

Um Estudo Comparativo de Três Técnicas de Tradução Automática para Processamento de Textos

Maria Alice O. C. Leal¹, Ana Lúcia L. M. Maia¹, Thiago D’Martin Maia¹

¹Departamento de Exatas – Universidade Estadual de Feira de Santana (UEFS)
44036-900 – Feira de Santana, BA – Brasil

{maocleal,anamarreiros}@gmail.com, tdmaia@ecomp.uefs.br

Resumo. *A Tradução Automática (TA) despontou como um agente relevante para possibilitar a comunicação humana entre diferentes culturas e línguas. No entanto, as ferramentas de tradução atuais, como o Google Translate, ainda não abarcam a complexidade de tradução em contextos restritos, com um mínimo aceitável de erros. Dessa maneira, o presente artigo revisa três técnicas populares de TA para processamento de textos, apresentando um estudo comparativo das mesmas e exemplos de tradutores que utilizam essas técnicas. Portanto, este trabalho pode ser utilizado como um material objetivo para passos iniciais no desenvolvimento de uma ferramenta de Tradução Automática.*

Abstract. *Machine Translation (MT) has emerged as one of the relevant agents that can make possible human communication among different cultures and languages. To date, known Machine Translation tools, as Google Translate, still have insufficient performance in distinguishing specific disambiguation contexts. For that reason, this paper reviews three popular MT techniques for written text, presenting a comparative study and also describing translator examples using these techniques. Therefore, this paper can be used as a concise material for initial steps in Machine Translation development.*

1. Introdução

O crescente avanço nas tecnologias de comunicação tem permitido interligar diferentes pessoas nos mais diversos locais do mundo. Para suprir a demanda por comunicação rápida e compartilhamento de informações mais eficientes entre falantes de línguas distintas, faz-se necessário o uso de ferramentas de Tradução Automática (TA). Há um grande interesse por estas ferramentas, que envolvem a tradução de uma língua natural para outra através de um computador [Specia e Rino 2002], permitindo a disseminação de informações multilíngues.

A pesquisa nesta área instiga inúmeros desafios oriundos do objetivo de se conseguir a completa automatização da tradução. Um dos desafios é realizar uma tradução dentro de um contexto específico. Harriehausen-Mühlbauer e Heuss (2012) demonstraram um cenário de teste com os tradutores: o Bing Translator¹, o Yahoo Babel Fish², o Google Translate³, o Moses⁴ e o LINGUATEC Personal Translator PT⁵. Essas ferramentas

¹Bing Translator. Link: <http://www.bing.com/translator>

²O Yahoo Babel Fish foi unido ao tradutor Bing Translator em maio de 2012. Link: <http://blogs.msdn.com/b/translation/archive/2012/05/30/welcoming-yahoo-babel-fish-users.aspx>

³Google Translate. Link: <http://translate.google.com/>

⁴Versão de demonstração do Moses. Link: <http://demo.statmt.org/>

⁵Versão de demonstração do LINGUATEC. Link: <http://www.linguatec.net/onlineservices/pt>

foram aplicadas na tradução do inglês para o alemão da frase “*Pages by Apple is better than Word by MS*”, no contexto de tecnologia da informação. Foi observado que todas as ferramentas citadas não obtiveram êxito, pois não levaram em conta o contexto.

Dado que existem lacunas em ferramentas atuais de Tradução Automática, este trabalho apresenta uma revisão bibliográfica de três técnicas mais utilizadas na tradução, com o intuito de contribuir com as pesquisas na área. Mais especificamente, destaca as características positivas e negativas dessas técnicas, a partir de critérios de comparação das mesmas, e descreve exemplos de tradutores que utilizam essas técnicas.

2. Metodologia

A metodologia aplicada ao presente trabalho consistiu de duas fases: revisão bibliográfica e análise comparativa das técnicas de TA.

A revisão bibliográfica foi iniciada com o levantamento de material teórico de referência geral sobre o tema do trabalho. Após a revisão inicial sobre TA e suas técnicas mais populares, com um enfoque sobre os conceitos envolvidos em cada uma técnicas, seus processos de tradução e princípios envolvidos, foram escolhidas três técnicas para estudo mais detalhado: Baseada em Regras (RBMT - *Rule Based Machine Translation*), Baseada em Conhecimento (KBMT - *Knowledge Based Machine Translation*) e Baseada em Estatística (SBMT - *Statistical Based Machine Translation*).

Para a análise comparativa, inicialmente foram escolhidos e descritos critérios de comparação para as técnicas. Os critérios foram definidos com base nas características mais documentadas nos artigos escolhidos para estudo, principalmente provenientes de artigos do repositório *Machine Translation Archive*⁶. Os critérios utilizados na comparação foram: disponibilidade de literatura, amadurecimento da tecnologia, complexidade de implementação, custo de manutenção e desempenho. Por fim, foram descritas três aplicações de tradução que utilizam *frameworks* e ferramentas baseados nas técnicas RBMT, KBMT e SBMT.

3. Técnicas de Tradução Automática

Silva et al. (2007) identificam dois tipos de técnicas de TA: a técnica fundamental (que utiliza conhecimento linguístico) e a técnica empírica (que utiliza dados empíricos). A seguir, são descritas duas técnicas fundamentais, a Baseada em Regras (RBMT) e a Baseada em Conhecimento (KBMT), e uma técnica empírica, a Baseada em Estatística (SBMT).

3.1. TA Baseada em Regras - RBMT

A técnica Baseada em Regras (RBMT) realiza a transferência do texto em língua fonte (LF) para a língua alvo (LA), por meio de regras que representam o conhecimento em diferentes níveis linguísticos [Harshawardhan 2011]. A abordagem por transferência, a mais popular, inicialmente cria uma representação da LF com todas as suas informações linguísticas. Em seguida, aplicam-se regras de conversão do modelo intermediário da LF em um outro modelo intermediário com as características linguísticas da LA. Por fim, o texto na LA é produzido, com a equivalência entre a nova representação intermediária e a língua alvo (Figura 1).

⁶Repositório mantido pelo pesquisador John Hutchins. Link: <http://www.mt-archive.info/>

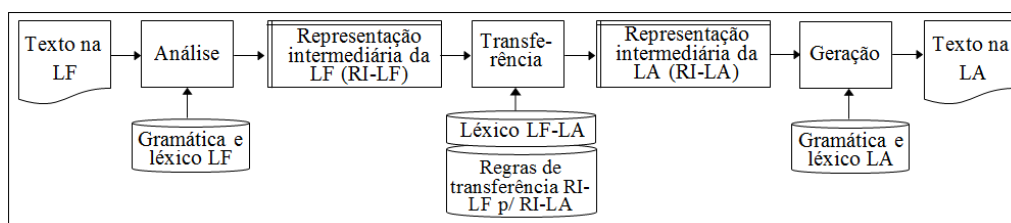


Figura 1. Arquitetura de um sistema Baseado em Regras (RBMT) com abordagem por transferência. Fonte: Adaptada de Silva et al. (2007).

Esta técnica tem como vantagem um bom desempenho e qualidade da tradução, capacidade de ser aplicada em contextos de propósito geral e em casos de línguas sem recursos de processamento linguístico. Como limitação principal associada a esta técnica, tem-se o fato de ser necessário criar e atualizar manualmente as regras de transferência do sistema. Os recursos envolvidos na criação de um tradutor automático com RBMT são as próprias regras de transferência lexicais, sintáticas e semânticas da língua.

3.2. TA Baseada em Conhecimento - KBMT

A técnica Baseada em Conhecimento (KBMT) associa conceitos mais profundos a um léxico por meio de modelos de domínio ou ontologias, como informações morfológicas, sintáticas e semânticas [Silva et al. 2007]. Através da consulta de termos na base de dados de semântica, organizada de forma sistemática para cada domínio, os sistemas KBMT realizam a tradução de uma língua fonte (LF) para uma língua alvo (LA) [Vertan 2005], passando por transferências intermediárias (Figura 2).

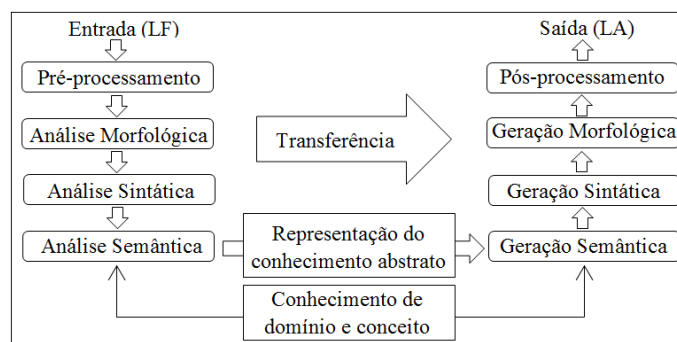


Figura 2. Arquitetura de um sistema KBMT. Fonte: Adaptada de Vertan (2005).

As vantagens desta técnica são uma tradução com mais qualidade, ao se encarar a ontologia como uma representação intermediária autônoma, e sua modularidade, o que permite substituir a base de dados sem muitas alterações no sistema. Como limitações, a KBMT apresenta o alto custo envolvido na criação das bases de conhecimento, a restrição destas bases a um domínio específico e a dificuldade de se delimitar a abrangência e o aprofundamento do domínio. Os recursos necessários à utilização desta técnica são uma ontologia no contexto desejado e regras de mapeamento lexical e gramatical.

3.3. TA Baseada em Estatística - SBMT

A técnica Baseada em Estatística (SBMT) é caracterizada pela obtenção automática de dados estatísticos da *corpora* paralela⁷ e do *corpus* monolíngue⁸ para a criação respectiva do modelo estatístico de tradução e do modelo linguístico [Silva et al. 2007]. Os modelos gerados serão treinados para realizar a tradução da LF para LA (Figura 3).

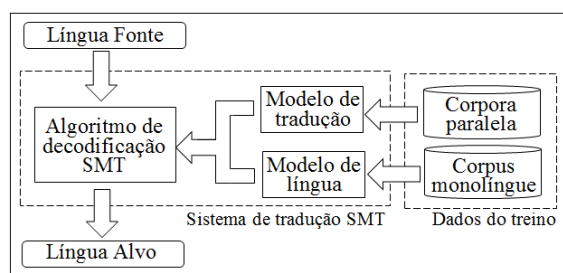


Figura 3. Arquitetura de um sistema SBMT. Fonte: Adaptada de NTT (2007).

As vantagens da SBMT são a existência de uma literatura vasta, o seu baixo custo (não é necessário contratar especialistas) e o rápido desenvolvimento, o que a faz se destacar dentre as técnicas. Seus pontos negativos são a restrição do tradutor ao domínio da *corpora* utilizada e a dificuldade em lidar com características específicas de um idioma. O recurso essencial para sua utilização é uma *corpora* paralela extensa para que seja possível treinar o tradutor em um grande conjunto de exemplos de sentenças nos idiomas escolhidos.

4. Análise Comparativa das Técnicas

Esta seção sistematiza uma comparação entre as técnicas estudadas, a partir dos pontos destacados, como vantagens, desvantagens, recursos necessários e área em que melhor se aplicam. Em seguida, estão descritos três tradutores que utilizam as técnicas selecionadas.

4.1. Sistematização da Análise

O Quadro 1 apresenta um resumo comparativo das técnicas revisadas, tendo sido criado a partir dos seguintes critérios:

- Disponibilidade de literatura, definida com base na distribuição de artigos por técnicas e por período, construída a partir das informações contidas no repositório *online Machine Translation Archive*. Foram definidos três graus de disponibilidade: baixa, para até 100 artigos; média, de 101 a 500 artigos; e alta, acima de 500 artigos;
- Amadurecimento da tecnologia, analisado com base no período de maior concentração da quantidade de artigos sobre cada técnica. Foram definidos os graus de amadurecimento: atual, com concentração de artigos entre 2001 e 2013; médio, com concentração entre 1991 e 2000; e antigo, com concentração anterior a 1990;

⁷Uma *corpora* paralela é um par de textos em duas línguas distintas equivalentes um ao outro.

⁸O *corpus* monolíngue é um conjunto de textos de uma única língua.

- c) Complexidade de implementação, analisado com base nas estruturas mínimas necessárias à aplicação de cada uma das técnicas. Foram definidos os graus de complexidade: alta, para técnicas com criação manual das regras de mapeamento lexicais, sintáticas e semânticas; média, para a necessidade de criação de estruturas híbridas de mapeamento; baixa, para a necessidade de apenas implementar tradutores com bases de dados e *frameworks* já disponíveis;
- d) Custo de manutenção, destaca as técnicas que oferecem uma maior dificuldade de atualização da ferramenta. Foram definidos os seguintes custos: alto, para as técnicas em que a atualização é de forma totalmente manual; médio, para as que dispõem de partes modularizadas e partes de atualização manual; baixa, para as que dispõem de um bom grau de modularidade ou flexibilidade;
- e) Desempenho, uma relação entre tempo de processamento da técnica e satisfação de resultados. Os graus de desempenho definidos foram: baixo, para técnicas que possuem alguma interação com o usuário durante a tradução; médio, para técnicas que realizam buscas em largas bases de dados; e alto, para técnicas com rápido tempo de processamento, normalmente personalizadas com regras de mapeamento linguísticas específicas.

Quadro 1. Resumo comparativo das técnicas de Tradução Automática estudadas.
 Fonte: Próprio Autor.

Crítérios de Comparação	RBMT	KBMT	SBMT
Disponibilidade de literatura	média	baixa	alta
Amadurecimento da Tecnologia	atual	médio	atual
Complexidade de implementação	alta	média	baixa
Custo de manutenção	alto	baixo	baixo
Desempenho	alto	médio	médio

Analisando o Quadro 1, com base nos critérios de disponibilidade de literatura, desempenho e amadurecimento da tecnologia, a técnica que se destaca é a técnica Baseada em Regras (RBMT). No entanto, se forem observados os critérios de disponibilidade de literatura em conjunto com o amadurecimento da tecnologia, a complexidade de implementação e o custo de manutenção, a técnica que mais se destaca é a Baseada em Estatística (SBMT), seguida pela Baseada em Conhecimento (KBMT).

4.2. Exemplos de Tradutores

Os seguintes tradutores foram selecionados como exemplos para ilustrar as técnicas estudadas: o Apertium (RBMT), o Jena (KBMT) e o Moses (SBMT).

4.2.1. Apertium - Tradutor Baseado em Regras

Nesta seção, é explicado um projeto de um tradutor Baseado em Regras por transferência com a ferramenta Apertium⁹. As etapas a serem seguidas são: a definição do par de idiomas a ser utilizado, levantamento de recursos linguísticos dentro do domínio desejado (dicionários monolíngues e bilíngues), configuração e personalização do Apertium, execução de testes e avaliação dos resultados obtidos com o padrão BLEU.

⁹Ferramenta de tradução Apertium. Link: <http://sourceforge.net/projects/apertium/>

Para se aplicar o Apertium a um contexto de domínio restrito, é necessário utilizar dicionários que abranjam o domínio desejado. Neste caso, pode-se realizar uma modificação dos dicionários monolíngues e bilíngue disponíveis na biblioteca do Apertium, adequando os vocábulos para o domínio a ser utilizado [François, Ribiczey e Ramírez-Sánchez 2010].

Na Figura 4, pode-se observar a arquitetura da ferramenta, com seus principais módulos. A ideia central do Apertium pode ser compreendida como uma sequência de traduções simples em diversos formatos por meio de módulos independentes [François, Ribiczey e Ramírez-Sánchez 2010]. Os módulos que compõem esta ferramenta são: os módulos de formatação, os módulos de processamento léxico, o módulo de desambiguação léxica e o módulo de transferência estrutural.

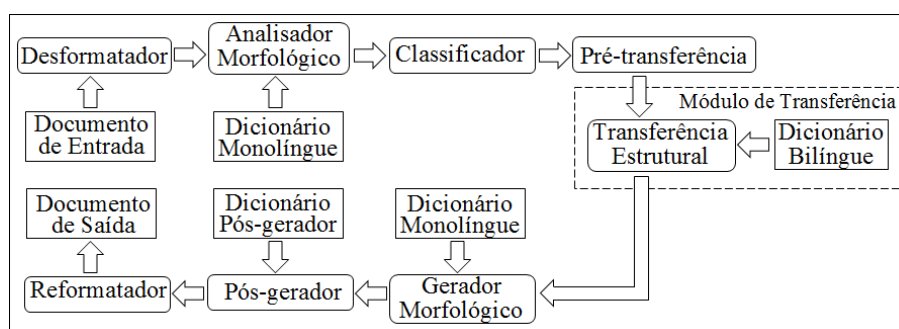


Figura 4. Tradutor Baseado em Regras com a ferramenta Apertium. Fonte: Adaptada de François, Ribiczey e Ramírez-Sánchez (2010).

Os módulos desformatador e reformatador realizam a separação entre o texto e sua formatação. O primeiro retira e armazena a formatação do texto em língua fonte (LF) que será processado nas próximas etapas, em contrapartida o segundo coloca de volta a formatação no texto em língua alvo (LA), para exibi-lo ao usuário [François, Ribiczey e Ramírez-Sánchez 2010].

Nos módulos do processamento léxico, o analisador morfológico identifica a forma lexical utilizada em cada palavra do texto de entrada [François, Ribiczey e Ramírez-Sánchez 2010]. Enquanto isso, o módulo de transferência converte palavras da LF para a LA. O módulo do gerador morfológico produz a forma lexical correta do trecho de texto traduzido pelo módulo de transferência. Por fim, o módulo pós-gerador faz correções ortográficas.

Segundo François, Ribiczey e Ramírez-Sánchez (2010), o módulo de desambiguação atua no módulo de transferência, quando o analisador léxico fornece mais de uma forma lexical possível para uma palavra. Cabe ao módulo de desambiguação determinar qual é a forma lexical mais provável para que o gerador morfológico finalize a geração do léxico na língua alvo.

O módulo de transferência estrutural é responsável por realizar algumas adequações estruturais entre a LF e a LA. São alguns exemplos de ações deste módulo: mudanças de tempos verbais, de preposições, de concordâncias, dentre outros [François, Ribiczey e Ramírez-Sánchez 2010].

Alguns recursos podem ser incorporados ao Apertium para melhorar a qualidade

da tradução. Por exemplo, textos técnicos podem ser usados como uma memória de tradução para influenciar nas regras do módulo de transferência e imprimir as características mais apropriadas ao estilo do texto a ser traduzido. Também podem ser feitas listas de significados para um conjunto de palavras, corrigindo problemas em que juntas essas palavras teriam um significado diferente do que se estivessem isoladas [François, Ribiczey e Ramírez-Sánchez 2010].

4.2.2. Jena - Tradutor Baseado em Conhecimento

Nesta seção, é detalhado um exemplo de como projetar um tradutor Baseado em Conhecimento da *web* semântica com a ferramenta Jena¹⁰, tendo como base o trabalho desenvolvido por Harriehausen-Mühlbauer e Heuss (2012). A ferramenta Jena é um *framework* utilizado para realizar o processamento do núcleo semântico do tradutor, através de manipulações de arquivos no padrão RDF (*Resource Description File* - padrão de descrição de conhecimento na *web* semântica) e suporte à linguagem de solicitações SPARQL.

Os recursos necessários ao projeto são uma base de domínio no padrão RDF, um dicionário bilíngue simples, um núcleo de processamento semântico e um módulo de pré-processamento e pós-processamento (Figura 5).

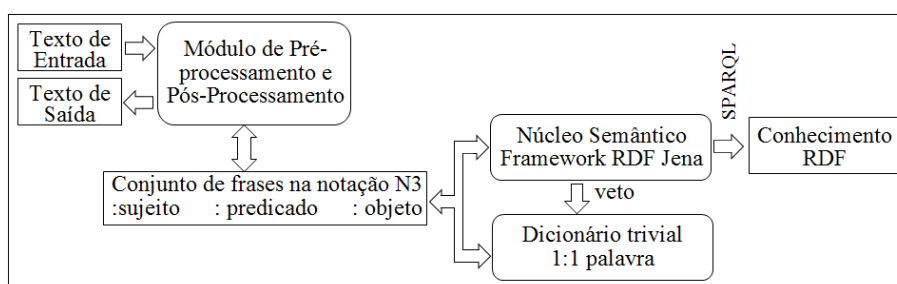


Figura 5. Tradutor Baseado em Conhecimento com a ferramenta Jena. Fonte: Adaptada de Harriehausen-Mühlbauer e Heuss (2012).

O módulo de pré-processamento realiza a conversão do texto em LF para um conjunto de frases composta por três partes, de acordo com a notação N3¹¹ do RDF: sujeito, predicado e objeto. Esta notação possui a vantagem de ser mais compreensível que outras notações formais. Em correspondência, o módulo de pós-processamento realiza a tarefa inversa, convertendo do formato N3 para o texto em LA, pela união das três partes da frase, sujeito, predicado e objeto [Harriehausen-Mühlbauer e Heuss 2012].

O dicionário é responsável pelas traduções palavra a palavra, sem verificações de contexto, nem restrições lexicais, sintáticas ou semânticas. Uma vez feita a tradução simples pelo módulo do dicionário, o núcleo semântico realizará a tradução contextualizada com a base RDF. Segundo Harriehausen-Mühlbauer e Heuss (2012), o núcleo semântico deve procurar uma tradução mais bem qualificada que a do dicionário, através de consultas com a linguagem SPARQL à base de conhecimento semântico RDF. Para obter um resultado satisfatório em termos de profundidade do conhecimento especificado, a base RDF deve ser criada de forma eficiente.

¹⁰Ferramenta de tradução Jena. Link: <http://jena.apache.org/>

¹¹Padrão de notação N3. Link: <http://www.w3.org/DesignIssues/Notation3.html>

A maior vantagem de utilizar essa abordagem está em solucionar problemas de caráter sintático e semântico que muitas vezes sequer são verificados em outras abordagens. Então, um benefício incremental seria obtido da combinação desta ferramenta Baseada em Conhecimento semântico com a abordagem Baseada em Estatística (discutida na próxima seção). Assim, a abordagem estatística poderia resolver o problema da determinação da melhor tradução, enquanto a Baseada em Conhecimento filtraria as traduções relevantes dentro do contexto especificado.

4.2.3. Moses - Tradutor Baseado em Estatística

Nesta seção, é descrito como projetar um sistema de tradução que utiliza a técnica Baseada em Estatística com a ferramenta Moses¹², tendo como referência a tarefa de tradução do *Workshop on Statistical Machine Translation - 2011 (WMT11)*¹³.

O primeiro passo é a obtenção de uma *corpora* paralela extensa no domínio em que se deseja realizar traduções. Após obtê-la, devem ser instaladas as seguintes ferramentas: Moses, responsável pela tradução; SRILM¹⁴, que cria e aplica o modelo estatístico; e GIZA++¹⁵ que treina o modelo de tradução estatístico da *corpora* paralela. A partir de então, seguem-se as próximas etapas: preparação dos dados, construção do modelo da linguagem, treinamento do modelo, otimização, execução do conjunto de testes e avaliação (Figura 6).

Na etapa de preparação dos dados, são utilizados *scripts*¹⁶ para realizar a tokenização da *corpora* (*tokenizer*), a separação do texto em frases (*tokens*), a conversão de todas as letras para o formato minúsculo (*lowercase*) e a exclusão de frases acima de um tamanho especificado (*clean-corpus*) [Koehn 2011].

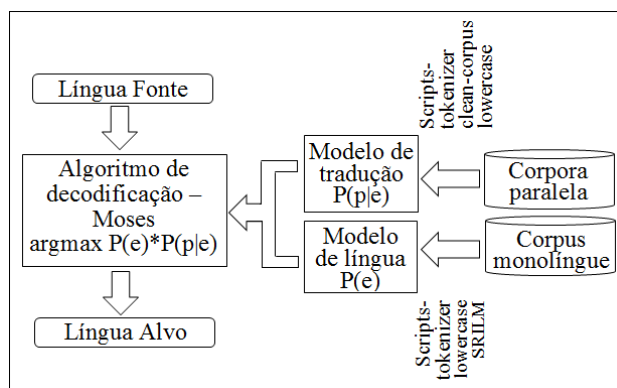


Figura 6. Tradutor estatístico com Moses, SRILM e GIZA++. Fonte: Adaptada de NTT (2007).

Para a construção do modelo de língua, são utilizados os mesmos *scripts* da etapa anterior para formatar os dados monolíngues da *corpora*. Então, a ferramenta SRILM cria o modelo que será utilizado no treinamento do tradutor estatístico [Koehn 2011].

¹²Ferramenta de tradução Moses. Link: <http://www.statmt.org/moses/>

¹³Workshop on Statistical Machine Translation 2011. Link: <http://www.statmt.org/wmt11/>

¹⁴Biblioteca SRILM. Link: <http://www.speech.sri.com/projects/srilm/>

¹⁵Biblioteca GIZA++. Link: <https://code.google.com/p/giza-pp/>

¹⁶Scripts para manipular os textos. Link: <http://www.statmt.org/wmt09/scripts.tgz>

Este modelo atribui probabilidades às sequências de n palavras (tamanho variável) que ocorrerem na base de dados monolíngue. O treinamento do modelo de língua criado na etapa anterior é feito através do Moses, utilizando os dados pré-processados da *corpora* paralela na primeira etapa [Koehn 2011].

Na fase de otimização, são tokenizados e convertidos para minúsculo os conjuntos de dados específicos de entrada e de referência de tradução do par de idiomas que se está utilizando. Depois é executado o *script* de refinamento com o objetivo de chegar a uma taxa de erro mínima [Koehn 2011].

Na etapa de testes, primeiramente são tokenizados e convertidos para minúsculo os textos de testes que se deseja traduzir, bem como as suas traduções de referência. Por fim, é executada a decodificação da língua fonte para a língua alvo com o Moses [Koehn 2011].

A última etapa avalia os resultados obtidos com os testes, para verificar o nível de acertos do tradutor. Nesta, o arquivo resultante dos testes é reformatado para então ser comparado com o conjunto de testes de referência [Koehn 2011]. A pontuação de acertos é dada pelo sistema de avaliação NIST BLEU [Papineni et al. 2002], que consiste em pontuar melhor as traduções automáticas mais próximas da tradução humana.

4.3. Discussões

A partir do estudo das técnicas de Tradução Automática levantadas, foi possível estabelecer um perfil de utilização de cada uma delas, determinar suas características marcantes, recursos necessários para sua aplicação, suas limitações, desempenho e custo.

Vários projetos vêm obtendo bons resultados na tradução de textos técnicos utilizando a técnica de tradução estatística, como, por exemplo, o tradutor Pluto [Tinsley, Way e Sheridan 2010]. Entretanto, em algumas situações não é viável sua aplicação, como nos casos das línguas que não dispõem de uma *corpora* paralela extensa de textos técnicos [Brandt e Tyers 2011]. Dois exemplos de línguas sem *corpora* paralela técnica, pública e extensa são o português-Br e o islandês, ambos em relação ao inglês [Brandt e Tyers 2011; Frankenberg-garcia e Santos 2002]. Nesses casos, normalmente é empregada a técnica de tradução Baseada em Regras, juntamente com alguma outra abordagem, em que a *corpora* paralela não é requerida.

Com relação ao exemplo de tradutor Baseado em Regras apresentado, é possível notar que os materiais presentes na literatura não são tão específicos quanto os referentes à técnica estatística. No caso da técnica RBMT, cada projeto desenvolvido constitui um projeto feito sob medida para os pares de línguas e tipo de linguagem do contexto utilizado. Dentre as ferramentas existentes, pôde-se identificar o Apertium [François, Ribiczey e Ramírez-Sánchez 2010], que tem como objetivo automatizar o processo da tradução Baseada em Regras, ficando a cargo do desenvolvedor a modificação e configuração dos dicionários bilíngues, monolíngues e módulos de desambiguação utilizados. Para experimentações simples, é possível utilizar dicionários em versões de desenvolvimento disponíveis na biblioteca do Apertium. No entanto, particularmente em testes de domínios específicos, é necessária a modificação da base de dicionários existente.

No caso do tradutor Baseado em Conhecimento (KBMT) apresentado, foi descrita a utilização da ferramenta Jena [Harriehausen-Mühlbauer e Heuss 2012], que re-

aliza o fluxo de processamento central de um tradutor. Para a criação de um tradutor técnico, seria necessário desenvolver ou buscar uma base de conhecimentos RDF da *web* semântica específica, para utilização como base no sistema. De acordo com Harriehausen-Mühlbauer e Heuss (2012), a personalização de uma base RDF e o processamento de uma base do mesmo tipo com relações complexas acarretam um custo de desenvolvimento e desempenho grandes, respectivamente. Uma abordagem que tem sido adotada é o uso desta técnica em conjunto com outra, por exemplo a estatística (SBMT), que pode suprir a carência da indecidibilidade da KBMT pela melhor tradução pesquisada na ontologia.

Apesar de viáveis, os exemplos de tradutores automáticos apresentados possuem vantagens e desvantagens em casos específicos. Tais técnicas exigem a execução de algumas tarefas custosas anteriores ao desenvolvimento do tradutor, como o desenvolvimento de uma *corpora* paralela, a adequação de dicionários ou o desenvolvimento e personalização de uma base de conhecimentos semânticos RDF, o que, por vezes, exige a participação de especialistas linguísticos.

5. Considerações Finais

Foi apresentado um estudo das três técnicas de Tradução Automática mais populares, bem como uma análise comparativa, apontando vantagens e deficiências de cada uma das técnicas. Além disso, com um enfoque prático, também foram apresentados exemplos de tradutores que utilizam estas técnicas. Dessa maneira, este trabalho contribui com uma introdução sucinta à Tradução Automática, a compreensão de suas principais técnicas e processos de tradução de exemplos práticos.

Alguns trabalhos futuros podem ser sugeridos como continuação do presente estudo: o desenvolvimento de uma *corpora* paralela para um domínio específico; o desenvolvimento de uma base RDF; a construção de uma ferramenta de busca e geração de bases de dados da *web* semântica; a implementação e testes de um tradutor automático utilizando algumas das técnicas e ferramentas apresentadas.

Referências

- Brandt, M. e Tyers, F. (2011). “Apertium-IceNLP: A rule-based Icelandic to English machine translation system.” Em: *Proceedings of the 15th Annual Conference of the European Association for Machine Translation*. Lovaina.
- François, M., Ribiczey, P. e Ramírez-Sánchez, G. (2010). “Using the Apertium Spanish-Brazilian Portuguese Machine Translation System for Localization.” Em: *Proceedings of the 14th Annual conference of the European Association for Machine Translation*. St. Raphael.
- Frankenberg-garcia, A. e Santos, D. (2002). “COMPARA, um corpus paralelo de português e inglês na Web.” Em: *Cadernos de Tradução* 1.9, pp. 61–79.
- Harriehausen-Mühlbauer, B. e Heuss, T. (2012). “Semantic Web based Machine Translation.” Em: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Avignon, pp. 1–9.
- Harshawardhan, R. (2011). “Rule Based Machine Translation System for English to Malayalam Language”. Diss. de mestrado. Coimbatore: Amrita School of Engineering.
- Koehn, P. (2011). *Baseline System: Moses*. <http://www.statmt.org/wmt11/baseline.html>. Junho, 2013.

- NTT (2007). *Statistical Machine Translation (SMT)*. <http://www.ntt.co.jp/RD/OFIS/active/2007pdf/hot/ct/06.html>. Junho, 2013.
- Papineni, K. et al. (2002). “Bleu: a Method for Automatic Evaluation of Machine Translation.” Em: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, pp. 311–318.
- Silva, B. C. D. et al. (2007). *Introdução ao Processamento das Línguas Naturais e Algumas Aplicações*. Rel. téc. NILC-TR-07-10. São Paulo: ICMC - USP.
- Specia, L. e Rino, L. H. M. (2002). *Introdução aos Métodos de Tradução e Paradigmas de Tradução Automática*. Rel. téc. NILC-TR-02-04. São Paulo: ICMC - USP.
- Tinsley, J., Way, A. e Sheridan, P. (2010). “PLuto: MT for Online Patent Translation”. Em: *Proceedings of AMTA 2010*. Denver.
- Vertan, C. (2005). *Knowledge Based Machine Translation*. <http://nats-www.informatik.uni-hamburg.de/pub/User/IntensiveCourseInMachineTranslation/kbmt.pdf>. Março, 2013.