu no programa de operacionalização, como mostra a pesquisa conduzida pela *CSA Research* de 2016. Os entrevistados foram empresas globais, *LSPs* (*Language Service Providers*, Provedores de Serviços de Idiomas) e tradutores *freelancer* que foram questionados sobre sua experiência com o uso de ferramentas de TA.

Segundo Lommel (2016), o resultado demonstrou um crescimento no uso da TA que pôde ser interpretado como uma prática cotidiana das empresas e projetava um crescimento ainda maior nos 3 anos seguintes. O receio em relação à máquina diminuir a demanda por serviços de tradutores humanos mostrou-se um mito, uma vez que em virtude da crescente necessidade por tradução global o trabalho de tradução humana, inclusive, demonstrou tendência de crescimento.

O estudo revelou ainda que pelo menos num espaço de 10 anos os entrevistados não acreditavam numa TA pura, ou seja, sem intervenção humana em algum dos processos, apesar da TA continuar transformando a perspectiva do mercado de quem trabalha com linguagem.

<p>Our research shows that MT is profoundly changing the landscape of the language industry. However, it will not eliminate the need for second languages or replace translators – certainly not any time soon and not within the ever-distant 10-year horizon beloved by tech journalists. The findings may not be as sexy as claims about how untested or experimental technology will up-end the whole way we interact across languages, but they present a much more realistic picture of the role that technology has to play. (LOMMEL, 2016, p. 1)</p>	<p>Nossa pesquisa mostra que a TA está mudando profundamente o perfil da indústria de idiomas. No entanto, não eliminará a necessidade do segundo idioma ou substituirá tradutores - certamente não tão cedo e não dentro do horizonte sempre distante de 10 anos, amado por jornalistas de tecnologia. As descobertas podem não ser tão atraentes quanto as afirmações sobre como a tecnologia não testada ou experimental vai superar toda a maneira como interagimos entre as línguas, mas elas apresentam</p>
--	---

	uma imagem muito mais realista do papel que a tecnologia tem que desempenhar. (Tradução nossa)
--	--

No programa de operacionalização a adoção da TA tomou uma perspectiva global. A abertura do mercado de tradução, bem como da disponibilidade de ferramentas gratuitas promoveu a TA em nível mundial. Em razão do crescimento sem precedentes do compartilhamento de conteúdo na internet, a necessidade por tradução aumentou num volume nunca visto e o apoio das ferramentas de TA para atender essa demanda demonstrou significativa aceitação.

O uso de TA é realidade no cotidiano de usuários da internet e de empresas, confirmando uma tendência de que a TA tornou-se parte fundamental e integrante na facilitação da comunicação mundial. O avanço tecnológico abre oportunidades de utilização da TA nas mais variadas áreas e possibilita o desenvolvimento de novas soluções no campo, isso motivado por um interesse mercadológico na comercialização de ferramentas de TA que levou a uma descentralização dos grupos de pesquisa, os quais alcançaram ramificações globais.

O impacto do avanço das tecnologias deste programa no papel do tradutor precisa ser visto do ponto de vista de aprimoramento no trabalho e da possibilidade de abertura para novas oportunidades na carreira, pois assim como em tantas outras áreas, a inovação faz parte da evolução natural do trabalho humano, e no caso da linguagem ainda precisa superar obstáculos, mas tem alcançado resultados surpreendentes.

CONSIDERAÇÕES FINAIS

O primeiro objetivo desta pesquisa foi revisitar alguns dos principais episódios da história da tradução, com destaque para o desenvolvimento de teorias, métodos e ferramentas, a fim de se buscar evidências sobre o surgimento da TA. A tradução é uma prática clássica com registros de trabalhos desde a Roma antiga de Cícero (106-43 a.C.) e Horácio (65-8 a.C.), passando pela tradução de São Jerônimo (da bíblia do hebreu em 390 d.C.) até chegarmos ao século XVIII com os primeiros estudos acerca de uma teoria da tradução, revelando um período em que a prática se dava pela simples experiência do ato de traduzir.

No entanto, foi no século XX que presenciamos a consolidação de uma área de estudos da tradução, com o desenvolvimento de um vocabulário específico para o campo a partir da década de 1940, assim como de pensamentos para introdução de uma teoria da comunicação e da linguística estrutural nos estudos da tradução, década em que também surgiu o interesse pela mecanização da tradução e os primeiros experimentos em TA. Aproximadamente duas décadas depois, uma visão metafísica é introduzida aos estudos da tradução e interpretação, que alcançaria seu reconhecimento como disciplina numa abordagem interdisciplinar dentro dos estudos da linguagem no final dos anos de 1970.

Essa consolidação como disciplina contou com teorias da tradução desenvolvidas por estudiosos contemporâneos, em que se destacam os trabalhos de Even-Zohar e G. Toury (Teoria dos Polissistemas), A. Lefevere (definição de um postulado para a disciplina), J. Holmes (definição dos estudos descritivos e estudos teóricos da tradução), A. Pym (arqueologia da tradução) e L. Venuti (modelos de tradução instrumental e de hermenêutica).

Como segundo objetivo, analisamos as razões do surgimento da TA na história da tradução, que se deu no final dos anos de 1940, num período em que os estudos da tradução ainda começavam se estruturar. A necessidade de traduzir de maneira rápida um grande volume de material científico foi a motivação inicial do investimento em pesquisas para desenvolver mecanismos de TA. O problema a ser resolvido ambientava-se num cenário pós-Segunda Guerra Mundial (1939-1945) em que a

necessidade, em especial dos EUA, de adquirir conhecimento sobre os estudos científicos de seus potenciais inimigos (naquele momento principalmente a URSS), motivou uma corrida no desenvolvimento de máquinas de tradução. A TA foi idealizada ao mesmo tempo em que os primeiros protótipos, do que futuramente seriam os computadores, começaram a ser desenvolvidos. Essas máquinas inicialmente projetadas para resolver operações matemáticas precisavam ser adaptadas para o processamento da linguagem natural, dando origem a disciplinas como a linguística computacional. Apesar do investimento em pesquisas por pouco mais de duas décadas, as limitações tecnológicas impediram um avanço significativo dos modelos de TA naquele período.

No terceiro objetivo proposto, buscamos interpretar historiograficamente as mudanças dos modelos de TA, bem como de visão de língua, em cada um dos programas de investigação da TA propostos pela autora desta dissertação de maneira inédita; programas esses definidos com base nos conceitos de programas de investigação propostos por Swiggers (1981) para os estudos da HL. Assim, os programas de investigação definidos especificamente para reconstruir uma história da TA nesta dissertação foram: os programas de mecanização, de matematização e de operacionalização.

O programa de mecanização apresentou um modelo de TA com base em regras, no qual adotava-se a perspectiva de uma gramática de estrutura sistematizada, priorizando uma visão de língua no nível da palavra, o que privilegiava uma tradução com base na incidência no léxico e de processamento mecânico, característica que deu origem ao nome desse programa.

No programa de matematização, a ruptura com padrões teóricos anteriores proporcionou o desenvolvimento de um modelo de TA com base estatística, que apoiado ao uso de corpus revolucionou os estudos em TA, definindo-se como um programa de base matemática por causa de sua abordagem probabilística. Nesse modelo, a visão de língua, influenciada por um conceito de gramática gerativa chomiskiana, passa a se apoiar no nível da sentença.

Por último, no programa de operacionalização, a possibilidade de uma tradução automatizada a partir do conceito da língua em uso, só seria possível em virtude do avanço computacional que viabilizou um modelo de TA com base neural. Nesse caso, treinava-se a máquina em um corpus a partir de uma abordagem do significado da função de uma unidade linguística na sentença, permitindo que a máquina aprendesse

a traduzir no nível da semântica. Esse modelo proporcionou um processo operacional mais eficiente que possibilitaria a integração com diversos recursos computacionais.

No quarto objetivo, procuramos relacionar as mudanças evidenciadas no terceiro objetivo aos aspectos históricos e sociais que serviram de apoio para as interpretações sobre os desenvolvimentos históricos nas práticas de tradução. Para isso lançamos mão da análise dos parâmetros externos que influenciaram os grupos de pesquisas, os quais foram responsáveis por cada programa de investigação da TA. Os programas foram se modificando de uma visão de língua a partir da palavra até atingirem a visão de texto, e isso foi ocasionado pelo clima intelectual onde circulavam as ideias linguísticas de cada programa.

Esse clima de opinião é provado pela mudança no perfil dos grupos de pesquisa ao longo das décadas; grupos que se iniciaram quase que exclusivamente no ambiente acadêmico e, em virtude de uma abertura do mercado de tradução, começaram a migrar para dentro de corporações, resultando em um negócio com objetivos comerciais consolidado num ambiente corporativo. A análise crítica de cada programa de investigação da TA auxiliou no entendimento das rupturas com padrões teóricos de um programa para outro, mostrando uma história de descontinuidades no desenvolvimento da TA, assim como nos estudos da tradução.

No quinto objetivo, elaboramos uma síntese interpretativa historiográfica sobre a presença das tecnologias no contexto da história da tradução, que foi comprovada a partir da consolidação de uma área de estudos da tradução, na qual se estimulou a formação de tradutores pela promoção de cursos técnicos e de nível superior na área. Isso tudo impulsionado por um mercado de tradução emergente entre as décadas de 1980 e 1990. Com a indústria da tradução se consolidando, em especial da tradução técnica e científica, tornou-se mais necessário o desenvolvimento de recursos para apoiar o trabalho de tradução, uma vez que se lidava com grandes volumes de conteúdo a ser traduzido, bem como novos projetos de tradução. Assim, na década de 1990 presenciamos o desenvolvimento das *CAT tools* e o avanço significativo no modelo de TA, que passaria a ser desenvolvido com base estatística e, posteriormente, a partir dos anos 2000, com base neural e *deep learning*.

No sexto e último objetivo definido para esta pesquisa, elaboramos uma reflexão sobre o impacto social do uso das ferramentas de tradução na sociedade, que demonstrou geração de trabalho e renda para tradutores, assim como o desenvolvimento de *softwares* para a área. Esses movimentos estimularam a

contratação de mais mão-de-obra para produção de programas, bem como para venda dos serviços de tradução. Outro aspecto social observado é em relação ao acesso à informação proporcionado pelo uso de ferramentas de TA gratuitas, que descentralizou a oportunidade de conhecimento apenas aos que dominavam determinados idiomas, promovendo a divulgação cultural, política e comercial de outros países.

Do ponto de vista filosófico, abre-se uma discussão sobre até que ponto o ser humano estaria dominado pela tecnologia, numa perspectiva em que a TA se definiria como um processo de geração de uma “linguagem artificial”. Neste momento, o perfil do tradutor começa ser questionado, com especulações sobre a máquina substituir o trabalho humano. No entanto ao longo deste estudo, a partir da visão de tradutora da autora desta dissertação, com experiência no uso de TA e ferramentas de mecanização da tradução, observamos que a mudança esperada no perfil de nós tradutores está no nível intelectual, em que nosso trabalho passa a alcançar novas perspectivas de atuação e especialização, transferindo para a máquina as tarefas repetitivas e operacionais.

A pergunta que motivou o desenvolvimento desta pesquisa (*qual a natureza essencial da evolução histórica das ferramentas tecnológicas de tradução automática e quais são seus impactos na sociedade?*) foi respondida. Na primeira parte, as evidências confirmadas na análise dos programas de investigação da TA provaram que o avanço tecnológico foi o principal motivo para o desenvolvimento de ferramentas de TA, nas quais apenas obtiveram êxito quando associadas a métodos e técnicas computacionais eficientes. Na segunda parte, observamos o impacto na sociedade em duas perspectivas, a de progresso da indústria da tradução e a de democratização do acesso à informação, essa última proporcionada pela disponibilidade de ferramentas de TA gratuitas na internet. Finalmente, graças à divulgação promovida pelo acesso global a ferramentas de TA na internet, a tradução tornou-se um tema de conhecimento comum para a sociedade e não apenas uma atividade exclusivamente intelectual e elitizada.

Assim, os processos de TA desenvolvidos em seus programas de investigação serviram de embasamento para compreender a natureza qualitativa desses mesmos processos e desenvolver uma análise sobre a visão de tradução adotada em cada programa, bem como os impactos na sociedade ao longo das décadas em virtude da automatização da tradução.

O perfil do tradutor no decorrer do avanço da mecanização da tradução e das ferramentas de TA teve de ser reinventado e ainda passa por transformações, em virtude das exigências de um mercado de tradução, agora consolidado, que busca por profissionais qualificados com novas habilidades. Por isso, o papel do tradutor passou a ser de interesse para futuros estudos, uma vez que levantam inquietações e questionamentos sobre o futuro da profissão, dos cursos técnicos e universitários, das novas habilidades esperadas pelo mercado e, finalmente, do interesse pela carreira de tradutor.

Por fim, com o objetivo de colocar em prática alguns dos recursos de TA discutidos nesta pesquisa, os textos provenientes do corpus analisado que foram traduzidos no capítulo IV contaram com apoio da ferramenta Google Tradutor, que mostrou-se bastante efetiva numa tradução de textos de naturezas distintas e épocas diferentes, servindo como instrumento para evidenciar a importância do aprofundamento da investigação sobre TA dentro do campo das humanidades. Com isso, esta pesquisa tem o objetivo de inaugurar uma linha de pesquisa pouco abordada, contribuindo com conceitos inéditos e específicos para o estudo da história da TA.

GLOSSÁRIO

Augmented Transition Network, Redes de Transição Ampliadas: sistema de representação gráfica usado na definição operacional de línguas formais e empregado, especificamente, na análise de línguas naturais.

CAT tools, Ferramentas de Tradução Assistida por Computador: programas ou softwares que auxiliam na tradução de textos, segmentando textos em sentenças para serem traduzidos de forma humana ou automática, com auxílio de memória(s) de tradução, banco de dados terminológicos e recursos de tradução automática.

Deep learning, Aprendizagem profunda: método de aprendizado de máquina com base em algoritmos de redes neurais artificiais que utiliza várias camadas de processamento de dados.

Ferramenta de tradução automática: programa ou *software* de tradução automática.

Inteligência artificial: campo interdisciplinar que estuda mecanismos computacionais, tais como, programas e softwares, entre outros, que executam tarefas as quais necessitariam de habilidades humana para execução. Os processos incluem a aprendizagem da máquina para execução de tarefas tidas, anteriormente, apenas como humanas.

Língua natural: é o formato puro da língua humana; formato de voz ou texto.

Linguística computacional: campo interdisciplinar que atua em conjunto com as áreas de inteligência artificial, estatística, linguística, entre outras, com o objetivo de utilização de processos computacionais para processamento e manipulação da linguagem humana.

Localização de softwares: processo de tradução de um programa que inclui não apenas a tradução textual da interface, mas as etapas que incluem a adaptação da interface para atender requisitos específicos de uma língua ou sistema linguístico.

Mainframe: é uma plataforma integrada de computadores capaz de processar grandes volumes de dados em um curto espaço de tempo. Essas plataformas oferecem serviços a milhares de usuários por meio de terminais conectados diretamente ou em rede. Usam programas modernos para o processamento de informações e são adotados por grandes corporações.

Machine learning, Aprendizagem de máquina: método que utiliza algoritmos para treinar um sistema desempenhar tarefas automaticamente com base na análise de dados.

Neural machine translation, Tradução de máquina neural: motor de tradução que utiliza modelos computacionais de base neural.

Parser (analisador): recurso computacional para realização da análise sintática e semântica na etapa de análise de sequências de caracteres (*strings*) e símbolos em PLN, linguagem de programação e estruturas de dados.

Parsing (análise): processo realizado pelo analisador sintático e semântico. Ver *parser*.

Phrase-based machine translation, Tradução de máquina com base em frase: motor de tradução que utiliza modelos computacionais com base em combinação de frases.

Plataforma: mesmo que ferramenta ou programa no contexto da ciência da computação.

Processamento de linguagem natural: área dentro da ciência da computação que atua em conjunto com as áreas de pesquisa sobre inteligência artificial e linguística na investigação de problemas gerados na automatização de línguas naturais.

Redes neurais artificiais: modelos computacionais desenvolvidos com base no modelo de funcionamento do sistema nervoso cerebral do ser humano.

Rule-based machine translation, Tradução de máquina com base em regras: motor de tradução que utiliza modelos computacionais com base em regras.

Statistical machine translation, Tradução de máquina com base em estatística: motor de tradução que utiliza modelos computacionais de base estatística.

Software open-source, Programa de acesso livre: também conhecido como “software livre” é quando um programa de computador é desenvolvido no conceito de código de programação aberto, em que pode ser usado por outros desenvolvedores de maneira livre. Isso facilita a integração com outros programas e recursos. O movimento de software livre ganha força a partir dos anos 2000 e podem ser encontrados programas desenvolvidos a partir desse conceito em praticamente todas as áreas da computação.

Tradução automática: tradução realizada por máquina, sem interação humana.

Tradução de máquina: ver Tradução automática.

Word-based machine translation, Tradução de máquina com base em palavra: motor de tradução que utiliza modelos computacionais com base em combinação de palavras.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALTMAN, C. *A pesquisa linguística no Brasil (1968-1988)*. 2ª ed., São Paulo: Humanitas FFLCH/USP, 2004.
- _____. *História, estórias e historiografia da linguística brasileira*. Todas as Letras, v.14, n. 1, p. 14-37, São Paulo: 2012.
- ARNOLD, D., BALKAN, L., MEIJER, S., HUMPHREYS, R. L., & SADLER, L. *Machine Translation: an introductory guide*. London: NCC Blackwell Ltd, 1994.
- AUROUX, S. *A filosofia da linguagem*. (Tradução: José Horta Nunes) Campinas, SP: Editora da Unicamp, 1998.
- BASSNETT, S. *Translation studies*. 3rd ed., USA and Canada: Routledge, 2002.
- BATISTA, R. *Introdução à historiografia da linguística*. São Paulo: Cortez, 2013.
- _____. *Historiografia da linguística*. São Paulo: Editora Contexto, 2019.
- HOLMES, J. *The name and nature of translation studies*. Third international congress of applied linguistics (pp. 66-80). Copenhagen: Rodopi, 1972.
- HUTCHINS, W. *Machine translation over fifty years*. Histoire Épistémologie Langage, v. 23, n.1, p. 7-31, 2001.
- KOERNER, E. *Quatro décadas de historiografia linguística: estudos selecionados*. (Tradução: Cristina Altman et al.) Vila Real: Universidade de Trás-os-Montes e Alto Douro, 2014.
- MILTON, J. *Tradução: teoria e prática*. 3ª ed., São Paulo: Martins Fontes, 2010.
- OUSTINOFF, M. *Tradução: história, teorias e métodos*. (Tradução: Marcos Marcionilo) São Paulo: Parábora, 2003.
- MURRAY, S. *Theory groups and the study of language in north america: a social history*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 1994.
- POIBEAU, T. *Machine translation*. London: The MIT Press, 2017.
- PYM, A. *Method in translation history*. 2nd ed., London and New York: Routledge, 1998.
- _____. *Humanizing translation history*. Hermes - Journal of Language and Communication Studies, v. 42, p. 23-48, 2009.
- RUTHERFORD, J. *O terceiro espaço: uma entrevista com Homi Bhabha*. Revista do Patrimônio Histórico e Artístico Nacional, p. 34-41, 1996.
- SOUZA, L. M. *Hibridismo e tradução cultural em Bhabha*. Margens da cultura: mestiçagem, hibridismo & outras misturas. 2004.

SWIGGERS, P. *The history-writing of linguistics: a methodological note*. General Linguistics, v. 21, n. 1, p. 11-16, 1981.

_____. *Linguistic historiography: a metatheoretical synopsis*. Todas as Letras, v. 19, n. 2, p. 73-96, 2017.

_____. *Linguistic historiography: object, methodology, modelization*. Todas as Letras, v.14, n.1, p. 38-53, 2012.

VENUTI, L. *The translation studies reader*. 3rd ed., London and New York: Routledge, 2012.

WILLIAMS, J., & CHESTERMAN, A. *The map: a beginner's guide to doing research in translation studies*. Manchester: St. Jerome Publishing, 2002.

Fontes primárias:

(ALPAC). *Language and machines: computers in translation and linguistics*. Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington D.C., 1966. Acesso em 25 de Novembro de 2020, disponível em <http://www.mt-archive.info/50/ALPAC-1966.pdf>

(TAUS). *What options do translators really have?* TAUS, 13 de Outubro de 2010. Acesso em 17 de Dezembro de 2020, disponível em <https://www.taus.net/think-tank/articles/translate-articles/what-options-do-translators-really-have?highlight=WyJtdCIslm10J3MiLCInbXQiXQ%3D%3D>

_____. *The future for translators looks bright, but they will have to reinvent the profession first*. TAUS, 13 de Setembro de 2011. Acesso em 17 de Dezembro de 2020, disponível em <https://www.taus.net/think-tank/articles/event-articles/the-future-for-translators-looks-bright-but-they-will-have-to-reinvent-the-profession-first?highlight=WyJtdCIslm10J3MiLCInbXQiXQ%3D%3D>

AZIZ, W.; DYMETMAN, M.; MIRKIN, S.; SPECIA, L.; CANCEDDA, N.; DAGAN, I. *Learning an expert from human annotations in statistical machine translation: the case of out-of-vocabulary words*. 14th Annual conference of the European Association for Machine Translation. Saint-Raphaël, 2010. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/10/EAMT-2010-Aziz.pdf>

BAR-HILLEL, Y. *The present state of research on mechanical translation*. American Documentation, v. 2, n.1, p. 153-165, 1951. Acesso em 23 de Novembro de 2020, disponível em <http://www.mt-archive.info/Bar-Hillel-1951.pdf>

BELIAEVA, L. *Machine translation methods, text structure and translator work*. International Journal of Translation, v.13, p. 119-146, 2001. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/00/IJT-2001-Beliaeva.pdf>

CARBONELL, J.; FARWELL, D.; FREDERKING, R.; HELMREICH, S.; HOVY, E.; KNIGHT, K.; NIRENBURG, S. *Pangloss. Technology partnerships for crossing the language barrier*. First Conference of the Association for Machine Translation in the Americas. Columbia, 1994. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/90/AMTA-1994-Carbonell.pdf>

CHO, K.; MERRIENBOER, B.; BAHDANAU, D. *On the properties of neural machine translation: encoder-decoder approaches*. Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. Doha, 2014. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/10/SSST-2014-Cho.pdf>

CHUNG, Y.; KANG, S.; MOON, K.; LEE, J. *Word sense disambiguation in a Korean-to-Japanese MT system using neural networks*. Coling-2002 workshop Machine translation in Asia. Taipei, 2002. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/Coling-2002-Chung.pdf>

COUGH, J. *Imagine you are a translation graduate*. TAUS, 21 de Janeiro de 2011. Acesso em 17 de Dezembro de 2020, disponível em <https://www.taus.net/think-tank/articles/translate-articles/imagine-you-are-a-translation-graduate?highlight=WyJtYWNoaW5lliwidHJhbnNsYXRpb24iLCIndHJhbnNsYXRpb24iLCJ0cmFuc2xhdGlvbiciLCIndHJhbnNsYXRpb24nliwidHJhbnNsYXRpb24ncyIsIm1hY2hpbmUgdHJhbnNsYXRpb24iXQ>

COUGHLIN, J. *Artificial intelligence and machine translation: present developments and future prospects*, 1988. Acesso em 30 de Novembro de 2020, disponível em <http://www.mt-archive.info/70/Babel-1988-Coughlin.pdf>

DOSTERT, L. *Systems of logic in machine translation*. Eighth Annual Round Table Meeting on Linguistics and Language Studies. Washington D.C., 1957. Acesso em 28 de Novembro de 2020, disponível em <http://www.mt-archive.info/50/GURT-1957-discussion-1.pdf>

DU, J.; WAY, A.; QIU, Z.; WASALA, A.; SCHALER, R. *Domain adaptation for social localisation-based SMT: a case study using the Trommons platform*. MT Summit XV: Fourth Workshop on Post-editing Technology and Practice. Miami, 2015. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/15/MTS-2015-W1-Du.pdf>

JIANG, J.; WAY, A.; HAQUE, R. *Translating user-generated content in the social networking space*. Tenth Biennial Conference of the Association for Machine Translation in the Americas. San Diego, 2012. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/10/AMTA-2012-Jiang-2.pdf>

LEHMANN, H. *MT for home and business users*. MT Summit V. Luxembourg, 1995. Acesso em 29 de Novembro de 2020, disponível em <http://www.mt-archive.info/90/MTS-1995-Lehmann.pdf>

LOMMEL, A. *MT is changing the industry, just not in the way mainstream media thinks it will*. CSA Research, 22 de Junho de 2016. Acesso em 17 de Dezembro de 2020, disponível em <https://csa-research.com/Insights/ArticleID/87/MT-is-Changing-the-Industry-Just-Not-in-the-Way-Mainstream-Media-Thinks-It-Will>

MASTERMAN, M.; KAY, M. *Mechanical pidgin translation: an estimate of the research value of "word-for-word" translation into a pidgin language, rather than into the full normal form of an output language*. Cambridge: Cambridge Language Research Unit, 1960. Acesso em 27 de Novembro de 2020, disponível em <http://www.mt-archive.info/50/CLRU-1960-Masterman.pdf>

NARUEDOMKUL, K.; CERCONE, N. *Steps toward accurate machine translation*. 7th International Conference on Theoretical and Methodological Issues in Machine Translation. Santa Fe, 1997. Acesso em 2 de Dezembro de 2020, disponível em <http://www.mt-archive.info/90/TMI-1997-Naruedomkul.pdf>

NELSON, B. *Machine Translation: Committe Skeptical over Research Support*. Science. v. 155, n. 1, 1967. Acesso em 25 de Novembro de 2020, disponível em <http://www.mt-archive.info/50/Science-1967-Nelson.pdf>

NISHIGAKI, T. *What can MT do for multilingualism on the Net?* Machine Translation Summit VII. Singapore, 1999. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/90/MTS-1999-Nishigaki.pdf>

OKITA, T. *Neural probabilistic language model for system combination*. Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT. Mumbai, 2012. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/10/Coling-2012-Okita-3.pdf>

OSWALD, V. *Word-by-word translation*. Conference on Mechanical Translation, Massachusetts Institute of Technology. Massachusetts Institute of Technology, 1952. Acesso em 23 de Novembro de 2020, disponível em <http://www.mt-archive.info/MIT-1952-Oswald-2.pdf>

_____. *An idioglossary for mechanical translation*. Modern Language Forum. v. 38, n. 3-4, 1953. Acesso em 23 de Novembro de 2020, disponível em <http://www.mt-archive.info/Oswald-1953.pdf>

STEIN, D. *Machine translation: past, present and future*. Translation: Computation, Corpora, Cognition, v. 3, n. 1, p. 5-13, 2013. Acesso em 2 de Dezembro de 2020, disponível em <http://www.mt-archive.info/10/TC3-2013-Stein.pdf>

TITUS, J. P. *The Nebulous Future of Machine Translation*. Communications of the ACM, v. 10, 1967. Acesso em 25 de Novembro de 2020, disponível em <http://www.mt-archive.info/CACM-1967-Titus.pdf>

VOLK, M.; HARDER, S. *Evaluating MT with translations or translators: what is the difference?* MT Summit XI. Copenhagen, 2007. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/05/MTS-2007-Volk.pdf>

WEAVER, W. *Correspondência*. New York: The Rockefeller Foundation, 1947. Acesso em 23 de Novembro de 2020, disponível em <http://www.mt-archive.info/Weaver-1947-original.pdf>

_____. *Correspondência*. New York: The Rockefeller Foundation, 1948. Acesso em 23 de Novembro de 2020, disponível em <http://www.mt-archive.info/50/Weaver-1948-1.pdf>

_____. *Correspondência*. New York: The Rockefeller Foundation, 1949. Acesso em 23 de Novembro de 2020, disponível em <http://www.mt-archive.info/Weaver-1949.pdf>

WILKS, Y. *Developments in MT research in the US*. Machine Translation Today: Translating and the Computer. v. 15. London. 1993. Acesso em 17 de Dezembro de 2020, disponível em <http://www.mt-archive.info/90/Aslib-1993-Wilks.pdf>

WITKAN, T. *A new road to automatic translation*. 90th Universal Kongreso. Utrecht, 2005. Acesso em 30 de Novembro de 2020, disponível em <http://www.mt-archive.info/05/Witkam-2005.pdf>

YNGVE, V. H. *Mechanical Translation*. Quarterly Progress Report of the Research Laboratory of Electronics. Massachusetts Institute of Technology, 1953. Acesso em 23 de Novembro de 2020, disponível em <http://www.mt-archive.info/50/MIT-RLE-1953-Yngve.pdf>