

UNIVERSIDADE FEDERAL DE ALAGOAS
PRÓ-REITORIA DE GRADUAÇÃO
INSTITUTO DE COMPUTAÇÃO

Trabalho de Conclusão de Curso

Tradução por Software: Estado da Arte e Análise Comparativa no Contexto do Falibras

Aluno:
MARCELO SANTA FÉ TODARO

Orientador:
PATRICK HENRIQUE DA SILVA BRITO

Maceió, março de 2013

Marcelo Santa Fé Todaro

Tradução por Software: Estado da Arte e Análise Comparativa no Contexto do Falibras

Monografia apresentada como requisito parcial
para obtenção do grau de Bacharel em Sistemas
de Informação do Instituto de Computação da
Universidade Federal de Alagoas.

Orientador:
PATRICK HENRIQUE DA SILVA BRITO

Maceió, março de 2013

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação do Instituto de Computação da Universidade Federal de Alagoas, aprovado pela comissão examinadora que abaixo assina.

Patrick Henrique da Silva Brito – Orientador
Instituto de Computação
Universidade Federal de Alagoas

Marcus de Melo Braga – Avaliador
Instituto de Computação
Universidade Federal de Alagoas

Olival de Gusmão Freitas Jr. – Avaliador
Instituto de Computação
Universidade Federal de Alagoas

Agradecimentos

A meu bondoso e compassivo Pai Celeste, sem cuja clemente ajuda eu jamais teria alcançado esta conquista.

A meus pais, que ao longo da vida lutaram incansavelmente por minha educação.

À minha ex-mulher Adriana, que me persuadiu a tentar um vestibular para o qual eu me julgava despreparado e cuja aprovação ajudou a restagar a fé em mim mesmo.

Resumo

As primeiras tentativas de desenvolvimento de sistemas automatizados de tradução remontam ao Séc. XVII, muito embora a forma não humana de converter textos de um idioma em outro tenha se tornado realidade somente a partir da segunda metade do Séc. XX. Além de todos os idiomas falados no mundo entre os quais se pode fazer traduções, há de se notar a existência de um adotado por uma pequena parcela da população: a língua de sinais dos deficientes auditivos brasileiros, chamada libras. Segundo o censo de 2010 do IBGE, 9,8 milhões de brasileiros, ou 5,2% da população, possuem algum tipo de deficiência auditiva. Destes, 2,6 milhões são surdos e 7,2 milhões ouvem com variados graus de dificuldade. Pessoas com tal deficiência demandam por acesso a recursos de informação, particularmente os de apoio à aprendizagem. Para atender tal demanda, nos últimos anos tem crescido a tendência de se utilizar sistemas de tradução automática para promover acessibilidade a pessoas surdas por meio da tradução automática de textos de português para libras. Foi então que em 2002 surgiu o sistema Falibras. Este trabalho aborda a temática da tradução automática de textos, com ênfase no uso desse tipo de sistema no apoio à aprendizagem pelos deficientes auditivos. Apresenta-se nele o estado da arte em termos de tradução automática, contextualizando-a com as estratégias de tradução já adotadas pelo Falibras no decorrer de seu desenvolvimento e aperfeiçoamento.

Abstract

The first attempts to deploy automated translation go back to the seventeenth century, although the human form of converting text from one language to another became reality only in the second half of the twentieth century. Besides all the languages spoken throughout the world among which translations can be performed, one should note the existence of a language spoken by only a small portion of the population: the sign language used by deaf Brazilians, called Libras. According to the IBGE 2010 census, 9.8 million Brazilians, or 5.2% of the population, have some form of hearing impairment. Of these, 2.6 million are deaf and 7.2 million hear with varying degrees of difficulty. People with such disabilities demand access to information resources, particularly learning support. To meet such demand, it has increased in recent years the trend of using machine translation systems to promote accessibility for deaf people through automatic translation of texts from Portuguese into Libras. Then, in 2002 came the Falibras system. This work addresses the automatic translation of texts, with emphasis on the use of this type of system in supporting learning from deaf students. It presents the state of the art in machine translation, contextualizing it with the translation strategies already adopted by Falibras during its development and improvement.

Sumário

Agradecimentos.....	1
Resumo.....	2
Abstract.....	3
1. Introdução.....	1
1.1. Motivação e justificativa.....	1
2. Fundamentação teórica.....	5
2.1. Resumo das principais arquiteturas.....	5
2.1.1. A arquitetura direta.....	5
2.1.2. A arquitetura de transferência.....	7
2.1.3. A arquitetura interlíngua (IL).....	8
2.2. Processamento de Linguagem Natural (PLN).....	10
3. Tradução automática de texto: estado da arte.....	11
3.1. Paradigmas da tradução automática.....	11
3.1.1. Paradigmas baseados em linguística.....	11
3.1.1.1. A técnica das restrições.....	12
3.1.1.2. A técnica do léxico.....	13
3.1.1.3. A técnica baseada em regras.....	15
3.1.1.4. A técnica “agitar e assar”.....	16
3.1.2. Paradigmas não baseados em linguística.....	17
3.1.2.1. A técnica baseada em estatística.....	17
3.1.2.2. A técnica baseada em exemplo.....	18
3.1.2.3. A técnica baseada em rede neural.....	19
3.1.3. Paradigmas híbridos.....	20
4. Histórico da evolução do algoritmo de tradução do Falibras.....	21
4.1. O sistema Falibras.....	21
4.2. Tradução baseada em regras.....	24
4.3. Transferência sintática.....	25
4.4. Tradução baseada em memória de tradução.....	26
5. Proposta de solução híbrida.....	28
6. Conclusões e trabalhos futuros.....	32
Referências.....	33

1. Introdução

1.1. *Motivação e justificativa*

Um provável primeiro registro da existência de uma multiplicidade de idiomas no mundo data de aproximadamente dois milênios antes de Cristo e está contido no Velho Testamento bíblico. Em Gênesis 11:1-9 o patriarca Moisés descreve como a civilização do mundo da época, antes falante de um único idioma, passou a falar diversos deles, tendo por isso sido dispersa e fragmentada em grupos que falavam o mesmo idioma. O episódio tornou-se historicamente conhecido como confusão das línguas na Torre de Babel.

Desde os tempos antigos, o único meio de tornar possível a comunicação entre pessoas que não falam o mesmo idioma é por meio do aprendizado do idioma do outro. Conforme a população mundial cresceu e o número de idiomas também, aumentou na mesma medida a necessidade de aprender novos idiomas para ser possível haver intercâmbio entre os diferentes povos, nações e culturas.

Aprender um novo idioma implica necessariamente em fazer traduções. É possível que as civilizações antigas já sonhassem com meios de tornar o processo de tradução o mais automático possível à medida que migrações, intercâmbio, comércio e turismo se desenvolviam. Antes mesmo do advento da Revolução Industrial ocorrida na Inglaterra, pensadores do Séc. XVII como Descartes, Leibniz e outros filósofos e acadêmicos imaginavam como tornar tais sistemas realidade (HUTCHINS, 2013). Como na época não existia eletricidade e a ciência não conhecia nada além de dispositivos meramente mecânicos, o sonho estava muito adiante de seu tempo para que pudesse vingar.

Foi só três séculos mais tarde que os primeiros rudimentos de tais sistemas vieram à luz. Em 1933 o engenheiro francês de origem armênia Georges Artsrouni patenteou na França um dispositivo eletromecânico capaz de armazenar e imprimir informações (Figura 1). Tal capacidade

rendeu-lhe muitas possíveis aplicações, como a produção automática de tabelas de horários de trens, listas telefônicas, códigos telegráficos, extratos bancários e até decriptação/criptação de mensagens. No fim, atribuiu-se à invenção de Artsrouni uma nova utilidade: fazer traduções (HUTCHINS, 2013).

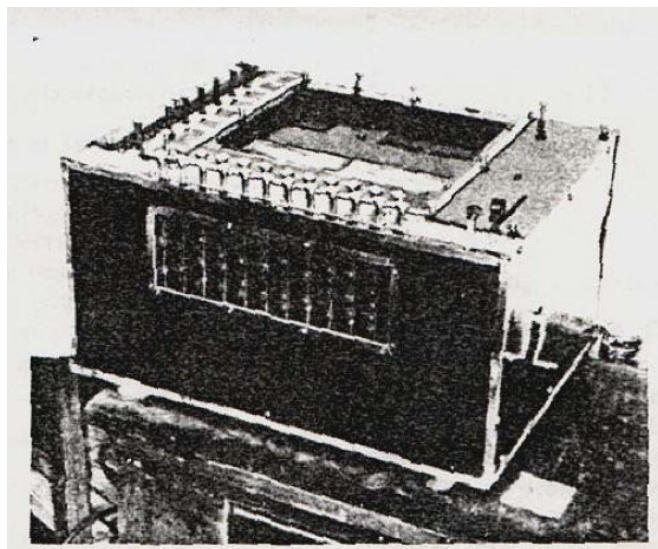


Figura 1: a máquina de Artsrouni (HUTCHINS, 2013)

Formas não humanas de converter textos de um idioma em outro só se tornaram alvo de pesquisas sérias a partir da segunda metade do Séc. XX. Várias foram as abordagens e técnicas computacionais desenvolvidas para essa finalidade, cada qual com seus prós e contras. A popularização da Internet no fim do Séc. XX contribuiu significativamente para o aumento na demanda por traduções à medida que colocou diante das pessoas oportunidades ilimitadas de estudo, trabalho e lazer. Era, pois, natural de se esperar que a tarefa de traduzir textos respondesse a essa demanda com sistemas automatizados cuja eficácia progride na velocidade do aumento da demanda, especialmente quando tal eficácia baseia-se não só em algoritmos cada vez mais evoluídos como também em métodos colaborativos.

Muito embora os modernos sistemas informatizados de tradução sejam capazes de produzir resultados cada vez melhores, ainda está longe o dia em que um programa de computador esteja apto a fazê-lo tal qual um competente tradutor humano, já que as sutilezas e nuances da linguagem humana estão além da atual capacidade computacional de análise. Quando muito, tais sistemas são capazes de produzir boas traduções de manuais técnicos, documentos científicos, prospectos comerciais, memorandos administrativos e relatórios médicos – ou seja, áreas em que o vocabulário

é limitado e não há ampla variação em torno de um mesmo tema.

Há de se notar também a existência de um idioma que geralmente passa despercebido pelo grande público por ser falado por apenas uma pequena parcela da população: a linguagem de sinais usada pelos deficientes auditivos brasileiros, chamada libras¹ (acrônimo de Língua Brasileira de Sinais). Segundo o censo de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE), 9,8 milhões de brasileiros, ou 5,2% da população, possuem algum tipo de deficiência auditiva. Destes, 2,6 milhões são surdos e 7,2 milhões ouvem com variados graus de dificuldade (BRASIL, 2013).

Pessoas com tal deficiência demandam por acesso a recursos de informação, particularmente os de apoio à aprendizagem. Para atender essa demanda, nos últimos anos tem crescido a tendência de se utilizar sistemas de tradução automática para promover acessibilidade a pessoas surdas por meio da tradução automática de textos de português para libras. Foi então que em 2002 surgiu o sistema Falibras (BRITO *et al.*, 2012).

Desenvolvido desde 2001, o Falibras provou ser uma eficiente ferramenta didática e pedagógica na educação de surdos, a ponto de despertar o interesse de uma empresa que desenvolveu um aplicativo para celular baseado no projeto, o Hand Talk. Em fevereiro de 2013 o aplicativo ganhou prêmio da ONU na categoria inclusão social no concurso WSA-Mobile, considerado o Oscar da tecnologia móvel, ocorrido em Abu Dhabi, nos Emirados Árabes.

Este trabalho pretende, pois, abordar a temática da tradução automática de textos, com ênfase no uso desse tipo de sistema no apoio à aprendizagem pelos deficientes auditivos. Apresenta-se nele o estado da arte em termos de tradução automática, contextualizando-a com as estratégias de tradução já adotadas pelo sistema Falibras no decorrer do seu desenvolvimento e aperfeiçoamento.

O restante desta monografia é apresentado como segue:

¹ Diversos autores grafam o nome libras com todas as letras maiúsculas – ou com pelo menos a inicial maiúscula – por representar uma sigla. Neste trabalho, contudo, será grafado com minúsculas por referir-se ao nome de um idioma, em concordância com as regras gramaticais da língua portuguesa que determinam que nomes de idiomas sejam escritos com inicial minúscula, assim como também dias da semana, meses do ano, etc.

- O capítulo 2 mostra a fundamentação teórica do trabalho, trazendo um resumo das principais arquiteturas de tradução por software, uma breve explicação sobre o Processamento de Linguagem Natural (PLN) e, por fim, o sistema Falibras;
- O capítulo 3 aborda o estado da arte em tradução automática de texto ao explorar os paradigmas da tradução automática, divididos em paradigmas linguísticos, não linguísticos e híbridos;
- O capítulo 4 mostra o histórico da evolução do algoritmo de tradução do Falibras, falando na tradução baseada em regras, na transferência sintática e na tradução baseada em memória de tradução;
- O capítulo 5 apresenta a proposta de solução híbrida adotada no Falibras;
- Por fim, o capítulo 6 expõe conclusões e trabalhos futuros.

A metodologia usada nesta pesquisa obedeceu a seguinte ordem:

1. Revisão bibliográfica com resumo dos principais trabalhos na área;
2. Estudo sobre processamento de língua natural e técnicas de tradução automática;
3. Estudo histórico das técnicas de tradução utilizadas no projeto Falibras;
4. Comparação das principais técnicas relatadas na literatura;
5. Proposta de abordagem híbrida para tradução automática de textos combinando as técnicas de memória de tradução e tradução baseada em regras;
6. Redação da monografia.

2. Fundamentação teórica

2.1. Resumo das principais arquiteturas

As arquiteturas da tradução automática de textos podem ser organizadas a grosso modo nas seguintes classes (DORR, 1998):

1. Direta
2. De transferência
3. Interlíngua (IL)

Essas três arquiteturas foram representadas num diagrama de “pirâmide” (Figura 2) que surgiu pela primeira vez de forma ligeiramente diferente na pesquisa de Vauquois (1968) e desde então tornou-se clássico. Dependendo da profundidade da análise provida pelo sistema, as três arquiteturas correspondem a diferentes níveis de transferência. A arquitetura direta está na base da pirâmide, consistindo na forma mais primitiva de transferência: substituição palavra-a-palavra. No topo está a arquitetura IL, que é a forma mais degenerada de transferência: o mapa de transferência essencialmente não existe. Em algum lugar desses dois extremos encaixa-se a maioria dos sistemas de tradução, variando de uma análise rasa (sintática) a uma análise profunda (semântica).

2.1.1. A arquitetura direta

Uma arquitetura direta produz um resultado que é um conjunto de palavras no idioma fonte diretamente substituídas pelas equivalentes no idioma alvo. De modo geral, a ordem das palavras nos idiomas fonte e alvo é a mesma, até nos casos em que as características do idioma alvo não permitem a mesma ordem, produzindo um resultado bem difícil de entender se o leitor não tiver bom conhecimento da estrutura do idioma alvo.

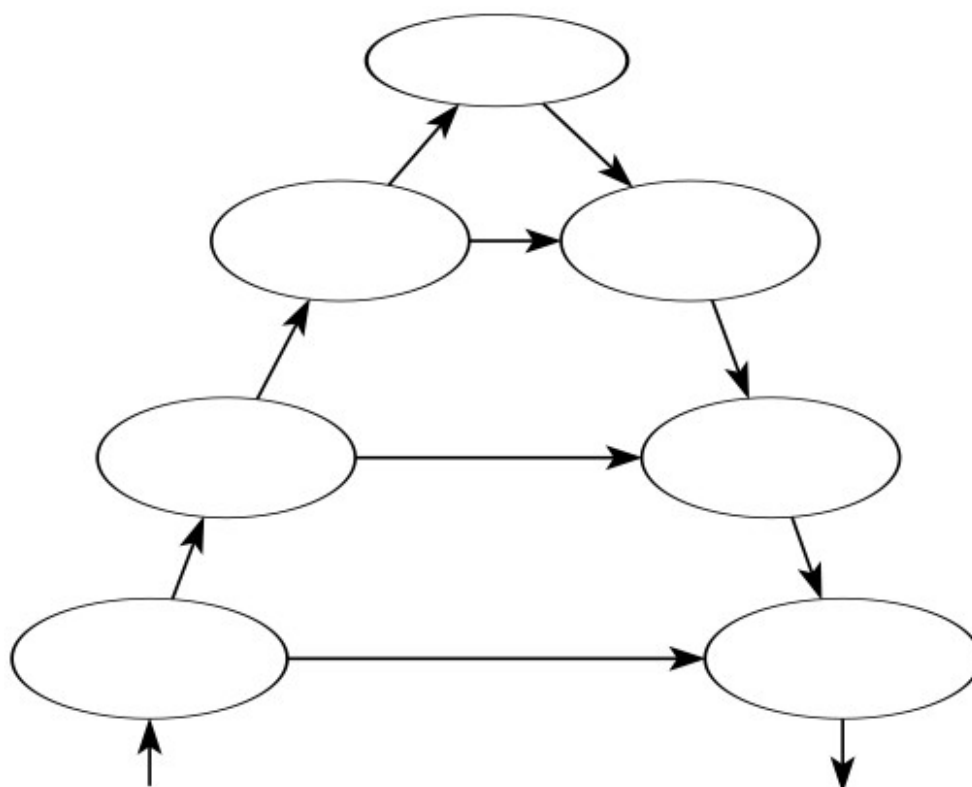


Figura 2: Tipos de sistemas de tradução automática (DORR, 1998)

Alguns sistemas de arquitetura direta são capazes de reconhecer formas sintáticas especiais no idioma fonte e, por isso, conseguem reordenar o conjunto de palavras no idioma alvo em formas aceitáveis, aumentando a legibilidade do texto traduzido. As regras que permitem a esses sistemas fazer essa reordenação são, em natureza, semelhantes às usadas nos sistemas de transferência, mas, como não efetuam análises sintáticas detalhadas, apenas formas simples podem ser reconhecidas. Assim, em casos contendo estruturas complexas como cláusulas e separações verbais – comuns em idiomas como o alemão –, a sintaxe da tradução fica como originalmente encontrada no idioma fonte. Em casos mais difíceis, como na frase em inglês *I like to eat*, é impossível construir regras de mapeamento direto, produzindo traduções muito literais, como *Ich esse gern* em alemão.

As ambiguidades léxicas representam problema ainda mais sério para os sistemas de arquitetura direta (assim como para algumas versões dos sistemas de transferência). Palavras como *ball* têm significados e traduções bastante distintos (homografia): pode ser um objeto esférico (*bola*) ou uma festa dançante (*baile*) – uma frase como *I like to dance at balls* pode acabar sendo traduzida

como *Gosto de dançar nas bolas*. Mais problemas surgem com palavras como *kill*, que tem sutis diferenças de significados relacionados em diferentes contextos (polissemia): pode significar matar uma pessoa (*kill a person*) ou encerrar um processo (*kill a process*). Nesses casos, os sistemas de arquitetura direta não sabem relacionar uma palavra à forma como é usada na sentença. Assim, o melhor que podem fazer é limitar o domínio textual e incluir apenas a tradução mais provável para aquele domínio. Em outras palavras, produzem traduções pobres.

2.1.2. A arquitetura de transferência

Originalmente, as arquiteturas de transferência tinham como intenção fornecer traduções sintaticamente corretas pela transformação das representações no idioma fonte em representações sintaticamente apropriadas no idioma alvo. Para fazer tal conversão era necessário usar regras de transferência adequadas a ambos os idiomas, embora algumas dessas regras requeriam ligeiras modificações quando um sistema de tradução automática fosse desenvolvido para um novo idioma linguisticamente relacionado a um já existente.

Os sistemas de transferência necessitam de “regras de ligação” servindo como mapa entre o assim chamado “texto de superfície” (nos idiomas fonte e alvo) e alguma representação interna. Assim, as regras de transferência são necessárias para criar um mapa entre essas duas representações. Ao contrário da arquitetura direta, a de transferência acomoda mais mapeamentos complexos, como o do exemplo *I like to eat*. No entanto, como alguns pesquisadores notaram (BENNETT *et al.*, 1986; DORR, 1993a), o problema com essa abordagem é que para cada par de idiomas fonte/alvo é preciso criar um grande número de regras de transferência. Por exemplo, um sistema de tradução criado para n idiomas requer n^2 conjuntos de regras.

Apesar dessa desvantagem, a arquitetura de transferência têm a vantagem de conseguir resolver com esforço mínimo as ambiguidades levadas de um idioma para outro. Por exemplo, na frase *John hit the dog with a stick* a ambiguidade não pode ser resolvida pela análise sintática por não haver como determinar sintaticamente se John usou uma vara para bater no cão ou se era o cão que portava a vara. Dependendo dos idiomas envolvidos na tradução, pode não ser necessário resolver tais ambiguidades: elas podem ser transferidas do idioma fonte para o alvo e ainda serem

compreendidas pelo leitor. Alguns sistemas de transferência tiram vantagem desse fenômeno e evitam realizar análises sintáticas dessas estruturas, resultando em melhor desempenho.

Outra vantagem das arquiteturas de transferência é sua capacidade de resolver certas ambiguidades léxicas, já que a determinação da categoria léxica de uma palavra no texto fonte pode ser determinada pela análise sintática. Por exemplo, dependendo do contexto, pode-se determinar se a palavra inglesa *book* deve ser traduzida como o substantivo *livro* ou como o verbo *reservar*.

Apesar disso, há muitas ambiguidades sintáticas e léxicas incapazes de serem resolvidas pelos sistemas de transferência, o que pode produzir traduções incompreensíveis. Para minimizar o problema, muitos desses sistemas adotaram análise semântica e regras de transferência semântica, o que resolveria ambiguidades como as dos casos *ball* e *with a stick*. Mesmo assim, isso ainda depende do par de idiomas envolvidos e, em alguma extensão, só funciona em pares específicos, significando que a adição de um novo idioma fonte pode requerer modificação de sua análise semântica.

Concluiu-se, contudo, que os sistemas de transferência geralmente produzem traduções de melhor qualidade que as dos sistemas de arquitetura direta, embora ao preço da necessidade de desenvolver extensas técnicas de análise do idioma fonte e grande quantidade de regras de transferência.

2.1.3. A arquitetura interlíngua (IL)

A ideia por trás da arquitetura IL consiste em analisar o texto no idioma fonte de modo a produzir uma representação desse texto independente de seu idioma. Isso tem a significativa vantagem de permitir que a análise do texto fonte e a geração do resultado no idioma alvo sejam feitos uma única vez para cada idioma. Então, um sistema de tradução baseado na arquitetura IL pode ser construído apenas juntando análise e geração na representação IL. Muito embora haja muitas pesquisas na área, apenas alguns sistemas comerciais foram criados usando essa arquitetura (FUJITSU, 1986; FARWELL, 1990; NYBERG, 1991; NYBERG, 1992).

Tomando a frase *I like to eat* como exemplo, a abordagem IL assume haver um único conceito subjacente para o significado do verbo principal tanto na sentença no idioma fonte quanto na tradução no idioma alvo, transmitindo a ideia de que algo ou alguém (“comer”) agrada alguém (“eu”). Essa ideia pode ser escrita na representação IL da seguinte forma:

(1) like/gostar: [CAUSE (X, [BE (Y, [PLEASED])])])²

Tal abordagem dispensa regras de transferência, já que a representação é a mesma tanto no idioma fonte quanto no alvo. A única necessidade é a de se definir regras de ligação visando criar um mapa entre o texto de superfície (nos idiomas fonte e alvo) e a representação IL.

Todavia, a abordagem IL levanta uma questão: considerando que a sintaxe do idioma fonte não influi na representação IL, a geração do texto no idioma alvo baseada nessa representação geralmente assume a forma de paráfrase em vez de tradução (JOHNSON, 1985; ARNOLD, 1987; HUTCHINS, 1987). Em outras palavras, perde-se o estilo e a ênfase do texto original. No entanto, essa característica está mais para falta de compreensão do discurso e pragmatismo requeridos para reconhecer estilo e ênfase do que para falha da abordagem IL. Há casos em que ignorar o estilo do autor pode até ser uma vantagem. Há também quem defenda a tese de que, fora do contexto dos textos artísticos, é totalmente supérfluo preservar a forma sintática do texto fonte na tradução (WHITELOCK, 1989; NIRENBURG, 1990).

Outro problema posteriormente apontado diz respeito ao fato de que os autores dos textos fonte geralmente assumem que seus leitores têm bom conhecimento geral e estão bem informados sobre o assunto de seus escritos. Assim, Nirenburg (NIRENBURG, 1992) e uma boa parcela de pesquisadores da abordagem IL acham necessário implementar uma profunda análise semântica compatível com um profundo conhecimento do mundo a fim de tornar as traduções mais adequadas. É da viabilidade de representar, coletar e armazenar eficientemente grande quantidade de conhecimento e do domínio de interesse que depende o desempenho de um sistema com tais

2 Essa representação em forma simplificada e genérica apareceu em diversos trabalhos (SCHANK, 1972; SCHANK, 1973; SCHANK, 1975; JACKENDOFF, 1983; JACKENDOFF, 1990). Dorr (1993) apresenta um tratamento mais detalhado de tais casos.

características, tarefa que consome grandes esforços dos pesquisadores da área da inteligência artificial.

2.2. Processamento de Linguagem Natural (PLN)

Processamento de Linguagem Natural (PLN) é um ramo da Inteligência Artificial (IA) que reúne técnicas para interpretar e gerar linguagem natural e consiste em avaliar e criar estruturas capazes de serem utilizadas tanto pelos componentes de um sistema automático de tradução quanto por humanos (BARROS, 2001).

As bases de conhecimento necessárias ao funcionamento de um sistema PLN são:

- Léxico – obra de compilação (em maior ou menor extensão) do repertório de vocábulos de uma língua, registrados com suas definições e relacionados segundo alguns critérios, dentre os quais se inclui a ordem alfabética (OBJETIVA, 2002). Pode ser monolíngue ou bilíngue – neste último caso, o léxico contém também a tradução de cada termo para outro idioma;
- Gramática – conjunto de regras individuais usadas para um determinado uso de uma língua, não somente da norma culta, mas também de variantes não padrão;
- Modelo do domínio – representação do contexto, provendo conhecimento na forma de algum formalismo de Inteligência Artificial, tal como lógica de predicados, frames, etc (BARROS, 2001).

O PLN é dotado de três níveis de processamento, a saber:

- Morfológico – determina a classificação das palavras nas frases por meio de consultas ao léxico, graças ao que é possível obter todos os possíveis significados de cada termo. As ambiguidades são resolvidas geralmente por meio de métodos estatísticos. Depois, a frase classificada passa para os cuidados do analisador sintático;
- Sintático – conforme a gramática utilizada, a análise sintática gera uma estrutura sintática

(geralmente uma árvore) que resulta na determinação das categorias da frase;

- Semântico – determina o significado da frase dentro de seu contexto.

3. Tradução automática de texto: estado da arte

3.1. Paradigmas da tradução automática

Os diferentes sistemas de tradução automática podem ser comparados por diversos aspectos, dentre os quais a arquitetura. Mas um dos aspectos mais importantes é o paradigma de pesquisa. É importante entender que um não pressupõe o outro: enquanto a arquitetura diz respeito ao método de processamento – direto, de transferência ou IL, como visto no item 2.1 –, o paradigma de pesquisa refere-se aos componentes da informação que auxiliam o processamento.

Os paradigmas mais importantes podem ser divididos em três categorias (DORR, 1998):

1. Os que propõem o uso mais pesado de técnicas linguísticas;
2. Os que não usam nenhuma técnica linguística;
3. Os que usam uma combinação de ambos.

3.1.1. Paradigmas baseados em linguística

O paradigma baseado em linguística diz respeito à tradução baseada em princípios bem fundamentados de teoria linguística. Sistemas que adotam esse paradigma aplicam restrições de sintaxe, léxicas e semânticas para produzir no idioma alvo uma percepção apropriada da sentença no idioma fonte. Esta seção apresentará alguns dos principais paradigmas baseados em linguística.

3.1.1.1. A técnica das restrições

Técnicas baseadas em restrições (CBMT, do inglês *Constraint-Based Machine Translation*) surgiram em diferentes abordagens de tradução automática (KAPLAN, 1989; WHITELOCK, 1991; ARNOLD, 1992; EBERLE, 1992; SADLER, 1992; WHITELOCK, 1992; KAPLAN, 1993). Uma das primeiras a utilizar restrições na combinação de itens léxicos surgiu no sistema LFG-MT (KAPLAN, 1989; KAPLAN, 1993), que faz tradução bidirecional entre inglês, francês e alemão usando uma gramática léxica funcional (KAPLAN, 1982). O sistema LFG-MT é capaz de lidar com casos difíceis de tradução como o seguinte (DORR, 1998):

- (2) Inglês: *The baby just fell.*
Francês: *Le bébé vient de tomber.* (“*The baby just [verb-past] of fall*”)

Neste exemplo, o termo inglês *just* é traduzido como o verbo principal francês *venir*, fazendo do evento de cair complemento de *tomber*. Por basear-se em representações específicas, as operações de mapeamento requeridas na transferência pelo sistema LFG-MT precisam ser feitas por equações de transferência que relacionam as estruturas funcionais nos idiomas fonte e alvo. São elas que identificam *venir* como o predicado francês correspondente e restringem o argumento de *just* para que seja um complemento encabeçado pelo complementador preposicional *de*.

O sistema LFG-MT tem duas falhas, uma menos e outra mais séria. A menos séria diz respeito à associação entre as estruturas sintática e funcional usando um conjunto de restrições codificadas como entradas léxicas. Essa técnica tem a desvantagem de criar um forte ligação entre as estruturas funcional e sintática, de modo que, se uma certa ideia puder ser sintaticamente expressa em mais de uma forma, haverá mais de uma estrutura funcional. A falha mais séria refere-se à forma de lidar com casos como o do exemplo 2 no contexto de cláusulas embutidas (SADLER, 1992). Se a frase em inglês do exemplo 2 fosse *I think that the baby just felt*, não seria possível gerar a tradução em francês, pois o sistema LFG-MT quebraria a sentença em relações predicado-argumento que (grosseiramente) obedecem à seguinte estrutura lógica:

(3) think(I, fall(baby))
just(fall(baby))

Como o constituinte lógico `fall(baby)` é predicado de duas cabeças lógicas, *think* e *just*, o sistema é incapaz de determinar como compor tais conceitos e gerar uma tradução.

Esse entrave foi resolvido (KAPLAN, 1993) e refinado (DORNA, 1998) com o uso de um novo operador de descrição de idioma, *restriction*, tendo sido depois adotado no projeto TM Vermobil (DORNA, 1996).

3.1.1.2. A técnica do léxico

A técnica baseada em léxico (LBMT³, do inglês *Lexical-Based Machine Translation*), é a mais pesadamente usada em conjunto com algumas outras abordagens, incluindo a baseada em regras (RBMT, *Rule-Based Machine Translation*) (APPELO, 1986; ARNOLD, 1987; ARNOLD, 1988; ARNOLD, 1990), a baseada em frases (PBMT, *Phrase-Based Machine Translation*) (DORR, 1991; DORR, 1992a; DORR, 1993a) e a assim chamada “agitar e assar” (S&BMT, *Shake-&-Bake Machine Translation*) (BEAVEN, 1992a; BEAVEN, 1992b). No geral, qualquer sistema que forneça regras para relacionar as entradas léxicas de um idioma às de outro é um sistema baseado em léxico. Ainda que em diferentes graus de generalidade, diversos pesquisadores adotaram a técnica baseada em léxico (ABEILLÉ, 1990; DORR, 1990; FARWELL, 1990; WILKS, 1990; FARWELL, 1991; TSUJII, 1991; BLÁASER, 1992; DORR, 1992b; FARWELL, 1992; SANFILIPPO, 1992; TRUJILLO, 1992; DORR, 1993a; DORR, 1993b; FARWELL, 1993).

Um exemplo de sistema que adota o paradigma baseado em léxico é o LTAG (ABEILLÉ, 1990), usado para traduções bilíngues inglês-francês. Trata-se de um sistema de arquitetura de transferência que utiliza gramáticas síncronas de árvores contíguas (SHIEBER, 1990) para fazer um mapeamento simples de derivações gramaticais de árvores contíguas (JOSHI, 1985) de um idioma

3 O acrônimo LBMT também costuma ser usado na tradução automática baseada em linguística, sendo um termo de sentido mais genérico e que também se refere a abordagens de alguns outros paradigmas linguísticos. Neste trabalho, contudo, é usado em sentido mais específico, referindo-se apenas aos sistemas baseados em linguística calcados principalmente no léxico.

para outro. Um um léxico bilíngue que associa diretamente as árvores fonte e alvo através de ligações entre os itens léxicos e seus argumentos é usado para fazer esse mapeamento. Em termos gerais, um mapeamento entre uma sentença no idioma fonte e uma no idioma alvo está contido em cada entrada bilíngue.

Esse tipo de abordagem é capaz de lidar com casos como este (DORR, 1998):

- (4) Inglês: *John is fond of music.*
Francês: *John aime la musique.* (“*John loves the music.*”)

Neste caso, a ideia central da frase no idioma fonte é percebida como a forma adjetiva *be fond of*, enquanto a tradução no idioma alvo percebe essa ideia como o verbo *aimer*. A regra de transferência usada neste mapeamento liga diretamente a frase adjetiva *fond of* na árvore do idioma fonte com o verbo *aimer* na do idioma alvo.

A vantagem dessa abordagem é saber lidar com frases modificadas como a seguinte (DORR, 1998):

- (5) Inglês: *John is very fond of music.*
Francês: *John aime beaucoup la musique.* (“*John loves very much the music.*”)

No exemplo, o advérbio inglês *very* está ligado ao predicado *fond of* ao invés de ao verbo principal *be*, enquanto que o correspondente adverbial francês *beaucoup* está associado ao verbo principal *aimer*. É a ligação entre a frase adjetiva *fond of* e o verbo *aimer* que permite que essa modificação funcione. Não havendo ligação alguma com o verbo principal *be*, o modificador precisa estar associado à frase adjetiva *fond of*.

Mas essa abordagem tem uma desvantagem: exigir que, para cada par de idiomas, árvores inteiras estejam armazenadas no dicionário de transferência. Quando o número de pares começa a aumentar, isso pode se tornar oneroso.

3.1.1.3. A técnica baseada em regras

O paradigma da tradução automática baseada em regras (RBMT, de *Rule-Based Machine Translation*) faz parte de sistemas que adotaram níveis linguísticos diversos para efetuar tradução entre pares de idiomas (DANLOS, 1982; APPELO, 1986; ARNOLD, 1987; ARNOLD, 1988; KAPLAN, 1989; LANDSBERGEN, 1989; McCORD, 1989; ARNOLD, 1990; FUJII, 1990; GRISHMAN, 1990; THURMAIR, 1990; VAN NOORD, 1990; OKUMURA, 1992; KAPLAN, 1993; ROSETTA, 1994). O Rosetta (ROSETTA, 1994) era um protótipo que dividia as regras de tradução em duas categorias: as “regras sem significado”, que mapeiam itens léxicos para árvores sintáticas, e as “regras que preservam significados”, que fazem o mapa entre as árvores sintáticas para as estruturas de significados subjacentes.

Essa separação feita pelo Rosetta é herança da noção de “composicionalidade relaxada” presente em sistemas de transferência como o Eutrota e seu descendente MiMo (ARNOLD, 1988; ARNOLD, 1990). O comportamento entrada-saída de ambos os sistemas são similares, mas o Rosetta produz uma representação IL para caracterizar o significado da frase de entrada, de modo a permitir que eventuais ambiguidades sejam removidas interativamente depois.

O seguinte exemplo mostra caso de divergência entre as categorias de regras no qual duas cabeças de frase são trocadas (DORR, 1998):

- (6) Inglês: *Mary happened to come.*
Holandês: *Mary kwam toevallig.* (“*Mary came by chance.*”)

Neste caso, o Rosetta usa uma representação IL que é uma forma canônica correspondente a ‘by-chance(Mary, come)’, de modo que a sintaxe inglesa torna-se equivalente à forma canônica: a construção verbal correspondente a ‘by-chance’ (*happen to*) usa como argumento a cláusula correspondente a ‘(Mary, come)’ (*came*). O mesmo não ocorre na sintaxe holandesa por não estar em sincronia com a forma canônica, já que o verbo correspondente a ‘(come, Mary)’ (*kam*) usa como argumento a construção adverbial correspondente a ‘by-chance’ (*toevallig*).

A maneira de lidar com tais casos é interromper o processamento normal e invocar uma “regra de troca”, passando o controle a um módulo que origina uma nova categoria de regra que domina a regra da cabeça sintática. Mas essa técnica tem o problema de deixar em aberto a questão de como as interrupções feitas gramaticalmente interagem com os requisitos idiossincráticos dos itens léxicos individuais. Um aspecto ainda mais problemático dessa abordagem é a dificuldade, senão a impossibilidade, de acomodar casos de troca de cabeça nos quais o “desviante” serve como cabeça na sintaxe e não na forma canônica (DORR, 1998).

Por causa disso, os desenvolvedores do Rosetta foram forçados a considerar tais casos de incompatibilidade como puramente gramaticais. O projeto básico do Rosetta, que era um sistema bem desenvolvido e cobria uma ampla gama de fenômenos linguísticos, podia ser aperfeiçoado estendendo ao léxico a noção de composicionalidade.

3.1.1.4. A técnica “agitar e assar”

A técnica *Shake-and-Bake Machine Translation* (S&BMT) (WHITELOCK, 1991; WHITELOCK, 1992; BEAVEN, 1992a; BEAVEN, 1992b; BREW, 1992) é um exemplo perfeito da razão pela qual faz-se distinção entre paradigma e arquitetura. Embora os criadores da abordagem S&BMT afirmem ser uma alternativa às arquiteturas de transferência e IL, ela só funciona graças justamente às regras de transferência. A única diferença é o algoritmo de mapeamento dos itens léxicos e a subsequente formação da sentença no idioma alvo (DORR, 1998).

As “entradas léxicas bilíngues” são a base da definição das regras de transferência. Após a análise da frase no idioma fonte, as palavras dessa frase são mapeadas no idioma alvo por meio de entradas bilíngues pelo algoritmo que combina as palavras no idioma alvo e tenta ordená-las com base nas restrições sintáticas do idioma alvo.

A razão da existência da técnica S&BMT é a necessidade de lidar com traduções complexas como a do exemplo *I like to eat* mostrado anteriormente. Ao contrário do procedimento da abordagem de transferência, o algoritmo S&BMT resolve o problema selecionando palavras no dicionário léxico do idioma alvo e tentando diferentes ordenações dessas palavras (o “agitar” do

nome da técnica) até produzir uma sentença que satisfaça todas as restrições sintáticas (o “assar”) (DORR, 1998).

A ideia por trás dessa abordagem é que, uma vez que os elementos bilíngues identifiquem corretamente os índices das entradas léxicas, o algoritmo S&BMT assume a tarefa de “combinar” esses elementos. A vantagem é que o lexicógrafo bilíngue só precisa especificar o conhecimento contrastante entre os dois idiomas, a partir do que as gramáticas monolíngues para análise e geração fazem o resto (DORR, 1998). Segundo os criadores dessa técnica, os mapeamentos bilíngues são suficientemente restritos para permitir a aquisição automática de correspondentes bilíngues de corpos alinhados (BEAVEN, 1992b; WHITELOCK, 1992).

No entanto, essa técnica tem uma desvantagem: não há um algoritmo geral no S&BMT que funcione em todos os casos. Mesmo assim, é possível estabelecer restrições no idioma alvo que possibilitem formar a entrada para a geração. Consideráveis melhorias podem ser obtidas por meio de uma abordagem heurística baseada em propagação de restrições (BREW, 1992). Para lidar com ambiguidades na tradução, outros refinamentos foram propostos (EMELE, 1998).

3.1.2. Paradigmas não baseados em linguística

O rápido desenvolvimento do poder computacional e a disponibilidade de dicionários e corpos léxicos monolíngues e bilíngues capazes de serem lidos por computadores permitiu aos pesquisadores investigar os paradigmas da TM que não são baseados em teorias linguísticas. Grandes corpos de texto são usados para treinamento e como base de dados de traduções existentes. É da existência deles que dependem todas as abordagens. Algumas das pesquisas em paradigmas não baseados em linguística são descritas nesta seção.

3.1.2.1. A técnica baseada em estatística

As análises estatísticas de corpos bilíngues paralelos são pesadamente usadas na produção de traduções baseadas em técnicas de predição estatística (SBMT, de *Statistical-Based Machine*

Translation). Houve um esforço inicial de investigação dessa técnica em meados da década de 1950 (KING, 1956), mas ela só ganhou impulso a partir de 1988, quando a IBM iniciou seu Candidate French-English Machine Translation Project (BROWN, 1988a; BROWN, 1988b; BROWN, 1990; BROWN, 1992). Outras pesquisas em SBMT também foram relatadas (GRISHMAN, 1990; NOMIYAMA, 1992; SU, 1992; CHANG, 1993; KOMATSU, 1993; MARUYAMA, 1993; DOI, 1994).

A abordagem SBMT deriva das técnicas de processamento de fala (DORR, 1998). Uma variante da regra de Bayes é usada para mostrar que a probabilidade de um conjunto de palavras (T) ser uma tradução de outro conjunto de palavras (S) é proporcional ao produto da probabilidade de um conjunto de palavras no idioma alvo ser uma expressão legal e a probabilidade de um conjunto de palavras no idioma fonte ser uma tradução do conjunto no idioma alvo. Ou seja:

$$(7) \quad P(TS) \propto P(T) \times P(S - T) -$$

Sendo conhecidas as probabilidades da direita, obtém-se a tradução escolhendo T tal que a probabilidade da esquerda seja maximizada. Considerando que as probabilidades de ambos os lados não podem ser sempre conhecidas, precisam ser estimadas para que a abordagem funcione. Geralmente isso se faz definindo-se modelos probabilísticos aproximados obtidos de probabilidades que possam ser diretamente estimadas a partir do que já se conhece.

3.1.2.2. A técnica baseada em exemplo

A técnica baseada em exemplo (EBMT, de *Example-Based Machine Translation*) surgiu na década de 1980 e faz traduções “por analogia”, ou seja, emula a tradução humana ao reconhecer a semelhança de uma frase no idioma fonte com traduções previamente feitas (NAGAO, 1984). Uma demonstração simples de traduções japonês-inglês produzidas com essa técnica foi feita no fim daquela década (SATO, 1990) e diversas outras investigações adicionais foram relatadas (GRISHMAN, 1990; SUMITA, 1990; FURUSE, 1992; JONES, 1992; MARUYAMA, 1992; McLEAN, 1992; NOMIYAMA, 1992; OKUMURA, 1992; SOMERS, 1992; NIRENBURG, 1993; RICHARDON, 1993; SUMITA, 1993; YASUHARA, 1993).

O conceito básico da técnica é a busca por traduções semelhantes num banco de dados. As traduções são então modificadas e combinadas para formar uma tradução mais próxima da adequada para a sentença no idioma fonte, técnica bem parecida com a Case Based Reasoning usada em inteligência artificial (KOLODNER, 1993). A similaridade é determinada semanticamente por alguma métrica baseada em léxico ou ontologia.

Se essa tradução é precisa e de qualidade depende fortemente da qualidade do banco de dados. Embora esse banco não precise ser tão grande quanto o requerido para a SBMT (uma vez que não é necessário cobrir todo o vocabulário), dificuldades podem surgir em função de divergências sintáticas e semânticas (DORR, 1998). A proximidade da correspondência requer pelo menos uma análise sintática e semântica básica das traduções paralelas⁴. Sugeriu-se pesar as traduções frasais por meio de sua frequência no banco de dados para evitar a correspondência de divergências impróprias (COLLINS, 1997).

Além da correspondência de frases, a EBMT requer que haja correspondência também na estrutura sintática das sentenças no idioma fonte e no banco de dados. Embora uma correspondência total tenha sido possível (FURUSE, 1994), a EBMT enfrenta problemas com palavras de função (SUMITA, 1993), frases nominais (SATO, 1993) e frases preposicionais (SUMITA, 1993).

3.1.2.3. A técnica baseada em rede neural

Análise (JAIN, 1991), remoção de ambiguidades léxicas (HIGINBOTHAM, 1990) e aprendizado de regras gramaticais foram alguns dos experimentos feitos com tecnologia de rede neural para tradução automática de textos. Alguns testes com pequenos vocabulários e sintaxe simples foram feitos (CASTAÑO, 1997). Pequenos porque grandes gramáticas inflam dramaticamente o tamanho das redes neurais e o gasto de tempo em treinamento. A representação da rede neural fica ainda mais complicada com as representações extras de tempo que as sequências de palavras requerem. Por esse motivo, usou-se um pequeno vocabulário (30 palavras) e sentenças curtas para encontrar sistemas similares via rede neural num sistema EBMT (McLEAN, 1992). Ao contrário de outras abordagens descritas neste trabalho, não há registro de algum sistema realístico

⁴ Essa análise frustraria o propósito da EBMT, propondo-se que essa correspondência seja feita usando sinônimos e hipônimos (NIRENBURG, 1993).

de tradução automática baseado inteiramente em rede neural.

3.1.3. Paradigmas híbridos

Como visto na discussão sobre os paradigmas não baseados em linguística, muitos desses paradigmas têm dificuldade em lidar com alguns dos processos da tradução automática. Por exemplo: o SBMT não lida bem com longas dependências contextuais e o EBMT tem problemas com estruturas de sentenças complexas. Logo, chegou-se à conclusão de que esses paradigmas não linguísticos poderiam ser combinados com os linguísticos, aproveitando o melhor de cada um (GRISHMAN, 1990; CARBONELL, 1992; FURUSE, 1992; LEHMANN, 1992; NIRENBURG, 1993). O paradigma híbrido é um misto dos paradigmas e das arquiteturas da tradução automática. A abordagem mais comum é usar métodos linguísticos para obter análises do texto fonte e usar técnicas estatísticas ou de exemplo para efetuar as traduções (FURUSE, 1992). Modelos de trigramas estatísticos de idioma alvo foram usados para seleção léxica (BROWN, 1995). Árvores de decisão geradas estatisticamente foram usadas para inserir artigos em inglês em traduções de textos em japonês (KNIGHT, 1994). O sistema Pangloss (NIRENBURG, 1994) é um híbrido dos paradigmas e das arquiteturas de tradução automática.

4. Histórico da evolução do algoritmo de tradução do Falibras

4.1. O sistema Falibras

O projeto Falibras foi concebido para ser um tradutor da linguagem falada em português para libras. Seu funcionamento é simples, porém interessante: uma pessoa fala ao microfone de um computador e o Falibras exibe no monitor, em tempo real, a tradução na língua de sinais por meio de uma animação gráfica. A Figura 3 oferece uma amostra da interface do Falibras.

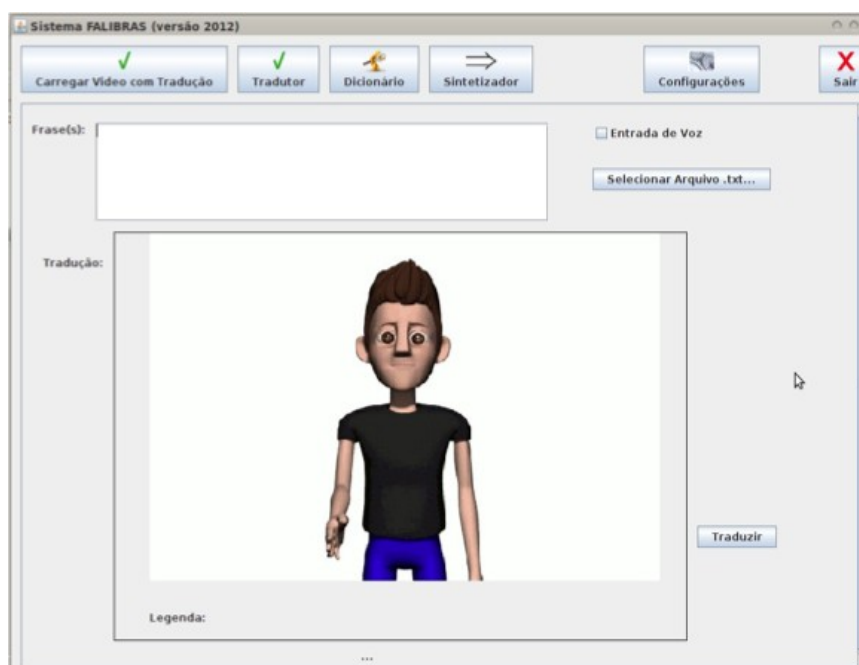


Figura 3: janela da interface do Falibras (BRITO *et al.*, 2012)

Mais do que simplesmente um tradutor automático português-libras, o Falibras permite inclusive realizar projetos de educação especial para portadores de deficiência auditiva, especialmente crianças, tornando-o ferramenta valiosa na integração delas à escola e auxiliando em seu aprendizado cognitivo. Trabalhos similares, como o Tlibras, são inspirados no conceito do Falibras (LIRA, 2013).

O Falibras está sendo desenvolvido utilizando tecnologias de reconhecimento de fala,

gerenciamento de bancos de dados e elaboração de animações gráficas. Além disso, a análise da estrutura do idioma libras é essencial e faz parte do desenvolvimento do projeto. Técnicas de inteligência artificial também são necessárias para lidar com as diferenças semânticas entre os idiomas.

Lançado em 2001 (CORADINE, 2001; CORADINE, 2002), o Falibras utilizava em sua primeira versão um módulo interpretador de fala ou de textos digitados cujas limitações impediam-no de processar a maior parte do contexto das frases, aproveitando apenas palavras isoladas e expressões ou frases curtas. O resultado do processamento feito pelo módulo interpretador era enviado ao módulo de exibição, que fazia o boneco animado “falar” a tradução em libras.

A segunda versão do Falibras, desenvolvida entre 2003 e 2004, teve seus esforços de desenvolvimento concentrados na melhoria da qualidade da tradução. Lançou-se mão de técnicas trazidas do Processamento de Linguagem Natural (PLN), área da Inteligência Artificial dedicada a estudar a compreensão e/ou geração automática de textos em línguas naturais (GAZDAR, 1989).

O período entre 2005 e 2007 viu surgir a terceira versão do Falibras, caracterizada pela ênfase na análise sintática dos textos no idioma fonte. A ideia por trás desta terceira versão era construir árvores sintáticas baseadas em gramática livre de contexto. Como resultado da aplicação de regras de tradução, obtinha-se as estruturas de frases em libras que eram exibidas pela animação gráfica. Assim foi que o Falibras ganhou a capacidade de efetuar traduções baseadas em sintaxe, cuja qualidade foi aprimorada por meio da aplicação de técnicas de tradução automática baseada no método de tradução por transferência sintática. Mesmo assim, o resultado ainda ficava limitado em desempenho e no tamanho da gramática pela rigidez das regras sintáticas e de tradução para mapeamento entre português e libras devido ao uso da linguagem de programação Prolog.

Uma posterior parceria com professores da Universidade Federal do Espírito Santo (UFES) permitiu o desenvolvimento de uma versão do Falibras dotada de memória de tradução. Essa nova versão foi batizada de Falibras-MT. Sua principal característica era possuir recurso usado principalmente em tradutores automáticos para textos técnicos por meio do qual não se faz uma tradução direta de textos, mas consulta-se um banco de dados formado por traduções previamente

preparadas por um tradutor humano e que servem de base para traduções similares – daí “memória de tradução”. O software então escolhe o excerto mais adequado ao contexto do texto original e faz a tradução ser “falada” em libras pela animação gráfica – esta gerada por meio de um dicionário de videos. Diferente do Falibras original, que não se baseava em banco de dados de traduções, o Falibras-MT é dotado de um módulo que permite editar o banco, aprimorando-o.

A versão atual do Falibras é mais modular e fácil de desenvolver. Dentre as melhorias da nova versão destacam-se:

- Animação 3D;
- Módulo de tradução mista – combinação de tradução por transferência sintática com memória de tradução;
- Módulo de apoio a intérpretes – tem por objetivo facilitar a preparação de intérpretes em libras para um evento específico e aperfeiçoar a qualidade da tradução do sistema;
- Módulo de acesso à Internet – propõe-se a ajudar os deficientes auditivos no acesso à informação online e funciona na forma de *plugin* para o navegador Firefox.

A proposta inicial do Falibras foi tomada com base na arquitetura apresentada na Figura 4 e consiste em duas partes: a interface tradutora e o analisador morfossintático.

O Falibras é formado por analisador léxico, analisador sintático, analisador de contexto e gerador de tradução em libras (CORADINE, 2002), além de utilizar técnicas baseadas em processamento de linguagem natural com a linguagem de programação Prolog (COELHO, 1988) e de gramáticas que fazem o processamento de expressões para ser possível a criação de analisadores sintáticos. Uma aplicação denominada InterProlog para a comunicação das aplicações com interpretadores Prolog permite implementar uma gramática livre de contexto (GLC) através de redes de transição recursivas.

O código em Prolog define um autômato finito que reconhece linguagens geradas por gramáticas regulares, fazendo com que as regras de tradução sejam especificadas, percorram o

autômato e reconheçam o texto, gerando a representação em libras.

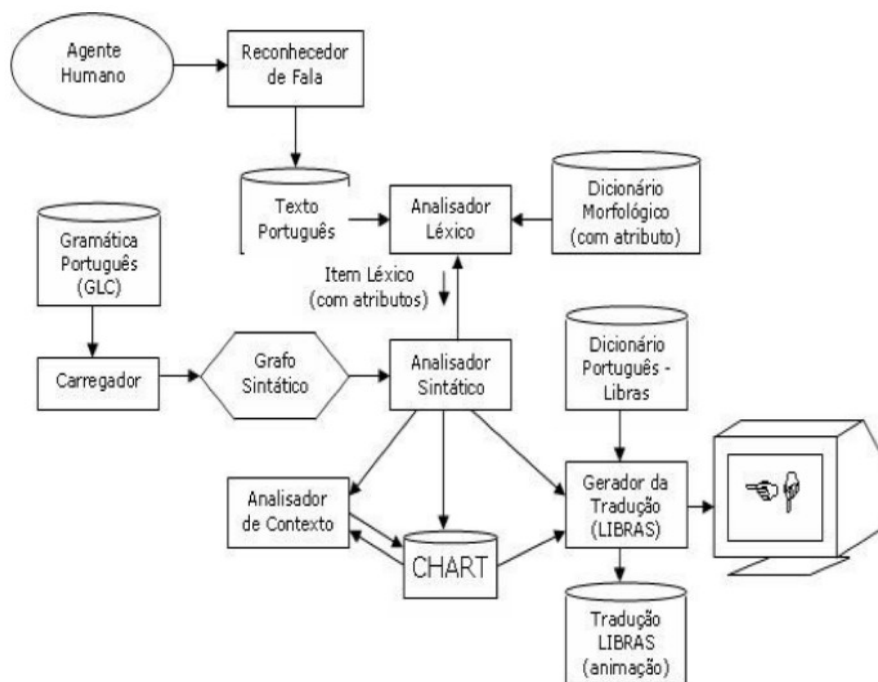


Figura 4: diagrama da arquitetura original do Falibras (RODRIGUES *et al.*, [200-?])

Uma outra abordagem desenvolvida foi o Falibras-MT, que possui como componentes os dicionários de símbolos e de vídeos, com a interface tradutora baseada em memória de tradução. O sistema é constituído pelos módulos de tradução e autoria da memória de tradução (CORADINE, 2002). Assim, nessa arquitetura são encontrados o texto parcial e a memória de tradução, que é constituída pelos dicionários de símbolos e de vídeos.

4.2. Tradução baseada em regras

O Falibras permite o aprendizado de idiomas e a generalização do processo de tradução, possibilitando a tradução de qualquer idioma escrito para qualquer outro idioma, seja escrito, falado ou sinalizado. Essa generalização também pode ser aplicada ao processo de inferência gramatical, o que permite inferir as regras de tradução dos exemplos contidos no banco de dados de tradução. Assim, embora o Falibras use o banco para traduzir textos de português para libras, os mecanismos utilizados pela tradução não limitam os idiomas envolvidos, apenas os formatos de entrada e saída.

Como exposto anteriormente, a partir de sua terceira versão o Falibras permitiu a construção de árvores sintáticas por meio das quais, através de regras de tradução, produz-se as estruturas frasais em libras. Com essa abordagem, o Falibras passou a fazer traduções baseadas em sintaxe, ainda que as regras sintáticas e de tradução entre português e libras fossem fixas e especificadas diretamente pelo sistema, resultando em restrições práticas de desempenho e limitações relativas ao tamanho da gramática.

4.3. Transferência sintática

A transferência sintática implementada no Falibras é uma aproximação da ideia por trás das *synchronous TAGs*, que são encaradas como léxicos de transferência, como mostra a Figura 5 adaptada (ABEILLÉ, 1990) para o contexto do Falibras (RIBEIRO, 2007). O processo de transferência é a própria derivação. As regras de transferência são as ligações entre as árvores num par. O processo de transferência é ilustrado na Figura 6.

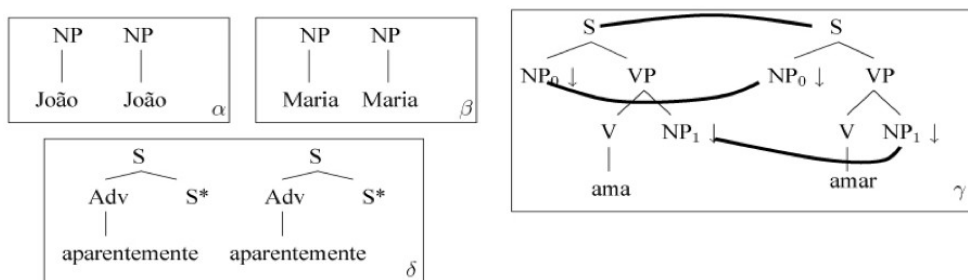


Figura 5: um léxico de transferência (RIBEIRO, 2007)

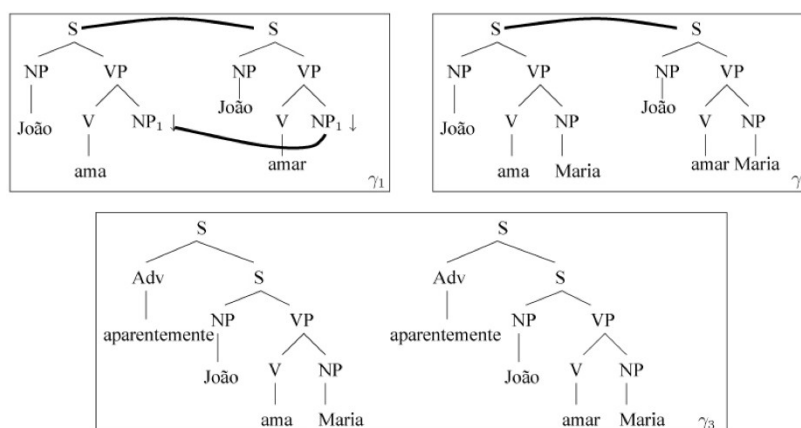


Figura 6: derivação do léxico de transferência (RIBEIRO, 2007)

A análise sintática provoca a “quebra” da árvore em componentes triviais. Então é escolhida

uma nova árvore derivada da principal, que geralmente é a que possui o verbo e que equivale ao par de árvores onde ocorrem as derivações em *synchronous TAGs*. Nesta etapa não se define uma ligação entre as árvores por não existir uma árvore de destino.

A criação de um identificador para a árvore principal – uma *string* contendo uma representação linear da árvore – é o ponto-chave da aproximação e é utilizado para a escolha da regra de transferência, que deve:

1. Construir a árvore alvo obedecendo a gramática do idioma alvo, rearranjando a árvore principal;
2. Realizar a operação de substituição nos nós da árvore principal com os componentes triviais gerados na análise.

Assim fazendo, gerar a tradução torna-se uma simples tarefa de percorrer as folhas da árvore alvo. Para tanto, informação léxica suficiente deve estar contida nas folhas. Como o idioma libras requer também fixação verbal, a necessária tarefa de fixá-la é realizada fornecendo tempo, modo, pessoa e número do verbo.

A arquitetura aqui descrita apresenta duas dificuldades (DORR, 1998):

1. Forte dependência entre a análise e transferência. Uma implementação baseada fortemente em *synchronous TAGs* ajudaria a simplificar essa dependência, já que a atividade de derivação é a ocorrência simultânea de análise e transferência;
2. O número de regras de transferência é enorme.

4.4. Tradução baseada em memória de tradução

O acesso ao módulo tradutor se faz por meio de uma janela na qual o usuário seleciona a memória de tradução desejada dentre as disponíveis no banco de dados, digita o texto de entrada no idioma fonte e solicita a tradução em libras. Como resultado, um campo mostrará os símbolos em libras e um vídeo com uma animação gráfica desses símbolos será exibido.

Quando o sistema é acessado por um colaborador, oferece-se a ele duas opções: traduzir textos ou editar a memória de tradução. Mesmo durante o uso do sistema, o colaborador pode alternar entre as tarefas de tradução de texto e edição da memória de tradução, o que facilita o teste do conhecimento embutido nessa memória. Essa opção não está disponível para o usuário comum.

O colaborador pode acessar o menu principal do editor, bem como quatro abas de dicionários: símbolos, vídeos, regras e termos. Pode também acessar as opções de navegação na ajuda do sistema. Nas janelas dos dicionários de símbolos e de vídeos, o colaborador pode gerenciar os respectivos conteúdos desses dicionários na memória de tradução. A janela do dicionário de regras pede ao colaborador para informar a classificação ou tradução de um termo quando não for capaz de inferir sua classificação ou traduzi-lo. O colaborador pode ainda adicionar novos classificadores gramaticais quando necessário.

As opções do dicionário de termos permitem ao colaborador gerenciar o conteúdo dos termos da memória de tradução, bem como adicionar novos classificadores gramaticais. Quando o colaborador aciona o processo de inferência na aba do dicionário de regras, a janela é preenchida com todos os termos adicionais cuja classificação e tradução foram informados ao sistema pelo colaborador.

5. Proposta de solução híbrida

Conforme exposto no item 3.1.3, o paradigma híbrido é um misto dos paradigmas e das arquiteturas da tradução automática, tendo como abordagem mais comum a combinação de métodos linguísticos para obter análises do texto fonte com técnicas estatísticas ou de exemplo para efetuar as traduções (FURUSE, 1992).

A solução híbrida, lançada em 2012, é a mais atual. Tal solução combina as vantagens das soluções baseadas em regras e em memória de tradução. Foi desenvolvida no contexto de um trabalho de mestrado do aluno Vitor Marcolino, em fase de finalização. Conforme já exposto, além das técnicas de tradução baseada em regras e em transferência sintática, o Falibras baseia-se também num banco de dados – chamado de memória de tradução – formado por traduções previamente inseridas. Quando requisitado, o software busca traduções semelhantes nessa memória. As traduções são então modificadas e combinadas para formar uma tradução mais próxima da adequada para a sentença no idioma fonte.

A Figura 7 apresenta a solução híbrida adotada no Falibras. Como mostra a figura, sua arquitetura sofreu influência de três estilos e padrões arquiteturais: MVC, Centrada em Dados e Cliente-Servidor. A Figura 8, por sua vez, mostra o detalhamento dos componentes internos da arquitetura.

Uma analogia entre as Figuras 7 e 8 mostra que as interfaces do sistema estão na camada de visão, enquanto que funcionalidades básicas como editores de regras de tradução, editor gramatical, analisador gramatical e semântico, corretor ortográfico e gramatical, sintetizador automático de texto e tradutor estão na camada de controle. Na camada de modelo, por sua vez, ficam os bancos de dados.

Nessa arquitetura, o usuário interage com o Falibras na camada de visão, recebendo do usuário sua entrada de texto, seja escrito ou falado. A camada de visão, por sua vez, aciona a de controle, localizando os editores que farão a busca no banco de dados localizado na camada de

modelo. O resultado desse processo passará pelos analisadores e estes acionarão a saída na forma de animação gráfica.

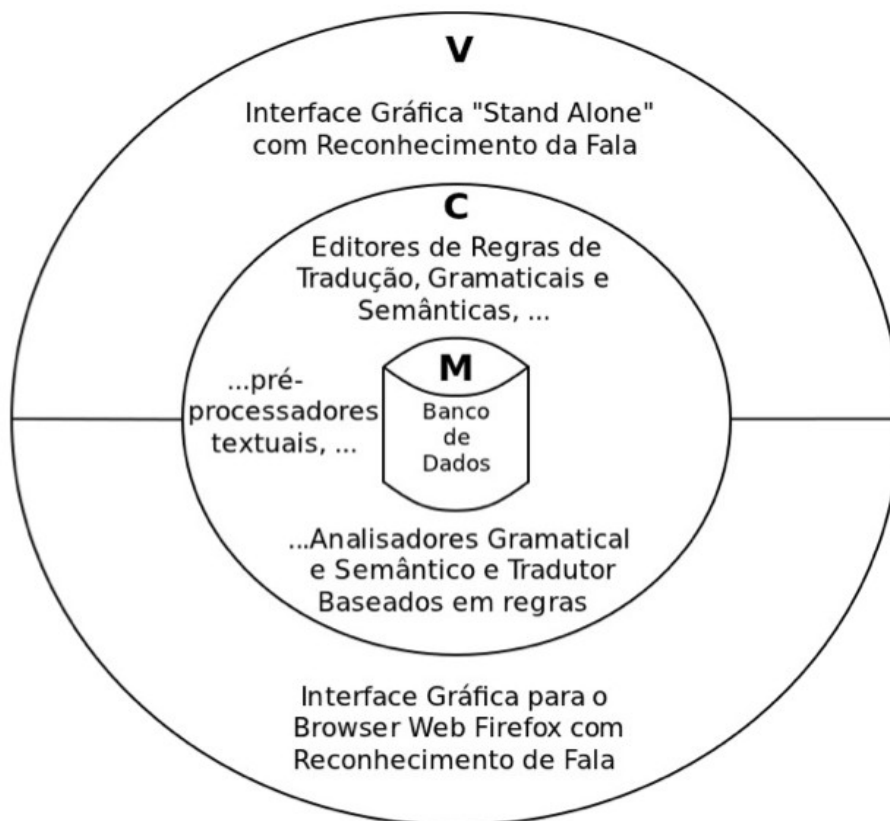


Figura 7: arquitetura de referência do Falibras (BRITO *et al.*, 2012)

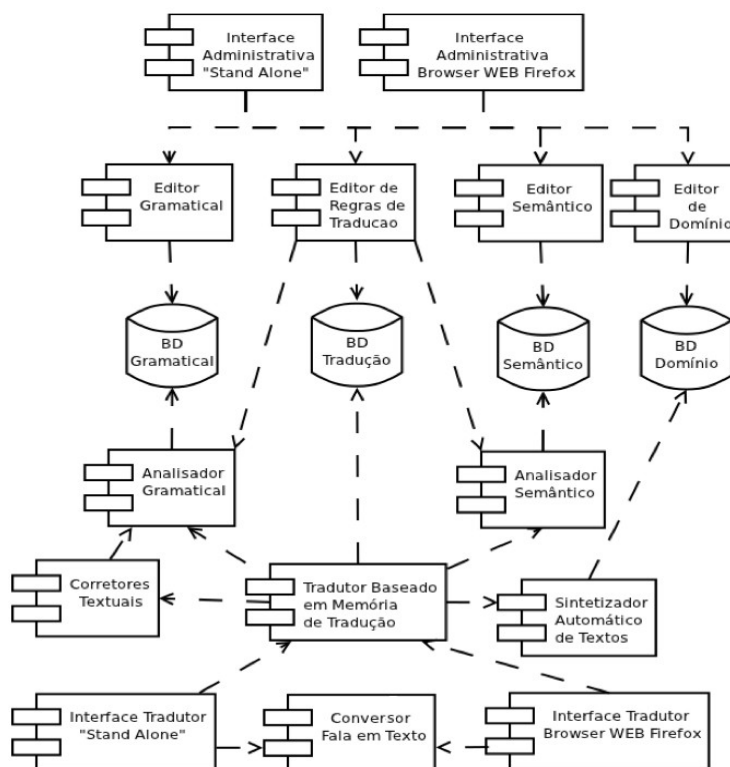


Figura 8: detalhamento dos componentes do Falibras (BRITO *et al.*, 2012)

Uma ilustração da interação entre os componentes pode ser dada pela descrição do funcionamento do componente Tradutor Baseado em Memória de Tradução, visto na Figura 6. O componente solicita de Corretores Textuais o texto no idioma fonte já corrigido ortográfica e sintaticamente. Depois, o componente Analisador Gramatical classifica as palavras do texto segundo o contexto sintático das frases. Finalmente, o texto está pronto para ser traduzido. Antes, porém, o usuário pode optar por traduzir o texto na íntegra ou gerar uma síntese automática dele segundo algum domínio específico, síntese essa realizada pelo componente Sintetizador Automático de Textos. Só então acontece a tradução, feita frase a frase, com cada palavra classificada sintática e semanticamente. Essa classificação é importante, uma vez que as regras de tradução são influenciadas por elas, refinando o princípio da memória de tradução (BREDA, 2005).

A adoção de um modelo híbrido garantiu ao Falibras o aproveitamento do melhor de cada paradigma – linguístico e não linguístico –, resultando num sistema de tradução automática português-libras suficientemente confiável para ser oferecido como ferramenta de apoio na educação de deficientes auditivos, inclusive crianças.

A Figura 9 apresenta o processo de tradução do componente tradutor do Falibras. Após receber o texto de entrada no idioma fonte, avalia-se a possibilidade do texto representar uma exceção à regra. Se não, verifica-se a existência de exemplos de tradução que possam servir de analogia para a tradução. Nos casos em que a base de regras não contenha nem exceções nem regras pelas quais possa-se inferir a estrutura de tradução do texto de entrada, a tradução será processada de acordo com as regras de tradução padrão, utilizando regras de tradução envolvendo as classes morfossintáticas das palavras da frase.

O módulo de apoio a intérpretes do Falibras também se vale de regras para desempenhar seu papel no do sistema. Ele permite ao intérprete carregar o texto a ser interpretado e verificar as regras sintáticas do português e as regras de tradução que são ativadas para cada frase. A ausência de regras implica na necessidade de estudo prévio e, se julgar oportuno, adicionar novas regras (sintáticas e/ou de tradução).

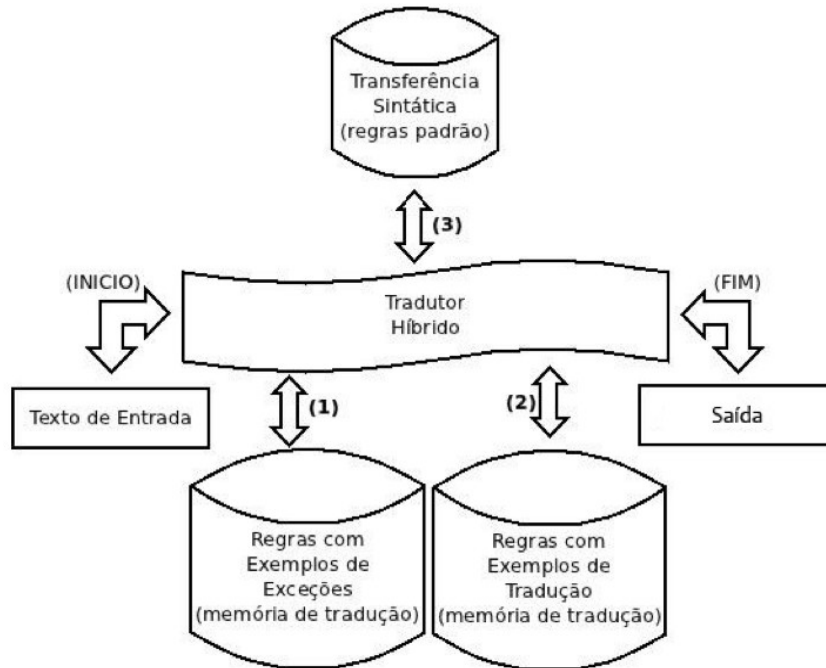


Figura 9: processo de tradução do Falibras (BRITO *et al.*, 2012)

6. Conclusões e trabalhos futuros

Este trabalho teve o intuito de explicar e exemplificar arquiteturas e paradigmas da tradução automática bilíngue de textos dentro do contexto do sistema Falibras de tradução português-libras.

O trabalho ressaltou o potencial do Falibras como ferramenta de apoio pedagógico na educação e aprendizagem de pessoas portadoras de deficiência auditiva, às quais o Falibras é voltado, propiciando-lhes um meio eficaz e aprender os mais diversos idiomas, em especial português e libras.

A monografia mostrou ainda detalhes da arquitetura do Falibras, seus componentes e sua arquitetura híbrida de tradução. Tais características permitem-lhe alcançar resultados relevantes na facilitação da aprendizagem de português e libras e na promoção do exercício da cidadania pelos deficientes auditivos ao dar-lhes acesso à informação de que precisam para efetuar sua inclusão social e digital.

Por seu caráter continuamente evolutivo, o Falibras tende a ter sua eficiência e relevância permanentemente aumentadas. É nosso propósito fomentar esse progresso e desfrutar da satisfação de contribuir para elevar a qualidade de vida dos portadores de uma deficiência cujas implicações na vida dessas pessoas só podemos imaginar. Muito mais que um mero programa de computador, cremos ser o Falibras também um projeto social do qual nos orgulhamos em participar.

Como trabalho futuro ligado diretamente a este, pretendemos refinar o comparativo entre as técnicas de tradução a fim de analisar cada uma das soluções em termos de consumo de memória, desempenho, usabilidade, facilidade de desenvolvimento e flexibilidade para regionalismos.

Referências

- ABEILLÉ, A. *et al.* Using Lexicalized TAGs for machine translation. In: *Proceedings of the 13th conference on Computational Linguistics*, páginas 1–6, Morristown, NJ, USA. Association for Computational Linguistics, 1990.
- ABEILLÉ, A. Using Lexicalized Tags for Machine Translation. In: *Proceedings of Thirteenth International Conference on Computational Linguistics (COLING-90)*, páginas 1-6, Helsinque, Finlândia, 1990.
- APPELO, L. A Compositional Approach to the Translation of Temporal Expressions in the Rosetta System. In: *Proceedings of Eleventh International Conference on Computational Linguistics*, páginas 313-318, Bonn, Alemanha, 1986.
- ARNOLD, D. *et al.* “Relaxed” Compositionality in Machine Translation. In: *Proceedings of the Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Carnegie Mellon University, Pittsburgh, PA, 12-14 de junho de 1988, Sessão 3: EUTROTA Perspectives.
- ARNOLD, D., DES TOMBE, L. Basic Theory and Methodology in Eutrotra. In: S. Nirenburg, editor, *Machine Translation: Theoretical and Methodological Issues*, páginas 114-135. Cambridge University Press, Cambridge, England, 1987.
- ARNOLD, D., DES TOMBE, L. Basic Theory and Methodology in Eutrotra. In: S. Nirenburg, editor, *Machine Translation: Theoretical and Methodological Issues*, páginas 114-135. Cambridge University Press, Cambridge, England, 1987.
- ARNOLD, D., SADLER, L. Rationalism and the Treatment of Referential Dependencies. In: *Proceedings of the Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 195-204, Montreal, Canadá, 1992.
- ARNOLD, D., SADLER, L. Theoretical Basis of MiMo. *Machine Translation*, 5(3):195-222, 1990.
- BARROS, F., ROBIN, J. Processamento de linguagem natural, *Revista Eletrônica de Iniciação Científica* 1(2), 1–58, 2001
- BEAVEN, J. *Lexicalist Unification-Based Machine Translation*. Tese de Doutorado, Universidade de Edinburgh, Edinburgh, 1992a.
- BEAVEN, J. Shake and Bake Machine Translation. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, páginas 603-609, Nantes, França, 1992b.
- BENNETT, P. A. *et al.* *Multilingual Aspects of Information Technology*. Gower, Bookfield, VT, 1986.
- BLÁASER, B. *et al.* A Reusable Lexical Database Tool for Machine Translation. In: *Proceedings of the Fourteenth International Conference on Computational Linguistics*, páginas 510-516, Nantes, França, 1992.
- BRASIL. Ministério do Planejamento, Orçamento e Gestão. Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2010. Disponível em: <http://loja.ibge.gov.br/censo-demografico-2010-caracteristicas-gerais-da-populac-o-religi-o-e-pessoas-com-deficiencia.html>. Acesso em: março de 2013.

- BREDA, W. *Falibras-MT – Um Sistema para Autoria e Uso de Tradutores Automáticos Português-Libras baseados em Memória de Tradução*. Projeto de Final de Engenharia de Computação. CCEC/UFES, 8 de abril de 2005.
- BREW, C. Letting the Cat out of the Bag: Generation for Shake-and-Bake MT. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, páginas 610-616, Nantes, França, 1992.
- BRITO, P. et al. *FALIBRAS: Uma Ferramenta Flexível para Promover Acessibilidade de Pessoas Surdas*. International Conference on Computers in Education (TISE), 2012.
- BROWN, P. A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2):79-85, 1990.
- BROWN, P. et al. A Statistical Approach to French/English Translation. In: *Proceedings of the Second International Conference on Theoretical and Methodological Issues on Machine Translation of Natural Languages*, Carnegie Mellon University, Pittsburgh, PA, 12-14 de junho de 1988a. Painele 2: Paradigms for MT.
- BROWN, P. et al. A Statistical Approach to Language Translation. In: *Proceedings of Twelfth International Conference on Computational Linguistics*, páginas 71-76, Budapeste, Hungria, 1988b.
- BROWN, P. et al. Analysis, Statistica Transfer and Synthesis in Machine Translation. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 83-100, Montreal, Canadá, 1992.
- BROWN, R., FREDERKING, R. Applying Statistical English Language Modeling to Symbolic Machine Translation. In: *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 221-239, Leuven, Bélgica, 1995.
- CARBONELL, J. et al. The KANT Perspective: A Critique of Pure Transfer (and Pure Interlingua, Pure Statistics, ...). In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 225-235, Montreal, Canadá, 1992.
- CASTAÑO, M. et al. Machine Translation using Neural Networks and Finite-State Models. In: *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 160-167, Santa Fé, Novo México, 1997.
- CHANG, J., SU, K. A Corups-Based Statistics-Oriented Transfer and Generation Model for Machine Translation. In: *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 3-14, Kyoto, Japão, 1993.
- COELHO, H., COTTA, J. *Prolog by Examples*. Springer-Verlag, Berlin, 1988.
- COLLINS, B., CUNNINGHAM P. Adaptation Guided Retrieval: Approaching EBMT with Caution. In: *Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 119-126, Santa Fé, Novo México, 1997.
- CORADINE, L. et al. (2002a). Interfaces tradutoras do sistema falibras. Vitória/ES, Brasil. In: *IV Congresso Ibero-Americano Sobre Tecnologias de Apoio a Portadores de Deficiência – CIIEE*, 2002.
- CORADINE, L. et al. Levantamento de Hipóteses para Implementação do Projeto Falibras. In: *'Workshop de Informática na Escola' do XII Simpósio Brasileiro de Informática na Educação (SBIE 2001) - Pôster*. Vitória/ES, novembro de 2001.
- CORADINE, L. et al. Sistema Falibras: Interpretação Animada, em LIBRAS, de Palavras e Expressões em Português. In: *III Congresso Ibero-Americano de Informática na Educação*

Especial (CII- EE 2002) - Demonstração. Fortaleza/CE, de 20 a 23 de Agosto de 2002. Anais do III Congresso Ibero-Americano de Informática na Educação Especial (CIIEE 2002), ago. 2002.

DANLOS, L., SAMVELIAN, P. Translation of the Predicative Element of a Sentence: Category Switching, Aspect and Diathesis. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 21-34, Montreal, Canadá, 1992.

DOI, S., MURAKI, K. Translation Ambiguity Resolution Based on Text Corpora of Source and Target Languages. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, páginas 525-531, Nantes, France, 1992.

DORNA, M. *et al.* Sintatic and Semantic Transfer in F-Structures. In: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL '98)*, páginas 341-347, Montreal, Quebec, Canadá, 1998.

DORNA, M., EMELE, M. Semantic-Based Transfer. In: *Proceedings of the 16th International Conference on Computational Linguistics*, páginas 316-321, Copenhagen, Dinamarca, 1996.

DORR, B. Interlingual Machine Translation: a Parameterized Approach. *Artificial Intelligence*, 63(1&2):429-492, 1993.

DORR, B. *et al.* *A Survey of Current Paradigms in Machine Translation*. Language and Media Processing Laboratory, Institute for Advanced Computer Studies, College Park, dezembro de 1998.

DORR, B. Parametrization of the Interlingua in Machine Translation. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, páginas 624-630, Nantes, França, 1992a.

DORR, B. Principle-Base Parsing for Machine Translation. In: R.C. Berwick, S. P. Abney e C. Tenny, editores, *Principle-Based Parsing: Computation and Psycholinguistics*, páginas 153-184. Kluwer Academic Publishers, Norwell, MA, 1991.

DORR, B. Solving Thematic Divergences in Machine Translation. In: *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, páginas 127-134, University of Pittsburgh, Pittsburgh, PA, junho de 1990

DORR, B. The Use of Lexical Semantics in Interlingual Machine Translation. *Machine Translation*, 7(3):135-193, 1992b.

DORR, B., VOSS, C. Machine Translation of Spatial Expressions: Defining the Relation between an Interlingua and a Knowledge Representation System. In: *Proceedings of the Twelfth Conference of the American Association for Artificial Intelligence*, páginas 374-379, Washington, DC, 1993b

EBERLE, K. *et al.* Contextual Constrains for Machine Translation. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 213-224, Montreal, Canadá, 1992.

EMELE, M., DORNA, M. Ambiguity Preserving Machine Translation During Packed Representations. In: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98)*, páginas 365-371, 1998.

FARWELL, D. *et al.* The Automatic Creation of Lexical Entries for a Multilingual MT System. In: *Proceedings of the Fourteenth International Conference on Computational Linguistics*, páginas 532-538, Nantes, França, 1992.

FARWELL, D. *et al.* Using Machine Readable Dictionaries for the Creation of Lexicons. In: *Working Notes, Building Lexicons for Machine Translation*, AAAI-93 Spring Symposium, Technical Report SS-93-02, página 106, Stanford, CA, 1993.

- FARWELL, D., WILKS, Y. Ultra: A Multi-Lingual Machine Translator. In: *Proceedings of Machine Translation Summit III*, páginas 19-24, Washington, DC, julho de 1991.
- FARWELL, D., WILKS, Y. *Ultra: a Multi-lingual Machine Translator*. Technical Report MCCS-90-202, Computing Research Laboratory, New Mexico State University, 1990.
- FUJII, Y. *et al.* Analysis of Long Sentence in Japanese-English Machine Translation System. In: *Proceedings of Information Processing Society of Japan*, 1990.
- FUJITSU, documento de vendas. *ATLAS II: A Machine Translation System for Multilingual Translation Japanese to German*, 1986.
- FURUSE, O., IIDA, H. An Example-Based Method for Transfer-Driven Machine Translation. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 139-150, Montreal, Canadá, 1992.
- FURUSE, O., IIDA, H. Constituent Boundary Parsing for Example-Based Machine Translation. In: *Proceedings of the 15th International Congress on Computational Linguistics (COLING-94)*, 1994.
- GAZDAR, G., MELLISH, C. *Natural Language Processing in Prolog: An Introduction to Computational Linguistics*. Addison-Wesley. Reino Unido. 1989.
- GRISHMAN, R., KOSAKA, M. Combining Rationalist and Empiricist Approaches to Machine Translation. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 263-274, Montreal, Canadá, 1992.
- HIGINBOTHAM, D. Semantic Cooccurrence Networks. In: *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, LRC University of Texas at Austin, Austin, Texas, 1990.
- HUTCHINS, J. *Two precursors of machine translation: Artsrouni and Trojanskij*. Disponível em <http://www.hutchinsweb.me.uk/IJT-2004.pdf>. Acesso em: 15 de março de 2013.
- HUTCHINS, W. Prospects in Machine Translation: Proceedings of MT Summit I. In: *Proceedings of Machine Translation Summit*, Japão, 1987.
- JACKENDOFF, R. *Semantic Structures*. The MIT Press, Cambridge, MA, 1990.
- JACKENDOFF, R. *Semantics and Cognition*. The MIT Press, Cambridge, MA, 1983.
- JAIN, A. Parsing Complex Sentences with Structured Connectionist Networks. *Neural Computation*, 3:110-120, 1991.
- JOHNSON, R. *et al.* EUTROTA: A Multilingual System under Development. *Computational Linguistics*, 11:155-169, abril-setembro de 1985.
- JONES, D. Non-Hybrid Example-Based Machine Translation Architectures. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 163-172, Montreal, Canadá, 1992.
- JOSHI, A. How Much Context-Sensitivity is Necessary for Characterizing Structural Descriptions – Tree Adjoining Grammars. In: D. Dowty e L. Karttunen e A. Zwicky, editores, *Natural Language Processing: Theoretical, Computational and Psychological Perspectives*, Cambridge University Press, New York, 1985.
- KAPLAN, R. *et al.* Translation By Structural Correspondences. In: *Proceedings of Thirteenth International Conference on Computational Linguistics (COLING-90)*, páginas 272-281, Manchester, 1989.

- KAPLAN, R., BRESNAN, J. Lexical-Functional Grammar: A Formal System for Grammatical Representations. In: J. Bresnan, editor, *The Mental Representation of Grammatical Relations*, páginas 173-281. The MIT Press, Cambridge, MA, 1982.
- KAPLAN, R., WEDEKIND, J. Restriction and Correspondence-base Translation. In: *Proceedings of Sixth Conference of the European Chapter of the Association for Computational Linguistics*, páginas 193-202, Manchester, 1993.
- KING, G. Stochastic Methods of Mechanical Translation. *Mechanical Translation*, 1956.
- KNIGHT, K., LUK, S. Building a Large Knowledge Base for Machine Translation. In: *Proceedings of AAAI-94*, 1994.
- KOLODNER, J. *Case-Based Reasoning*. Morgan Kaufman, San Mateo, California, 1993.
- KOMATSU, E. *et al.* A Mono-lingual Corpus-based Machine Translation of the Interlingua Method. In: *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 24-46, Kyoto, Japão, 1993.
- LANDSBERGEN, J. *et al.* The Power of Compositional Translation. *Libertary and Linguistics Computing*, 4(3), 1989.
- LEHMANN, H., OTT, N. Translation Relations and the Combination of Analytical and Statistical Methods in Machine Translation. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 237-248, Montreal, Canadá, 1992.
- LIRA, G. *O Impacto da Tecnologia na Educação e Inclusão Social da Pessoa Portadora de Deficiência Auditiva: TLibras Tradutor Digital Português x Língua Brasileira de Sinais-Libras*. Boletim Técnico Senac, 2003. Disponível em: <http://www.senac.br/BTS/293/boltec293d.htm>. Acesso em: 21/02/2013.
- MARUYAMA, H. *et al.* The Mega-Word Tagged-Corpus Project. In: *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 15-23, Kyoto, Japão, 1993.
- MARUYAMA, H., WATANABE, H. Tree Cover Search Algorithm for Example-Based Machine Translation. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 173-184, Montreal, Canadá, 1992.
- McCORD, M. Design of LMT: A Prolog-Based Machine Translation System. *Computacional Linguistics*, 15:33-52, 1989.
- McLEAN, I. Example-Based Machine Translation Using Connectionist Matching. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 35-43, Montreal, Canadá, 1992.
- NAGAO, M. A Framework of a Mechanical Translation Between Japanese and English By Analogy Principle. In: A. Elithorn e R. Banerji, editor, *Artificial and Human Intelligence*. North Holland, 1984.
- NIRENBURG, S. A Direction of MT Development. In: *Proceedings of Machine Translation Summit IV*, páginas 189-194, Kobe, Japão, 1993.
- NIRENBURG, S. *et al.* *Machine Translation: A Knowledge-Based Approach*. Morgan Kaufmann Publishers, San Mateo, CA, 1992.
- NIRENBURG, S. *et al.* Multi-Engine, Adaptative MT. In: *Proceedings of Fifteenth International*

Conference on Computational Linguistics, 1994.

NIRENBURG, S. *et al.* Two Approaches to Matching in Example-Based Machine Translation. In: *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 47-57, Kyoto, Japão, 1993.

NIRENBURG, S., GOODMAN, K. Treatment of Meaning in MT Systems. In: *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, páginas 171-187, Linguistic Research Center, University of Texas, 1990.

NOMIYAMA, H. Machine Translation by Case Generalization. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, páginas 714-720, Nantes, França, 1992.

NYBERG, E., MITAMURA, T. An Efficient Interlingua Translation System for Multilingual Document Production. In: *Proceedings of Machine Translation Summit III*, páginas 55-61, Washington, DC, julho de 1991.

NYBERG, E., MITAMURA, T. The KANT system: Fast, accurate, high-quality translation in practical domains. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, Nantes, France, 1992.

OBJETIVA, Editora. *Dicionário Eletrônico Houaiss da Língua Portuguesa*, versão 1.0.5a, novembro de 2002.

OKUMURA, A. *et al.* A Pattern-Learning Based, Hybrid Model for the Syntactic Analysis of Structural Relationships among Japanese Clauses. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 45-54, Montreal, Canadá, 1992.

RIBEIRO, M. *Tradução Automática Utilizando Arquitetura de Transferência*. Trabalho de Conclusão de Curso de Graduação em Ciência da Computação, UFAL, 2007.

RICHARDON, S. *et al.* Combining Dictionary-Based and Example-Based Methods for Natural Language Analysis. In: *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 69-79, Kyoto, Japão, 1993.

RODRIGUES, V. *et al.* *Remodelagem da Arquitetura do Sistema Falibras*. Universidade Federal de Alagoas, Campus Arapiraca, [200-?]

ROSETTA, M. *Compositional Translation*. Kluwer Academic Publishers, Dordrecht, Holanda, 1994.

SADLER, L., ARNOLD, D. A Constraint-Based Approach to Translating Anaphoric Dependencies. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, páginas 728-734, Nantes, França, 1992.

SANFILIPPO, A. *et al.* Translation Equivalence and Lexicalization in the AC-QUILEX KLB. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 1-11, Montreal, Canadá, 1992.

SATO, S. Example-Based Translation of Technical Terms. In: *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 58-68, Kyoto, Japão, 1993.

SATO, S., NAGAO, M. Toward Memory-Based Translation. In: *Proceedings of Thirteenth International Conference on Computational Linguistics (COLING-90)*, páginas 247-252, Helsinki, Finlândia, 1990.

SCHANK, R. C. *Conceptual Dependency: A Theory of Natural Language Understanding*.

Cognitive Psychology, 3:552-631, 1972.

SCHANK, R. editor. *Conceptual Information Processing*. Elsevier Science Publishers, Amsterdam, The Netherlands, 1975.

SCHANK, R. Identification of Conceptualizations Underlying Natural Language. In: Roger Schank e Kenneth Colby, editores, *Computer Models of Thought and Language*, páginas 114-151. Freeman, San Francisco, CA, 1973.

SHIEBER, S., SCHABES, Y. Synchronous Tree Adjoining Grammars. In: *Proceedings of Thirteenth International Conference on Computational Linguistics*, páginas 253-258, Helsinki, Finlândia, 1990.

SOMERS, H. Interactive Multilingual Text Generation for a Monolingual User. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 151-161, Montreal, Canadá, 1992.

SU, K., CHANG J. Why Corups-Based Statistics-Oriented Machine Translation. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 249-262, Montreal, Canadá, 1992.

SUMITA, E. *et al.* Example-Based Approach in Machine Translation. In: *Proceedings of InfoJapan*, 1990.

SUMITA, H. *et al.* An Example-Based Disambiguation of Prepositional Phrase Attachment. In: *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, páginas 80-91, Kyoto, Japão, 1993.

THURMAIR, G. Complex Lexical Transfer in METAL. In: *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 91-107, Linguistics Research Center, The University of Texas, Austin, 1990.

TRUJILLO, A. Locations in the Machine Translation of Prepositional Phrases. In: *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 13-20, Montreal, Canadá, 1992.

TSUJII, J., FUJITA, K. Lexical Transfer Based on Bilingual Signs: Towards Interaction During Transfer. In: *Proceedings of the European Chapter of the Association for Computational Linguistics*, páginas 275-289, Berlin, Alemanha, 1991.

VAN NOORD, G. *et al.* The MiMo2 Research System. In: *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, páginas 213-233, Linguistics Research Center, The University of Texas, Austin, TX, 1990.

VAUQUOIS, B. A Survey of Formal Grammars and Algorithms for Recognition and Transformation in Machine Translation. In: *IFIP Congress-68*, páginas 254-260, 1968.

WHITELOCK, P. *A Lexicalist Unification Grammar of Japanese*. Tese de Doutorado, Idioma e Linguística, UMIST, Manchester, 1991.

WHITELOCK, P. Shake-and-Bake Translation. In: *Proceedings of Fourteenth International Conference on Computational Linguistics*, páginas 784-791, Nantes, França, 1992.

WHITELOCK, P. Why Transfer and Interlingua Approaches to TM are Both Wrong: A Position Paper. In: *MT Workshop, Into The 90's*, Manchester, England, 1989.

WILKS, Y. *et al.* Providing Machine Tractable Dictionary Tools. *Machine Translation*, 5(2):99-154, 1990.

YASUHARA, H. An Example-Based Multilingual MT System in a Conceptual Language. In: *Proceedings of Machine Translation Summit IV*, páginas 189-194, Kobe, Japão, 1993.