

**JÚLIO ALBERTO PINTO DE AGUIAR**

**DESENVOLVIMENTO DE ALGORITMO PARA APERFEIÇOAMENTO DO  
PROCESSO DE TRADUÇÃO AUTOMÁTICA  
(TRADUTOR UNIVERSAL)**

Monografia de Conclusão do CEAI - Curso de  
Especialização em Administração Industrial da Escola  
Politécnica da Universidade de São Paulo

São Paulo

2008

3º Quadrimestre

**JÚLIO ALBERTO PINTO DE AGUIAR**

**DESENVOLVIMENTO DE ALGORITMO PARA APERFEIÇOAMENTO DO  
PROCESSO DE TRADUÇÃO AUTOMÁTICA  
(TRADUTOR UNIVERSAL)**

Monografia de Conclusão do CEAI - Curso de  
Especialização em Administração Industrial da Escola  
Politécnica da Universidade de São Paulo

Área de Concentração: Comunicação multilíngüe e  
Tecnologia da Informação

Orientador: Prof. Dr. Antônio Cantizani Filho

São Paulo

2008

3º Quadrimestre

## **AGRADECIMENTOS**

Agradeço em primeira instância aos competentes mestres que com paciência tornaram a trilha até este ponto menos tortuosa e consideravelmente mais satisfatória, agradeço a minha família e a luz que obtive durante toda minha vida, emanada de minha querida mãe, Jupory Vaz Curado, que sempre soube conduzir meus passos em direção ao aprendizado e ao conhecimento.

## **RESUMO**

Este trabalho visa traçar uma linha de ação em busca de patamares inéditos da eficiência da tradução lingüística automática, utilizando-se de técnicas heurísticas e tomando como subsídio as entradas em massa do processo de Tradução Assistida, com o subsequente uso das funções obtidas em portal de comunicação entre grupos de usuários de diferentes línguas nativas.

Trata-se da quebra da última grande barreira de resistência na comunicação global que, não utilizará na verdade qualquer produto novo, mas une diversos deles na ordem adequada, criando um apelo de marketing irresistível sobre a queda da última barreira: a linguagem.

Portais de comunidades de relacionamento poderiam com a utilização do recurso apresentado estender sua abrangência, tornando possível a formação de grupos com participantes em qualquer local do mundo.

# SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E REFERÊNCIAS

1. INTRODUÇÃO .....	1
1.1. Apresentação do Problema .....	8
1.2. Objetivos do Trabalho .....	9
1.3. Relevância do Tema .....	11
1.4. Estrutura do Trabalho .....	13
2. FUNDAMENTAÇÃO TEÓRICA .....	14
2.1. Inteligência Artificial.....	14
2.2. Técnicas Heurísticas.....	19
2.3. Desenvolvimento Algorítmico.....	22
2.4. Tradução Assistida .....	25
2.5. Tradução Automática .....	27
2.6. Tecnologia Existente .....	30
2.7. Atividade Remota Remunerada .....	32
2.8. Processamento Distribuído .....	34
3. COLETA DE DADOS .....	36
4. DISCUSSÃO DOS RESULTADOS.....	39
5. CONCLUSÕES.....	40
6. ANEXO 1: Computing Machinery and Intelligence.....	42
7. ANEXO 2: Soneto 17 na íntegra.....	44
8. REFERÊNCIAS BIBLIOGRÁFICAS .....	45

## LISTA DE FIGURAS

Figura 1- Linguas mais utilizadas para acesso à Internet.....	2
Figura 2 - Diagrama heurístico básico .....	20
Figura 3 - Exemplo de algoritmo básico .....	23
Figura 4 - Capa do primeiro documento visando alcançar a Tradução Automática .....	27

## **LISTA DE TABELAS**

Tabela 1- Potencial de cada Linguagem para uso à Internet.....	3
Tabela 2 - Evolução da disseminação da Internet .....	4
Tabela 3 - Ferramentas de Tradução Assistida.....	30
Tabela 4 – Ferramentas de Tradução Automática existentes para uso via Web .....	31

## LISTA DE SIGLAS E REFERÊNCIAS

*Rogeriano*: método psicoterapêutico centrado no próprio paciente.

*Engenharia reversa*: a partir de uma solução pronta, retirar todos os possíveis conceitos novos ali empregados.

*Metalinguagem*: código que descreve uma outra linguagem ou código.

*Morre*: Gordon Moore, fundador da Intel.

*Engine*: motor, núcleo de sistemas computacionais responsável pela transformação de entradas em saídas.

Web Browser: programa de computador que permite acessar páginas Internet escritas em codificação padrão HTML



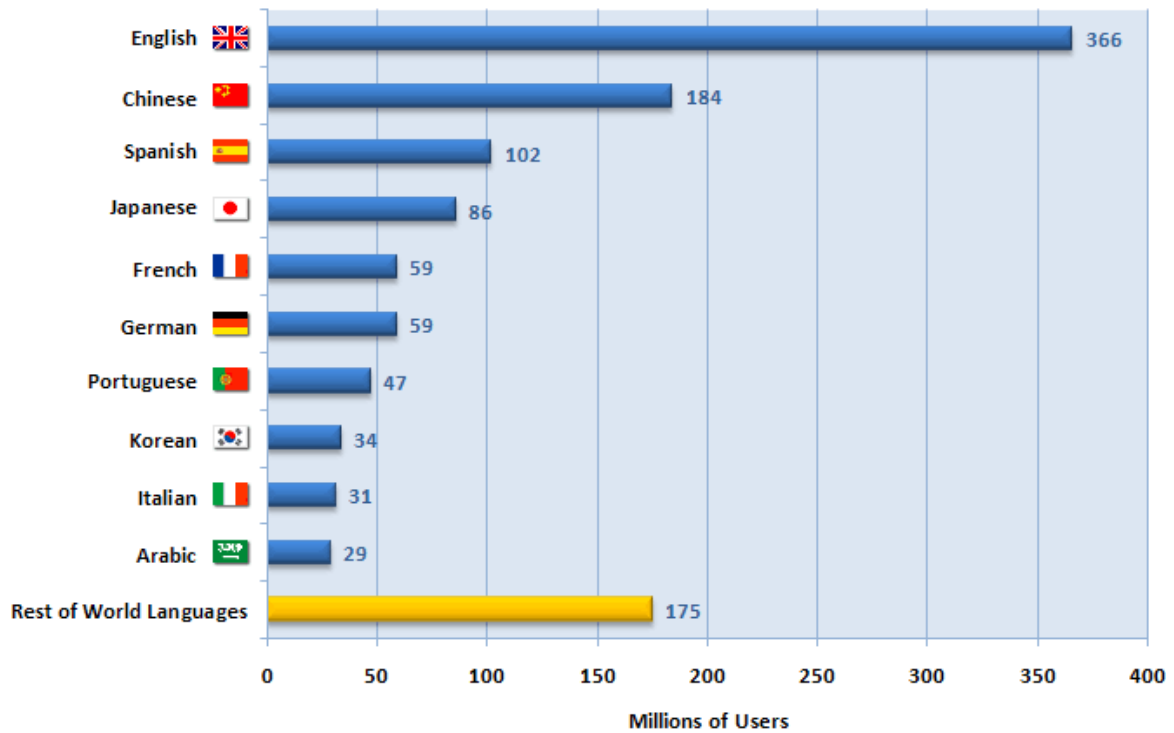
## 1. INTRODUÇÃO

A tecnologia tem sido a base para profundas revoluções na forma como nossa sociedade se organiza e com maior ascendência sobre a forma como nos comunicamos, por exemplo: o Telégrafo elétrico (~1850) possibilitou que a velocidade da propagação da informação pela primeira vez se separasse do meio físico necessário para movê-la, fazendo com que as estradas perdessem sua designação original de “meios de comunicação”; o Telefone (1876) permitiu pela primeira vez que houvesse comunicação instantânea entre pessoas em grandes distâncias ou o Rádio (1896) que durante décadas foi o meio de comunicação mais utilizado como propagação de informações e como entretenimento.

Na ponta desta linha de tempo, vemos a Internet, introduzida com a designação de Arpanet em 1969 com objetivos inicialmente militares, se desenvolveu nas últimas três décadas como ferramenta de profunda transformação em nossa sociedade, a partir dela o termo globalização ganhou sentido, fazendo com que qualquer informação esteja disponível em tempo real, para qualquer pessoa do planeta, informações passam a ter caráter absolutamente virtual, sem a necessidade de qualquer meio de propagação física da informação, elas são simplesmente armazenadas e disponibilizadas para o acesso em qualquer momento que se deseje. Uma única restrição se mantém: a linguagem.

O diagrama a seguir, produzido a partir da Internet World Stats, apresenta as dez línguas mais utilizadas atualmente para acesso à Internet, onde vemos que somados o inglês, o espanhol e o português chega-se a impressionante marca de mais de 500 milhões de pessoas:

## 10 Top Internet Languages



**Figura 1- Línguas mais utilizadas para acesso à Internet**

No quadro a seguir que demonstra a penetração de cada língua na Internet e a participação dos usuários nesta língua perante o total da estimada população da mesma, confirma-se com clareza o potencial de crescimento existente para cada uma delas, de mais de dois bilhões e meio de pessoas nas três línguas citadas (inglês, português e espanhol), apenas 513 milhões estão hoje na Internet:

Top Ten Languages Used in the Web ( Number of Internet Users by Language )					
TOP TEN LANGUAGES IN THE INTERNET	% of all Internet Users	Internet Users by Language	Internet Penetration by Language	Language Growth in Internet ( 2000 - 2007 )	2007 Estimated World Population for the Language
<a href="#">English</a>	31.2 %	365,893,996	17.9 %	157.7 %	2,042,963,129
<a href="#">Chinese</a>	15.7 %	184,001,513	13.6 %	469.6 %	1,351,737,925
<a href="#">Spanish</a>	8.7 %	101,539,204	22.9 %	311.4 %	442,525,601
<a href="#">Japanese</a>	7.4 %	86,300,000	67.1 %	83.3 %	128,646,345
<a href="#">French</a>	5.0 %	59,207,849	15.3 %	385.4 %	387,820,873
<a href="#">German</a>	5.0 %	58,981,592	61.1 %	112.9 %	96,488,326

<b>Portuguese</b>	4.0 %	47,326,760	20.2 %	524.7 %	234,099,347
<b>Korean</b>	2.9 %	34,120,000	45.6 %	79.2 %	74,811,368
<b>Italian</b>	2.7 %	31,481,928	52.9 %	138.5 %	59,546,696
<b>Arabic</b>	2.5 %	28,782,300	8.5 %	940.5 %	340,548,157
<b>TOP TEN LANGUAGES</b>	<b>85.0 %</b>	<b>997,635,142</b>	<b>19.3 %</b>	<b>203.7 %</b>	<b>5,159,187,766</b>
Rest of World Languages	15.0 %	175,474,783	12.4 %	440.3 %	1,415,478,651
<b>WORLD TOTAL</b>	<b>100.0 %</b>	<b>1,173,109,925</b>	<b>17.8 %</b>	<b>225.0 %</b>	<b>6,574,666,417</b>

**Tabela 1- Potencial de cada Linguagem para uso à Internet**

O crescimento do uso da comunicação eletrônica e dos recursos permitidos pela Internet, seja através do correio eletrônico, da navegação em páginas ou da participação em listas de discussão é indiscutível, conforme comprovamos através do digrama abaixo, reproduzida a partir de informações das Nações Unidas, claramente de pode ver a evolução da penetração da Internet nos mais diferentes países, com as mais diferente culturas:

País	1997	1998	1999	2000	2001	2002	2003	2004
Australia	1600000	4200000	5600000	6600000	7700000	10500000	11300000	13000000
Austria	760000	1230000	1840000	2700000	3150000	3340000	3730000	3900000
Belgium	500000	800000	1400000	3000000	3200000	3400000	4000000	4200000
Brazil	1310000	2500000	3500000	5000000	8000000	14300000	18000000	22000000
Canada	4500000	7500000	11000000	12971000	14000000	15200000	17600000	20000000
Chile	156875	250000	625000	2537308	3102200	3575000	4000000	4300000
China	400000	2100000	8900000	22500000	33700000	59100000	79500000	94000000
China, Hong Kong	675000	947000	1400000	1855200	2601300	2918800	3212800	3479700
Colombia	208000	433000	664000	878000	1154000	2000113	2732201	4050243
Czech Republic	300000	400000	700000	1000000	1500000	2600180	3100280	5100000
Ecuador	13000	15000	100000	180000	333000	537881	569727	624579
Egypt	60000	100000	200000	450000	600000	1900000	3000000	3900000
Finland	1000000	1311000	1667000	1927000	2235320	2529000	2560000	3286000
France	2485000	3704000	5370000	8460000	15653000	18716000	21900000	25000000
Germany	5500000	8100000	17100000	24800000	26000000	28000000	33000000	35200000
Greece	200000	350000	750000	1000000	915347	1485281	1718435	1955000
Hungary	200000	400000	600000	715000	1480000	1600000	2400000	2700000
Iceland	75000	100000	150000	168000	172000	186600	195000	225610
India	700000	1400000	2800000	5500000	7000000	16580000	18481044	35000000
Indonesia	384000	510000	900000	1900000	4200000	4500000	8080000	14508000
Iran	30000	65000	250000	625000	1005000	3168000	4800000	5500000
Ireland	150000	300000	410000	679000	895000	1102000	1260000	1198000
Israel	250000	600000	800000	1270000	1800000	2000000	2500000	3200000
Italy	1300000	2600000	8200000	13200000	15600000	19800000	22880000	28870000
Japan	11550000	16940000	27060000	38000000	48900000	59220000	61640000	64160000
Korea	1634000	3103000	10860000	19040000	24380000	26270000	29220000	31580000
Malaysia	500000	1500000	2800000	4977000	6346650	7842000	8643000	9879000
Mexico	595700	1222379	1822198	5058000	7410124	10764715	12218830	14036475
Mongolia	2600	3400	12000	30000	40000	50000	142800	200000
Morocco	6000	40000	50000	200000	400000	700000	1000000	3500000
Netherlands	2200000	3500000	6200000	7000000	7900000	8200000	8500000	10000000
Nicaragua	10000	15000	25000	50000	75000	90000	100000	125000
Niger	200	300	3000	4000	12000	15000	19000	24000
Nigeria	20000	30000	50000	80000	115000	420000	750000	1769661
Norway	900000	1000000	1100000	1200000	1319400	1398600	1583300	1792000

Oman	10000	20000	50000	90000	120000	180000	210000	245000
Pakistan	37800	61900	80000	300000	500000	1000000	1600000	2000000
Panama	15000	30000	45000	90000	168690	220000	260000	300000
Peru	100000	300000	500000	800000	2000000	2400000	2850000	3220000
Poland	800000	1581000	2100000	2800000	3800000	8880000	8970000	9000000
Portugal	500000	1000000	1500000	1680200	1860400	2267200	2674000	2951000
Puerto Rico	50000	100000	200000	400000	600000	677000	764000	862000
Romania	100000	500000	600000	800000	1000000	2200000	4000000	4500000
Russian Federation	700000	1200000	1500000	2900000	4300000	6000000	10000000	16000000
Rwanda	100	800	5000	5000	20000	25000	31000	38000
Saudi Arabia	10000	20000	100000	460000	1016208	1418880	1500000	1586000
Senegal	2500	7500	30000	40000	100000	105000	225000	482000
Singapore	500000	750000	950000	1300000	1700000	2100000	2300000	2421782
Slovakia	63000	144539	292359	507029	674039	862833	1375809	2276055
Slovenia	150000	200000	250000	300000	600000	750000	800000	950000
South Africa	700000	1266000	1820000	2400000	2890000	3100000	3325000	3566000
Spain	1110000	1733000	2830000	5486000	7388000	7856000	9789000	14332800
Sri Lanka	30000	55000	65000	121500	150000	200000	250000	280000
Swaziland	900	1000	5000	10000	14000	20000	27000	36000
Switzerland	548000	939000	1473000	2096000	2800000	3000000	3400000	3500000
Tunisia	4000	10000	150000	260000	410000	505500	630000	835000
Turkey	300000	450000	1500000	2500000	3500000	4300000	6000000	10220000
Uganda	2300	15000	25000	40000	60000	100000	125000	200000
Ukraine	100000	150000	200000	350000	600000	900000	2500000	3750000
United Arab Emirat.	90000	200000	458000	765000	976000	1175516	1373217	1384837
United Kingdom	4310000	8000000	12500000	15800000	19800000	25000000	34400000	37600000
United States	60000000	84587000	102000000	124000000	142823008	159000000	161632400	185000000
Uruguay	110000	230000	330000	350000	370000	380000	530000	680000
Uzbekistan	2500	5000	7500	120000	150000	275000	492000	880000
Venezuela	90000	322244	680000	820022	1152502	1274429	1934791	2312683
Zimbabwe	4000	10000	20000	50000	100000	500000	800000	820000

**Tabela 2 - Evolução da disseminação da Internet**

Dentro do contexto de globalização em que nossa época encontra-se inserida, a comunicação multilíngüe ganha importância crescente para diminuir as diferenças potenciais de acesso à informação, enquanto a quantidade total de informação disponível em meio eletrônico avança exponencialmente, não existem ainda ferramentas que permitam que usuários mono-lingüistas se comuniquem em línguas que não a sua original, é neste contexto que apresentamos esta monografia.

Muitos esforços existem na produção de ferramentas de tradução automática, que podem basicamente ser divididos em dois grandes grupos, aqueles de Tradução Automática pura e aqueles baseados em Tradução Assistida, com veremos adiante o objetivo final que procuramos é uma síntese que permita, a partir da produção de uma gigantesca base de dados, o aperfeiçoamento do processo de Tradução Automática.

Nosso objetivo base é a produção de uma Portal de Internet viabilizando listas de discussão (grupos de afinidade) temáticas, que permitam a comunicação de povos diferentes

em sua língua de origem, através de ferramenta automática de tradução e do monitoramento corretivo de operadores sobre as falhas da tradução automática. Inicialmente serão focadas as línguas: português e inglês, com o estágio seguinte focado no espanhol. Em outras palavras, usuários do nosso portal poderão escrever em seu idioma principal e o site se encarregará de traduzir o conteúdo da comunicação para outro idioma.

Trata-se de efervescente mercado, com mais de um bilhão de usuários em potencial, a própria gigante Microsoft acaba de investir no segmento, com a aquisição da francesa Systran, visando competir com o serviço Babelfish do Altavista, entretanto ambos esbarram na baixa qualidade do texto traduzido, sendo fundamental para a atividade proposta neste projeto, a ação de operadores.

Com esta iniciativa pioneira, estaremos derrubando a última grande barreira de resistência na comunicação global. Portais de comunidades de relacionamento como o Orkut tornaram-se extremamente populares em muito pouco tempo, com a utilização do recurso apresentado, sua abrangência se tornaria ainda maior, tornando possível a formação de grupos com participantes em qualquer local do mundo.

Através de listas de discussão temáticas, será possível ao usuário que domine apenas sua língua de origem, a virtual comunicação com pessoas do mundo inteiro, seja em uma atividade de lazer ou profissional, o acesso à informação global estará disponível, independente de restrições de comunicação. Com listas segmentadas a publicidade pode ser direcionada, permitindo que a cobrança usualmente utilizada por CPM (milhares de exibições), seja substituída pela cobrança por Page-views (cliques ou CPC), onde o anunciante paga exclusivamente pela quantidade de vezes que os usuários tiverem clicado sobre seus anúncios.

Segundo o Interactive Advertising Bureau norte-americano, em 2006 foram investidos 16,8 bilhões de dólares na internet, volume 34% maior que em 2005. Uma única empresa, o mega buscador Google, apresentou um faturamento no primeiro trimestre de 2007, de impressionantes 3,66 bilhões de dólares, com um crescimento de 63% em relação ao ano anterior.

Para avaliação do potencial da solução, um dos serviços mais populares para Grupos

na Internet é disponibilizado pelo Yahoo, em sua seção brasileira, sob a marca Yahoo Brasil, existem nada menos do que 500.224 grupos e subgrupos de discussão!

Os serviços serão automatizados através de ferramenta terceira de tradução automática, havendo a interação do operador para corrigir as imperfeições, uma vez que esta atividade não exige uma presença física em um mesmo ambiente dos operadores, pode-se implantar uma extensa rede de posicionamento, com operadores distribuídos em qualquer parte do país e do planeta.

O objetivo principal é oferecer um portal de comunicação (listas de discussão) que permita que sejam trocadas mensagens na linguagem original do usuário, alcançando-se através dele um algoritmo altamente eficaz de tradução automática.

Comercialmente, pode-se trabalhar em nichos específicos já vez que são listas segmentadas, assim a lista sobre automóveis pode receber anúncios do setor e a lista de esportes outros anunciantes, abre-se ainda a possibilidade de anunciantes diferentes para cada país de origem, ao menos sessenta países falam oficialmente inglês ou falam como segunda língua (508.000.000 milhões de pessoas), e sete falam português. Qual o alcance da solução quando ela atender também a comunidade hispânica ou o mandarim chinês? A escolha de novas linguagens será baseada na procura dos seus usuários e pelos incentivos oferecidos pelas representações dos países que terão seus povos beneficiados. O português será uma das duas línguas escolhidas no primeiro estágio exclusivamente por questões de logística operacional.

Listas Temáticas: Animais de Estimação, Cursos e Formação Técnica, Jogos Eletrônicos, Esportes, Artes, Relacionamento, Saúde e Beleza. Cada Grupo principal poderá ter subgrupos sem limites de quantidade, assim o grupo Artes, por exemplo, poderá se dividir em Cinema, Televisão, Música, Artesanato, Pintura, Teatro, dentre outros. Esta classificação será foco de estudo detalhado no momento de sua implantação.

A divulgação inicial do serviço poderá ser feita através, por exemplo, do *Google Adwords*, que permite que o anunciante registre uma combinação de palavras que obtenha uma maior probabilidade de despertar o interesse ao usuário. Isto é, uma empresa de

manutenção de automóveis, pode registrar as palavras “carro” e “peça”, fazendo com que apenas os usuários que combinaram as duas palavras recebam a publicidade, aumentando a chance do click gerado (e conseqüentemente pago pelo anunciante) se transformar em um negócio. Uma vez que temos listas temáticas, as publicidades já serão dirigidas a públicos específicos, ajustando-se desta forma ao perfil apresentado pela publicidade do *Google Adwords*.

Um grande diferencial do produto, que permitirá uma evolução na eficiência da tradução automática e uma diminuição da necessidade da interação humana será o registro das imperfeições e o desenvolvimento de uma estratégia heurística de aperfeiçoamento do resultado obtido, ou em outras palavras, a ferramenta virtualmente aprenderá com os erros corrigidos pelos operadores e poderá se desenvolver, assimilando e eliminando as falhas. Trata-se de iniciativa sem paralelo, todo o esforço atual para desenvolver algoritmos de tradução que obtenham uma eficiência ao menos tolerável, quando submetido ao usuário de destino mostraram ineficientes. O texto final traduzido é pobre e, apesar das tentativas de seu aprimoramento, por empresas com posição de destaque, quando o assunto é inovação e tecnologia (*Microsoft, Yahoo, Google*), ainda estamos distantes de um resultado final limpo.

Com a interação do operador, será viável ao software absorver os erros e adaptar-se de acordo com a necessidade, o trabalho que hoje se tenta aperfeiçoar de análise do texto a ser traduzido, será substituído simplesmente por um banco de dados. Em algum tempo teremos ferramentas automáticas de tradução com uma eficiência muito superior ao patamar atual, uma tecnologia de tamanha magnitude, que torna-se difícil projetar uma estimativa absoluta sobre seu valor comercial.

## **1.1. Apresentação do Problema**

Tradução Automática ainda pode ser considerada uma atividade arcaica, o resultado final ainda se distancia (e muito) de um trabalho inteligível e livre de recalques. Os esforços mundiais no sentido de desenvolver algoritmos de Tradução Automática mostraram-se insuficientes, mesmo com a imensa capacidade de processamento atualmente disponível. Vê-se como certo que será necessária por um longo espaço de tempo, a interação humana para a produção do resultado final.

Neste contexto, estimulou-se a criação de Bancos de Dados de Tradução, que servem de repositórios para as expressões usuais, permitindo que softwares pesquisem inicialmente o histórico de expressões já traduzidas e caso necessário, tentem a partir da utilização de algoritmos próprios identificar a tradução ideal.

Os Bancos de Dados criados atualmente, apesar de extensos e de altíssimo valor comercial, não conseguem contemplar um número suficiente de expressões que torne menos necessário o processamento algorítmico, tornando o resultado final de baixa qualidade, ou forçando a interação humana através da Tradução Assistida.



## 1.2. Objetivos do Trabalho

Apresentar fundamentação teórica que permita o desenvolvimento de um algoritmo de Tradução Automática com performance largamente superior aqueles atualmente existentes, discorrer sobre a conceituação teórica envolvida, inclusive sobre o ambiente no qual a solução está inserida e sobre os impactos que um produto desta natureza traria à sociedade. Não faz parte dos objetivos deste estudo uma análise financeira para o desenvolvimento e implantação do produto desenvolvido, se restringido à proposta para atingir a idéia inovadora quanto ao enfoque e resolução do problema.

O processo proposto para alcançar o objetivo principal de desenvolvimento do algoritmo é sintetizado a seguir:

- a. Criação de um portal de comunicação, inicialmente nas linguagens inglês e português, com listas de discussão variadas;
- b. As mensagens passarão por processo automatizado prévio (Tradução Automática) e imediatamente serão visualizadas pelos usuários das listas na sua própria linguagem, será necessário identificar parceiro comercial que tenha interesse na divulgação de sua marca como responsável pela execução do processo, como por exemplo a MicroPower, produtora do DeltaTranslator, dentre outras atuantes no mercado. Este parceiro contribuirá com o projeto, através do algoritmo original de tradução que se pretende aperfeiçoar;
- c. Análise superficial das mensagens por operadores, procurando eliminar grandes falhas de tradução, estimativa de uma média de 30 segundos por mensagem ou 120 mensagens/hora, esta operação é conhecida como Tradução Assistida e produzirá um Banco de Dados de altíssimo valor comercial, conhecido como *Translation Memory*;
- d. Pela natureza da atividade, os operadores poderão estar em qualquer ponto do Brasil ou do mundo, e serão remunerados conforme sua produtividade e qualidade de interpretação, inserindo o projeto no contexto de Atividade Remota Remunerada;
- e. A massificação do processo de alimentação do Banco de Dados decorrente do alto grau de uso de listas de discussão produzirá um incremento contínuo e

- substancial do grau de abrangência e eficiência do processo;
- f. Aperfeiçoamento da *engine*\* de Tradução Automática, a partir das inconsistências mais comuns e da otimização do algoritmo original, com base em técnicas heurísticas de adequação da Tradução Automática aos resultados obtidos pela Tradução Assistida;
  - g. Uso do conceito de Processamento Distribuído para análise dos resultados obtidos pelos dois conceitos de tradução (Automática e Assistida), permitindo um ajuste ao processo automatizado e seu contínuo desenvolvimento.

Com a aplicação das técnicas descritas, sobre as quais discorreremos em detalhes adiante, entendo que será aceitável entender como resultado final a obtenção de uma *engine* de Tradução Automática com performance inédita, trazendo conseqüências importantes, inclusive sobre a sociedade como um todo, já que a aplicação do produto obtido não será restrita a Listas de Discussão e sim poderá ser virtualmente aplicada a qualquer atividade que envolva comunicação escrita.

O potencial da solução é imenso, imagine um médico conversando com um colega na Austrália sobre a análise para determinado sintoma, um engenheiro conversando com alguém nos Estados Unidos sobre idéias para novos projetos ou um torcedor da seleção brasileira discutindo com um Inglês, o resultado do último confronto entre nossa seleção e a inglesa. Trata-se uma autêntica revolução na forma como a comunicação é feita atualmente, que não utiliza qualquer produto novo, mas une diversos deles na ordem adequada, criando um apelo de marketing irresistível sobre a queda da última barreira: a linguagem. Onde muitos vêem dificuldades, o autêntico empreendedor distingue a real oportunidade de sucesso, afinal “*a descoberta consiste em olhar para a mesma coisa como todo mundo e pensar algo diferente*” (Albert Szent Gyorgi).

### 1.3. Relevância do Tema

*“E o SENHOR disse: Eles constituem apenas um povo e falam uma única língua. Desta maneira, coisa nenhuma os impedirá, de futuro, de realizarem todos os seus projetos. Vamos, pois, descer e confundir de tal modo a linguagem deles que não consigam compreender-se uns aos outros. E o SENHOR dispersou-os dali por toda a superfície da Terra, e suspenderam a construção da cidade. Por isso, lhe foi dado o nome de Babel, visto ter sido lá que o SENHOR confundiu a linguagem de todos os habitantes da Terra, e foi também dali que o SENHOR os dispersou por toda a Terra”* Gênesis 11:6-9

A Bíblia, como diversos livros sagrados de diferentes religiões (Literatura Rabínica, Alcorão, Livros de Origem grega e dos povos Americanos, como Incas e Maias) apontam a criação de diferentes linguagens como um desastre de proporções catastróficas, essencialmente a comunicação entre os povos é historicamente vista como algo de interesse indiscutível.

Com a esperança de estabelecer um padrão internacional de comunicação e permitir que povos de qualquer língua possam se comunicar foram criadas as Linguagens Artificial, como a Novial em 1928, a Volapuque em 1880 e mais famosa e difundida entre elas, o Esperanto, criada pelo russo Ludwik Lejzer Zamenhof em 1887, entre as línguas artificiais criadas com o objetivo de simplificar conceitos lingüísticos e permitir a comunicação internacional, foi a única que saiu do terreno teórico e chegou a ser utilizado em determinados segmentos, embora com grandes restrições. Virtualmente é impossível eliminar a prática de povos durante anos e impor uma linguagem artificial de comunicação, por mais prático e produtivo que sua utilização possa se mostrar.

Estima-se que existam em todo o planeta mais de 5.000 línguas e dialetos diferentes, embora um percentual extremamente significativo possa ser concentrado nas principais línguas utilizadas internacionalmente, entre elas o Mandarim, o Inglês e Espanhol, permitir que povos de diferentes línguas possam se comunicar sem restrições, abre perspectivas avassaladoras, seja no tocante a disseminação de informações, como característica especificamente cultural, seja no enfoque de educação, seja com vista a interesses de caráter comercial.

Pode-se dizer com convicção que a ultrapassagem desta restrição possibilitará uma verdadeira revolução na forma como a sociedade hoje se apresenta, com uma diminuição extraordinária das diferenças potenciais entre indivíduos com línguas subalternas, não existirá mais limites para o acesso e uso da informação.

Um desafio que o século XX não conseguiu sequer se aproximar, mas que pode ser um feito histórico para identificar inclusive o século que se inicia. Este projeta quer contribuir nesta direção.

## 1.4. Estrutura do Trabalho

Levantamento inicial dos potenciais do produto, análise de alternativas para seu desenvolvimento e recursos necessários para alcançar um resultado final satisfatório.

Entre os conceitos essenciais para a produção do resultado procurado, listamos:

- Conceituação de Heurística;
- Tradução Assistida;
- Tradução Automática;
- Desenvolvimento Algorítmico;
- Tecnologias de acesso a Internet;
- Atividade Remota Remunerada;
- Processamento Distribuído.

Passa ainda a atividade a ser desenvolvida pelo processo fundamental de levantamento de dados e teste de performance e sobre o levantamento das ferramentas atualmente existentes, elaborando-se desta forma uma *Base Line* de trabalho e permitindo que o objetivo prático de elaboração de uma proposta seja ratificada pelo projeto.

Este será o caminho pelo qual trilharemos para seguirmos adiante na procura por resultado ímpar no desenvolvimento intelectual a nível mundial, a otimização e estabelecimento de estratégias para escrita de algoritmo que permita o aperfeiçoamento do processo de Tradução Automática em níveis inéditos.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. Inteligência Artificial

Em anos de avassaladora evolução tecnológica, onde barreiras antes consideradas como intransponíveis são superadas em cada piscar de olhos, nos deparamos com uma imensa lacuna no conhecimento humano, como conceituar inteligência? Consciência e noção de identidade seriam pré-requisitos essenciais para caracterizar inteligência ou podemos generalizar como fez o gênio Alan Turing em 1950, no célebre artigo *Computing Machinery and Intelligence?*

Turing, matemático britânico que comandou a quebra dos códigos secretos da comunicação nazista na Segunda Guerra Mundial, propôs no artigo citado, o seguinte jogo para avaliação se o interlocutor trata-se de unidade inteligente (independente de natural ou não): uma pessoa conversando com dois outros interlocutores em duas salas distintas, caso não consiga determinar com certeza qual deles é humano e qual é máquina, certamente podemos determinar que todos dois são inteligentes.

A previsão de Turing era que até o ano de 2000, máquinas conseguiriam ser aprovadas em seu teste, entretanto em competições anuais, como a mais famosa delas, disputada em Nova Iorque, EUA, e conhecida como prêmio Loebner, o desempenho das máquinas ainda é acanhado, sendo possível determinar com poucas perguntas, qual dos elementos não é humano. Independente das previsões originais de Turing não terem ainda se cumprido, é consenso entre estudiosos que tudo não passa de uma questão de tempo.

As restrições se impuseram sobre a eficiência do teste no decorrer das próximas décadas, e a elas Turing preparou um dossiê de contra-argumentos, entre todas apresentadas, uma pareceu sempre fazer grande sentido, a qual ele próprio nomeou de “*Argumento a partir da informalidade do comportamento*”:

“*não é possível produzir um conjunto de regras que possa descrever o que um homem deveria fazer em cada circunstância concebível*” (Turing, 1990,p. 58); e acrescenta: “*disto se conclui que nós não podemos ser máquinas*” (ibid.).

Sua contra-argumentação é de que máquinas também não podem ter seu comportamento previsto em detalhes, ou que ao menos não teríamos a tecnologia para mensurar as variáveis e planejadas, sem que isso interferisse no resultado final.

Antes de encontrar as potencialidades e horizontes abertos pela Inteligência Artificial, precisamos nos concentrar em sua conceituação, historicamente a distinguimos em dois nichos (Searle, 1990): o projeto “fraco”, onde o computador é visto como ferramenta que procura reproduzir comportamentos humanos, imitá-los e o projeto “forte” que visa desenvolvimento de uma real mente artificial. Vemos que a distinção entre ambos não é fácil, final qual seria a diferença entre imitar o pensamento e pensar?

Mesmo com a controvérsia sobre a conceituação de Inteligência Artificial e com a distância que estamos ainda de alcançar, vemos o conceito de comunicação lingüística fortemente ligado a ela e é neste ponto que fazemos a interligação com o estudo ora em questão.

O homem já produziu equipamentos com capacidade de armazenamento e velocidade de processamento muitas vezes superior à “mundana” capacidade humana, qual barreira nos separa então da distinção e compreensão do processo da fala? Podemos comandar um cérebro cibernético que vence um campeão mundial de Xadrez, mesmo havendo  $10^{120}$  combinações possíveis e não conseguimos compor 26 letras de forma a possibilitar aptidões de comunicação a equipamentos com o mesmo potencial?

Obviamente os desafios inerentes são absurdamente distintos, a comunicação (mesmo quando escrita) possui toda uma série de interpretações reflexas que limitam a performance de nossos melhores equipamentos eletrônicos atuais.

Estigmatizar as máquinas existentes com incapazes de se comunicarem é absolutamente perigoso, uma vez que uma das atividades mais frequentes de máquinas é justamente a comunicação, a troca de informações, seja com seus pares, seja mesmo com nós humanos, a distinção é que seu formato de comunicação em muito se distancia da fala inteligente, as máquinas usam seu próprio formato codificado de troca de dados, ou seja, de comunicação.

Indiscutivelmente um dos maiores problemas que temos para desvendar a comunicação humana, é a própria compreensão do mecanismo interno do cérebro, da consciência e dos processos mentais cognitivos, progressos constantes e significativos têm sido alcançados nesta direção, entre eles podemos destacar a criação de “atendentes virtuais”, que recebem internautas em páginas virtuais e tornam o acesso às informações muito mais atraente. Um dos robôs capazes de criar “atendentes virtuais” (ou *chatbots*) foi criado pela empresa brasileira Inbot ([www.inbot.com.br](http://www.inbot.com.br)) e já é responsável pela comunicação em páginas de grande empresas como a Petrobras, Tim e Unilever.

Abaixo apresentamos uma transcrição de um diálogo transcrito entre um usuário e o robô Ed, criado para uma página do Programa Nacional da Racionalização do Uso dos Derivados do Petróleo e do Gás Natural (CONPET) do Ministério de Minas e Energia:

**Ed:** Olá, eu sou Ed, o robô. Minha missão é ajudar na preservação de energia e de outros recursos naturais. Sobre o que você quer conversar?

**Usuário:** O que você faz?

**Ed:** O que faço? Leio muito, estudo bastante e converso mais ainda. Trabalho aqui no Conpet e estou sempre pronto para conversar sobre o que você quiser.

**Usuário:** Como posso economizar energia em casa?

**Ed:** Você pode economizar energia de várias maneiras, tomando banhos menos demorados, desligando a TV quando ninguém estiver assistindo, apagando as luzes quando as pessoas saírem dos ambientes, fechando a porta da geladeira após pegar alguma coisa nela, as lâmpadas incandescentes devem ser substituídas por lâmpadas fluorescentes ou compactas.

**Usuário:** Quais são as principais fontes alternativas de energia?

**Ed:** Energias alternativas, que têm baixo custo ambiental, como a energia solar, a energia eólica (dos ventos) e a produzida por óleos vegetais (biodiesel e biomassa), por exemplo.

**Usuário:** Mas, o que é a biomassa?

**Ed:** Biomassas são matérias de origem vegetal ou animal que podem ser aproveitadas como fonte de produção de calor ou eletricidade, tal como cana-de-açúcar, óleos vegetais, madeira, dejetos orgânicos ou resíduos das indústrias agrícola e alimentícia.



**Usuário:** Obrigado. Até logo, vou indo...

**Ed:** De nada, às ordens! Mas volte logo, pois adorei conversar com você!

Parece uma comunicação muita razoável, mas a real capacidade de interpretação logo aparece, vejamos mais uma tentativa:

**Ed:** Olá! Meu nome é Ed, estou aqui para ajudar na preservação de energia e de outros recursos naturais. Sobre o que você quer conversar?

**Usuário:** Protocolo de Kioto

**Ed:** O Protocolo de Kyoto é um tratado de combate ao aquecimento da Terra. Prevê reduzir gradativamente (de 2008 a 2012) as emissões de gases poluentes que destroem a camada de ozônio e causam o efeito estufa. O país que produzir emissões abaixo das normas tem o direito de vender cotas de emissão de certificados de Redução de Emissões de Carbono para aqueles que ultrapassem os limites estabelecidos. muitos países já ratificaram o tratado.

**Usuário:** O Brasil já ratificou o tratado?

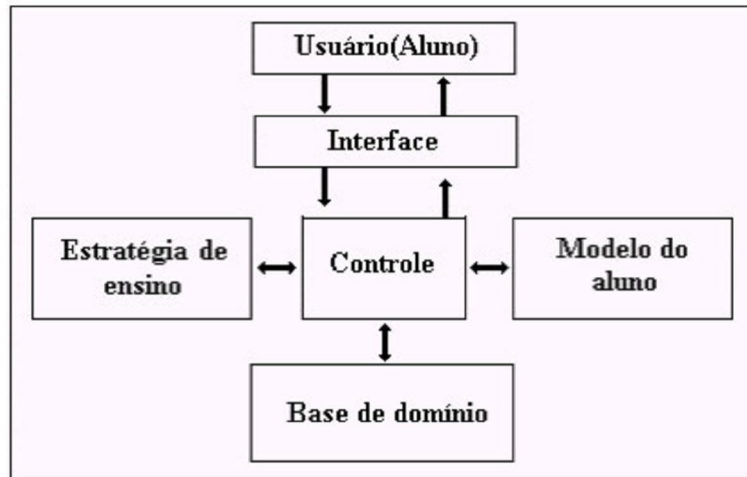
**Ed:** Brasil? Oito mil quilômetros de costa, uma média de 5 horas de sol por dia, ventos intensos e 12% de toda água doce do planeta: é o país com um dos maiores potenciais energéticos do mundo. Fontes de energia é o que não nos falta... (?)

Não foi muito difícil identificar as sérias restrições do programa, muitas perguntas simples recebem respostas em todo desconexas, entretanto outros programas como o Sete Zoom utilizado como atendente virtual da Close-up (<http://www.inbot.com.br/sete/>) e produzido pela mesma empresa é muito mais fácil de passar incólume, uma vez que simula um atendente adolescente e utilizado-se de uma comunicação informal, respostas as vezes dúbias não parecem tão absurdas.

A fundamentação teórica para este tipo de programa parte de 1966, com a criação do software Eliza, pelo alemão Joseph Weizenbaum, que tinha como objetivo a partir de poucas linhas de programação permitir a simulação de uma comunicação com um terapeuta rogeriano\*, seu algoritmo era tão simples que pode ser resumido em 100 linhas de código, entretanto os desdobramentos decorrentes da constatação do mundo, que efetivamente uma máquina poderia se apresentar como um humano e ser reconhecido como tal, teve impactos irreversíveis.

Uma das aplicações mais desejadas do desenvolvimento da Inteligência Artificial e de softwares como os discutidos anteriormente é a criação de Tutores Inteligentes e o

aperfeiçoamento da Instrução Auxiliada por Computador (CAI - *Computer Assisted Instruction*), com origem na década de 1960, já são largamente utilizadas ao redor do mundo e permitem que cada aluno receba um conteúdo didático próprio, direcionado às suas necessidade e particularidades.



**Figura XX** – Arquitetura Clássica de um Sistema Tutor Inteligente

Tecnologias com base nestes princípios norteiam sistemas inteligentes ao redor do mundo e podem em médio prazo alterar radicalmente a forma como se conduz processos educacionais, tornando-os extremamente personalizados e otimizados, e em última instância mais baratos para aplicação e disseminação.

O dilema filosófico decorrente do avanço da Inteligência Artificial possivelmente será um dos desafios deste século, a partir momento que efetivamente uma máquina conseguir sintetizar o pensamento humano sem que seja possível qualquer distinção, onde nos manteremos como essenciais? A próxima pergunta poderá ser para que servimos se os computadores podem até pensar por nós? São perguntas razoáveis e suas respostas hoje, soam exclusivamente como pura especulação.

Entretanto os inúmeros ganhos que a Inteligência Artificial certamente trará ao mundo não podem ser suprimidos por pré-conceitos ou idéias retrogradadas, as possibilidade, sejam em robótica, investigação científica pura, medicina, educação, dentre outros são inquestionáveis. Inclusive para este estudo ora em causa, de tradução automática entre linguagens.

## 2.2. Técnicas Heurísticas

Pode-se sintetizar heurística como a busca por soluções de problemas, que não conhecemos o caminho pré-definido para caminhar até elas. A análise desta técnica leva a determinação do método ideal para alcançar a solução. Para alguns, pode ser o inverso do Algoritmo, mas traremos neste estudo como um método para alcançá-lo, como veremos adiante. A origem do termo é grega, de “HEURISKEIN”, que significa descobrir.

Segundo o American National Standards Institute (ANSI), são “métodos ou algoritmos exploratórios para definição de problemas em que as soluções são descobertas pela avaliação do progresso obtido na busca de um resultado final”.

Modernamente, costuma-se segmentar a heurística em cada segmento em que aplica, seja investigação científica, matemática, biologia ou qualquer outro, entretanto sua popularização transcorreu a partir do lançamento em 1945 da obra “*How to solve it*”, do matemático húngaro George Pólya que sentenciava “*Se você não consegue resolver um problema, procure um outro semelhante que já tenha resolvido*”.

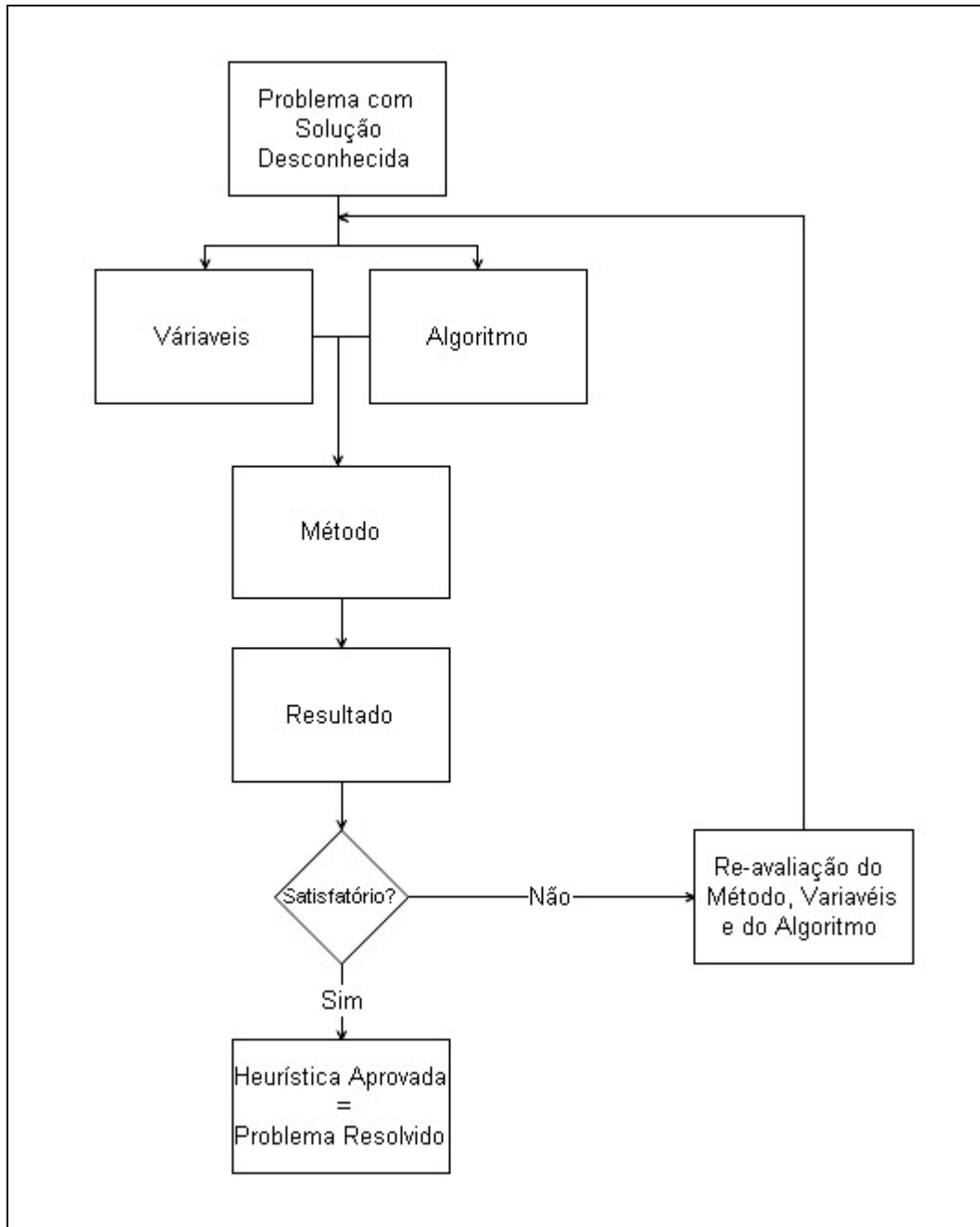
Seu livro tem quatro conceitos emblemáticos sobre Heurística:

- Caso não possa compreender um problema, monte um esquema que o ajude;
- Caso não encontre a solução, faça um mecanismo inverso (engenharia reversa\*);
- Caso o problema seja abstrato, tente propor um exemplo prático;
- Comece sempre pelo problema como um todo e divida-o até seccionar parte menores.

O sistema baseia-se em conceituação de prioridades e da influência de variáveis sobre o resultado obtido em cada etapa da resolução do problema, sua quantificação, o estabelecimento das variações obtidas e seu desvio e a retro-alimentação do ciclo, fazendo com que a cada etapa o método seja revisto e aprimorado para se aproximar do resultado desejado. No final do processo, o método escolhido como ideal pode ser distinto daquele sobre qual iniciou-se o estudo.

Tem-se o pensamento humano como um modelo classicamente heurístico, que através de experiências, muitas vezes sem conexões claras entre si, estabelece padrões e se re-

programa, ou seja nosso pensamento abstrato, imaginação e criatividade, convergem para produzir caminhos antes não imaginados.



**Figura 2 - Diagrama heurístico básico**

Em nosso problema de Tradução Automática, técnicas heurísticas serão fundamentais para que aplicadas à Tradução Assistida, possam nortear o re-processamento do ciclo e a calibração do algoritmo, até que o desvio da solução desejada, seja o menor quanto possível.

Iniciativas desta natureza vêm sendo tomadas em diversas partes mundo, embora com resultados ainda insatisfatórios, o que ratifica o potencial comercial da solução, seu retorno financeiro seria de monta absurdamente alto, um vez que o produto poderia transformar a forma como que a comunicação internacional hoje se mostra, trazendo alterações irreversíveis a sociedade como um todo.

### 2.3. Desenvolvimento Algorítmico

Antes de pensarmos em Desenvolvimento Algorítmico, precisar delinear em qual esfera pretendemos incluir o conceito de Algoritmo, uma vez que se trata de conceito intrinsecamente amplo. De forma simplista podemos apresentar algoritmo como sendo uma seqüência definida de instruções, cada uma das quais pode ser executada num período de tempo também definido e procurando uma saída em formato esperado ou como uma receita com entradas, processos que possam ser traduzidas em linguagem matemática prática e saídas.

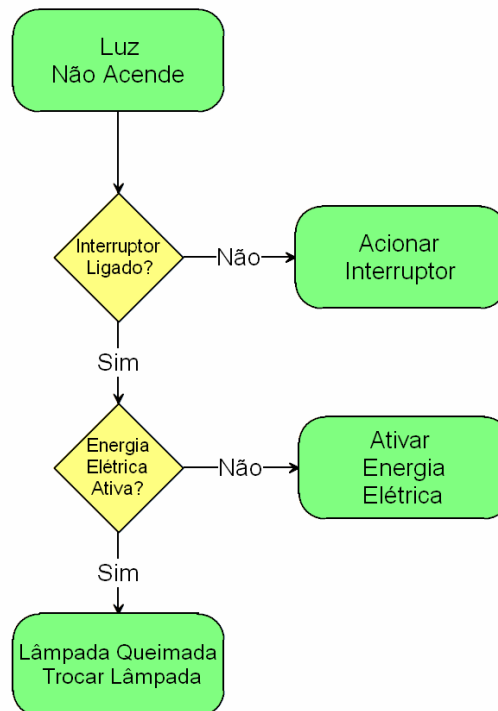
Academicamente, temos a clássica definição de Alan Turing, formulada em 1936, em seu artigo "*On Computable Numbers, with an Application to the Entscheidungsproblem*":

1. O algoritmo consiste de um conjunto finito de instruções simples e precisas, que são descritas com um número finito de símbolos;
2. O algoritmo sempre produz resultado em um número finito de passos;
3. O algoritmo pode, a princípio, ser executado por um ser humano com apenas papel e lápis;
4. A execução do algoritmo não requer inteligência do ser humano além do necessário para entender e executar as instruções.

O exemplo a seguir apresenta uma clássica demonstração de um algoritmo escrito em pseudocódigo, para o caso se trata de português estruturado, vê-se aqui claramente o conceito de entradas, processamento e saídas:

```
Algoritmo "SomaDeDoisValores"  
declare  
SOMA,A,B: inteiro  
inicio  
  Escreva("Digite dois numeros")  
  Leia(A,B)  
  SOMA <- A + B  
  escreva(SOMA)  
fim
```

O diagrama a seguir demonstra um outro clássico exemplo de conceito algorítmico:



**Figura 3 - Exemplo de algoritmo básico**

O desenvolvimento ou aperfeiçoamento de um algoritmo é a análise sistemática do resultado obtido, eliminando ou minimizando desvios do resultado final desejado, permitindo que a calibragem do processo efetuado pelo algoritmo evolua até o máximo resultado possível.

Tal como apresentamos, o conceito algorítmico é a base para a lógica computacional no desenvolvimento de *softwares* e sistemas computacionais complexos, não é possível imaginar a programação de computadores, que não com base na lógica apresentada pela conceituação algorítmica, encontramos aqui razão indiscutível pela qual tal desenvolvimento encontra exigência essencial para inserção no projeto ora apresentado de aperfeiçoamento do processo de Tradução Automática.

Essencialmente atualmente os algoritmos utilizados pelas *engines* de tradução são pobres e não encontram resultado satisfatório. Através da evolução do algoritmo utilizado, com a aplicação da técnica de processamento distribuído e com base em massiva Base de

Dados obtida através da Tradução Assistida, objetivamos viabilizar o desenvolvimento do algoritmo original a patamares inéditos.

Não se trata da aplicação de absolutamente nenhum conceito novo, todos eles são de domínio público e podem livremente ser estudados e aperfeiçoados, entretanto até esta data, todas as iniciativas trilharem pelo seu desenvolvimento individual, contrariamente ao apresentado neste estudo, que visa permitir a confluência de conceitos, que somados e utilizados adequadamente possuem um potencial muitas vezes maior.



## 2.4. Tradução Assistida

O desafio de desenvolvimento de um algoritmo de tradução automática que permite-se o intercâmbio entre povos de línguas distintas mostrou-se ao final do século XX como ainda distante de ser alcançado. O progresso em Inteligência Artificial e no desenvolvimento de algoritmos que atingissem um nível satisfatório de qualidade na tradução mostrou-se decepcionante.

A ciência da tradução lingüística, formada por milhares de profissionais e empresas de tradução ao redor do mundo, começou a prever que (ao menos por um longo tempo) a intervenção humana seria inevitável no processo de tradução, neste momento a atividade de tradução tomou um caminho específico na direção da Tradução Assistida ou Tradução auxiliada por computador, onde ferramentas de tradução elaboram um esboço inicial do texto final e a intervenção humana ajusta as imperfeições, tornando o resultado final mais digerível, tinha-se desistido da idéia do trabalho gerado integralmente a partir da simples Tradução Automática por computador.

Como auxiliar ao processo de Tradução Assistida, criaram-se as *Translation Memory* (TM) ou Memórias de Tradução, que consistem em bancos de dados que recebem traduções previamente elaboradas com o auxílio do operador que as armazenam para uso futuro, aumentando a qualidade e velocidade de processamento de traduções posteriores.

Grandes empresas criaram suas *Translation Memory*, como a IBM, através do seu *Translation Manager* e a Microsoft com o *Trados* e o *Translator's Workbench* associado ao *Microsoft Word*. Um problema crítico então se apresentou: compatibilidade.

Para que os bancos de dados gerados obtivessem valor comercial e pudessem ser utilizados por terceiros, cada um deles se utilizava de técnicas distintas de armazenamento e principalmente de acesso aos dados, procedimento imperativo para quem deseja sistematizar um processo de tradução. Visando minimizar o impacto deste problema, em 1988, foi criado o padrão TMX (*Translation Memory eXchange*), pelo OSCAR (Open Standards for Container/Content Allowing Re-use), a grupo especial da LISA (Localization Industry

Standards Association), um organismo que visa o estudo e criação de normatizações para utilização internacional.

O TMX utiliza o popular padrão de formatação de dados XML, permitindo desta forma uma universalização da informação e uma diminuição indiscutível dos custos para desenvolvimento ao seu acesso, os dados seguem padrão ISO e a codificação padrão Unicode.

Qualquer banco de dados de *Translation Memory* que não se utilize o padrão TMX, certamente terá sua potencialidade comercial razoavelmente reduzida, uma vez que os investimentos de terceiros para acessar e tratar os dados serão sempre mais dispendiosos. Criou-se um padrão de mercado de fato.

Internacional tem-se como certo no campo do Direito de guarda a propriedade intelectual, o legítimo direito sobre *Translation Memories*, uma vez que os dados gerados necessitaram forçosamente da intervenção de um operador para sua produção, entende-se que tanto um texto final gerado a partir de softwares que tomaram como base os Bancos de Dados, tanto quanto o próprio Banco de Dados em si, tem sim proteção de criação intelectual, devendo-se interpretar o utilizador do produto meramente como um usuário de uma licença de utilização temporária e não proprietário do software, como inclusive vemos um ferramentas de uso genérico, como editores de texto, planilhas de dados, dentre outros softwares. Em ultima análise tem-se a *Translation Memory* como uma criação intelectual e não apenas como um Banco de Dados, que como tal seria impossível de requerer propriedade.

## 2.5. Tradução Automática

Por mais impressionante que possa parecer (ao menos a mim, tive desta forma), as primeiras tentativas de desenvolvimento de um processo de tradução automática partem de meados do século XVII! Na Alemanha, o monge Johannes Becher elaborou um documento com fórmulas para aplicação de uma metalinguagem matemática que pretendia aplicar uma seqüência de números aos significados das palavras, em seguida outras seqüências para expressar as semânticas utilizadas e uma correspondência final ao latim. Um segundo documento aplicava equações aos resultados obtidos que levariam o texto final a qualquer linguagem desejada.



**Figura 4 - Capa do primeiro documento visando alcançar a Tradução Automática**

O interesse da humanidade em facilitar o processo de comunicação pode ser tão antigo quanto a própria comunicação em si e suas diferentes apresentações em diferentes partes do mundo.

Por volta de 1930, iniciaram-se esforços para a produção de máquinas de tradução ou como ficaram conhecidos: “cérebros eletrônicos”, o francês Georges Artsruni, construiu uma máquina leitora de bandas perfuradas que permitia o relacionamento com uma segunda banda perfurada, produzindo um resultado final de transcrição da língua de origem para a língua final.

Somente em 1940, com o advento da eletrônica e a criação de máquinas calculadoras, concebeu-se a idéia de que um circuito poderia, da mesma forma como lê números e os transforma em resultados, poderia ler letras e transcrevê-las em uma outra simbologia, o que nada mais é do que as diferentes linguagens. Estas idéias precursoras do processo de Tradução Automática restringiam-se a analogias simplórias entre origem e destino, sem qualquer análise semântica de contexto.

Durante as décadas seguintes os progressos foram tímidos, chegando o governo Norteamericano a suspender em 1966 totalmente os investimentos em Tradução Automática, como vemos abaixo pela transcrição da conclusão do Comitê de Assessoria no Processo da Automatização de Línguas:

*"Contudo, o comitê não justifica suficientemente um apoio de envergadura à tradução automática per se, considerando que este tipo de tradução é mais lenta, menos precisa e mais cara do que a realizada pelos tradutores humanos..."*

A expectativa sobre a possibilidade de sucesso da Tradução Automática estava em patamares nulos.

O fato é que o mundo ainda não tinha sido apresentado à famosa Lei de Moore\*: *“a cada 2 anos a capacidade de processamento dos computadores dobra, enquanto os custos permanecem constantes”*. Esta constatação permitiu que os avanços da Tradução Automática alcançassem níveis exponenciais desenvolvimento.

Atualmente o Sistema toma forma com base no processo já citado de *Translation Memories* ou Memórias de Tradução, Bancos de Dados onde o software responsável pela Tradução Automática faz a primeira pesquisa dos termos a serem traduzidas e somente caso não encontre uma seqüência completa já traduzida parte para a aplicação de algoritmos de tradução.

Muito se discute quando será possível a total Tradução Automática sem intervenção do operador humano, ou mesmo se isto será possível um dia, faz parte do objetivo deste projeto desenhar caminhos que possam nos dirigir nesta direção, havendo a prerrogativa que não se procura como conceitualmente apresentamos anteriormente de criarmos traduções que

terminantemente possam passar como produzidas por um operador, mas simplesmente produzir um resultado final que possa ser compreendido e que ao contrario do resultado final obtido atualmente por Traduções Automáticas, não gere um resultado virtualmente agressivo ao bom senso.

A produção sistemática e continua de um Banco de Dados de tradução permitirá a diminuição incessante da necessidade da intervenção humana, este processo será massificado pelo uso de sistemas de comunicação com abrangência ao mundo inteiro, possibilidade apenas atualmente disponível pelo advento da Internet e sua cobertura sem restrições.

## 2.6. Tecnologia Existente

Não existe no mercado mundial qualquer serviço semelhante, permitindo a utilização on-line de listas de discussão diretamente na língua original do usuário e transcrição na língua de destino, partimos então para detalhar os serviços que tem similaridade com o processo ou que servem de apoio para a realização de Tradução Assistida:

Para a atividade de Tradução Assistida, o software de suporte apropriado são as citadas *Translation Memories*, e as ferramentas que permitem seu acesso, conhecidas como CAT (*Computer Aided Translation*) onde destacamos os seguintes:






Produto	Logo	Empresa	Características
Catalyst		Alchemy Software Development	Ambiente visual integrado multi-plataforma
déjàvu		Atril	Permite que o próprio usuário expanda o Banco de Dados
Metatexis enterprise		Metatexis	Baixo Custo e integração ao Microsoft Word
Trados		SDL International	Suporte a todas as línguas correntes na Europa, inclui bases específicas de termos técnicos
AlignFactory		Terminotix, Inc.	Alto grau de compatibilidade com outros Bancos de dados
Heartsome Translation Studio		Heartsome Technologies Ltd	Utilização de bancos que podem ser utilizados livremente

**Tabela 3 - Ferramentas de Tradução Assistida**

Em Tradução Automática, podemos dividir as ferramentas em duas categorias distintas, aquelas direcionadas para tradução de palavras ou de pequenas sentenças e as que também pretendem traduzir textos longos.

Entre as primeiros pode-se listar o Babylon Pro, Lingoos, Power Translator.

Listamos a seguir softwares executados exclusivamente via *web browsers\**, diferentemente daqueles apresentados antes, se apresentam como prontos para receber textos longos, contudo a eficiência da tradução sofre grande decréscimo, sendo praticamente inviável a utilização com este fim:

Produto	Logo	Empresa	Línguas	Descrição
BabelFish		Altavista (Yahoo)	Inglês, Chinês, Alemão, Francês, Grego, Italiano, Japonês, Coreano, Português, Russo e Espanhol.	Aplicativo de tradução via Web, engine de tradução da norte-americana Systran
Windows Live Translator		Microsoft	Alemão, Espanhol, Francês, Italiano, Português, Chinês, e Japonês	Aplicativo de tradução via Web, engine de tradução proprietária
World Lingo		WorldLingo Translations LLC	Inglês, francês, espanhol, alemão, italiano, Português, holandês, grego, sueco, árabe, chinês (tradicional e Simplificado), japonês, coreano e Russo	Ferramenta proprietária que permite além da tradução via Web, o uso de APIs para a tradução de páginas inteiras e inclusive a interligação com Chats.
Google Translate		Google	Do Inglês para Árabe, Chinês (Simplificado), Chinês (Tradicional), Francês, Alemão, Italiano, Coreano, Japonês, Russo, Espanhol, Português, Do Chinês (Simplificado) para o Chinês (Tradicional), do Alemão para o Francês	Procura similaridades entre palavras com base em um gigantesco banco de dados e utilizando-se de ferramenta própria, seu grau de eficiência é muito baixo, já que minimiza a utilização de conceitos lingüísticos, privilegiando a similaridade com seu banco de dados
Amikai		Amikai, Inc.	Do Inglês para Japonês, Francês, Alemão, Espanhol, Italiano e Português	Permite a interligação com o Outlook e o Office

**Tabela 4 – Ferramentas de Tradução Automática existentes para uso via Web**

## 2.7. Atividade Remota Remunerada

O teletrabalho ou trabalho remoto já faz parte da realidade de diversos segmentos da sociedade, notadamente em vendas, consultoria e em tele-atendimento. Enquanto em países como Portugal e Chile, existem leis específicas para o tratamento desta modalidade de trabalho, no Brasil, ainda nos deparamos com a Consolidação da Leis do Trabalho (CLT), de 1943, que absolutamente não poderia conter nenhum tipo de direcionamento que vise tratar este tipo de atividade.

Somente em abril do corrente ano (2008), a Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática (CCT) do Senado Federal, emitiu parecer favorável sobre o Projeto de Lei da Câmara 102/07, que pretende reger a atividade remota remunerada, outros trâmites serão necessários para que possivelmente no próximo ano, o projeto possa ter sua aprovação final.

Independentemente das considerações quanto ao ajuste da CLT brasileira aos sessenta anos que se passaram desde sua criação, a atividade que é alvo deste projeto não se enquadra nos preceitos exigidos para caracterização do vínculo empregatício subordinado à CLT.

Entre estes preceitos estão incluídos: subordinação, horário, e frequência, todas não se aplicam ao caso em questão: remuneração mediante produtividade remota de mensagens traduzidas por meio assistido e pela qualidade da tradução.

Vemos então inteiramente a disposição todas possibilidades abertas ao projeto pela atividade remota. Teremos acesso a uma população gigantesca de possíveis candidatos à atividade, uma vez que tudo que a pessoa precisará é tempo, um microcomputador, acesso à Internet e conhecimento (mesma não profundo) de uma segunda língua.

Em um país com obstáculos imensos quanto à inserção social, distribuição de renda e eliminação de restrições de acesso ao emprego, seja quanto a raça, idade ou sexo, temos em um projeto como este, possibilidades inclusive de cunho sociológico direto, permitindo que abram-se possibilidades antes bloqueadas para toda uma seção de nossa população.



Quanto mais o operador se dedicar à atividade, maior remuneração ele poderá receber, o horizonte da população atingida que poderá usufruir desta atividade é tão extenso, que certamente os custos para remuneração terminarão se mostrando reduzidos.

Não existem restrições sequer de área geográfica limítrofe de países, operadores nas duas pontas, sejam da língua da origem, sejam da língua de destino poderão desempenhar o papel de auxiliares de Tradução, alimentando o Banco de Dados criado (*Translation Memories*), e disponibilizando seus equipamentos para agirem como auxiliares na execução do Processamento Distribuído de otimização do Algoritmo de interpretação.

Uma quantidade incalculável de pessoas, de certo estão ávidas por uma alternativa que se apresente desta forma: liberdade de horários para sua atividade, remuneração mediante produtividade e inexistência de local físico para sua realização.

## 2.8. Processamento Distribuído

Tecnicamente Processamento Distribuído é tido como o processamento em paralelo de informações em equipamentos absolutamente distintos, mas que convergirão a um mesmo resultado desejado, independentemente da plataforma local de processamento, será obtido um processo final consistente e homogêneo quanto à formatação dos dados.

Intrinsecamente existem restrições técnicas a construção de microcomputadores, exigindo investimentos exponenciais, a cada degrau significativa de aumento de velocidade ou de capacidade de processamento. O Processamento Distribuído traz soluções até simplistas para problemas desta natureza: será sempre muito mais barato ter 1.000 equipamento plataforma IBM-PC executando um processamento em paralelo, do que possuir um único equipamento com capacidade de processamento 1.000 vezes superior a cada unidade do sistema.

Com base neste conceito, criaram-se extensas redes mundiais de multi-processamento, onde usuários livremente permitem que seus computadores pessoais sejam utilizados em momentos em que estão ligados, mas ociosos, para executar um grande processamento central. Uma destas iniciativas é o Boinc (*Berkeley Open Infrastructure For Network Computing*), plataforma criada na Universidade de Berkeley para processamento oportunístico.

Exemplificando: o ciclo de processamento exigido de um microcomputador moderno durante a digitação de um texto, ou consulta a e-mails é um fração extremamente diminuta da capacidade total da máquina, nestas ocasiões o software do Boinc, identifica a ociosidade do processador e direciona o mesmo para colaborar com estudos que exigem um imensa capacidade de processamento.

Entre os processos que podem se beneficiar da lógica do Processamento Distribuído estão: desenvolvimento de modelos climáticos, pesquisas astronômicas, interações a nível molecular, estudos sobre a trajetória de partículas subatômicas, dentre outros.

O volume de processamento exigido por um sistema para análise de *Translation Memories* e elaboração de algoritmos que permitam sua utilização em variações semânticas futuras é característica própria de processa que pode ser fortemente auxiliado pelo Processamento Distribuído.

Uma vez que faz parte do processo de captação de Traduções Assistidas, a formação de uma extensa rede de colaboradores ao redor do mundo, necessariamente estaríamos adquirindo também a disponibilidade de um igualmente extensa rede de computadores a disposição do sistema central, visando obter um único resultado final: otimização do Algoritmo de Tradução Automática, permitindo que a importância da interação operacional seja continuamente minimizada.

### 3. COLETA DE DADOS

Como vimos, a Tradução por meio eletrônico pode atualmente ser dividida em três grandes grupos:

- Através de Operador Humano (conhecido como Tradutor);
- Por meio de Tradução Assistida, com auxílio humano;
- Tradução automática, inteiramente realizada pelo computador.

Nesta seção vamos analisar a performance de várias destas ferramentas e sua capacidade de adequação ao objetivo ora proposto.

Submeteremos às ferramentas um mesmo texto com características semânticas típicas da linguagem em questão, como também oferecemos uma sentença simples, a seguir apresentamos o texto submetido, com sua tradução apropriada:

Texto Original	Tradução
You must be patient, it wasn't my fault.	Você tem que ter paciência, não tive culpa disso (tradução de Ricardo Schütz)
Who will believe my verse in time to come. If it were filled with your most high deserts? (Soneto XVII, de William Shakespeare)	Quem crerá em meu verso na era futura. Se ele é cheio de tua mais alta verdade? (tradução de Fabrício Souza)

Coleta de dados mediante testes práticos:

Texto Original	Google Translate
You must be patient, it wasn't my fault.	Você deve estar doente, não foi culpa minha.
Who will believe my verse in time to come. If it were filled with your most high deserts?	Quem vai acreditar em meus versos tempo para vir. Se fosse preenchido com as suas mais altas desertos?

Texto Original	Babylon
You must be patient, it wasn't my fault.	É preciso ser paciente, não foi culpa minha.
Who will believe my verse in time to come. If it were filled with your most high deserts?	Quem vai acreditar minha poesia no tempo. Se fosse enchidas com o seu mais alto desertos?

Texto Original	BabelFish
You must be patient, it wasn't my fault.	Você deve ser paciente, ele wasn' t minha falha.
Who will believe my verse in time to come. If it were filled with your most high deserts?	Quem acreditará meu verso a tempo para vir. Se foi enchido com seus desertos mais elevados?

Texto Original	Microsoft Windows Live
You must be patient, it wasn't my fault.	Você deve ser paciente, não fosse culpa minha.
Who will believe my verse in time to come. If it were filled with your most high deserts?	Quem acreditarão meu verso em tempo para entrar. Se ele foram preenchido com seus desertos mais alto?

Texto Original	World Lingo
You must be patient, it wasn't my fault.	Você deve ser paciente, ele não era minha falha.
Who will believe my verse in time to come. If it were filled with your most high deserts?	Quem acreditará meu verso a tempo para vir. Se for enchido com seus desertos mais elevados?

Texto Original	Amikai
You must be patient, it wasn't my fault.	Você deve ser paciente, não era minha falta.
Who will believe my verse in time to come. If it were filled with your most high deserts?	Que acreditará meu verso futuramente. Se estivesse cheio com seus desertos mais altos?

Texto Original	Catalyst
You must be patient, it wasn't my fault.	Deve ter paciência, não foi culpa minha.
Who will believe my verse in time to come. If it were filled with your most high deserts?	Qual disposição acreditar meu poema em tempo por vir. Se aquilo era preenchido com a maioria alto desertos?

#### **4. DISCUSSÃO DOS RESULTADOS**

Conforme a expectativa produzida pelos registros apresentados durante este trabalho, os resultados obtidos através das ferramentas de tradução foram absolutamente distantes de qualquer interpretação inteligível do texto de origem.

Uma simples frase, sem qualquer dubiedade quanto ao seu sentido, obteve interpretações das mais esdrúxulas, quanto pior a interpretação obtida do estrofe do poema de Shakespeare.

Fossem os dados mais ricos em seu resultado, poderia se delinear análises minuciosas do seu conteúdo e se colocar em foco, a discussão dos dados obtidos, quanto a sua acuracidade frente ao objeto final desejado pela tradução. Entretanto em vista às discrepâncias colossais tal análise na verdade se mostra improdutiva, ou em termos gerais podemos declarar sem temor de engano, que as ferramentas disponíveis de Tradução Automática obtêm um resultado distante do desejado em um processo de tradução, exigindo ainda a intervenção de um operador para que o resultado final obtenha uma qualidade aceitável.

## 5. CONCLUSÕES

Nestas páginas apresentamos o conceito proposto de geração de Banco de Dados de Tradução ou *Translation Memories*, a partir da massificação proporcionada pela alimentação de traduções por operadores, acionados através de requisições provenientes de Listas de Discussão temáticas. A natureza desta operação permite a aplicação do conceito de Atividade Remota Remunerada, onde cada operador poderá virtualmente estar fisicamente instalado em qualquer localidade do planeta, proporciona ainda a utilização dos equipamentos individuais como unidades de processamento, a partir da aplicação do conceito de Processamento Distribuído.

Em conjunto estas atividades poderão tornar possível a otimização do algoritmo de Tradução Automática tomado como base para o início do trabalho, desenvolvendo-o e permitindo com que o resultado final obtenha uma qualidade muito mais próxima do que se procura, do que aquela obtida pelas ferramentas hoje existentes. Fato que evidenciamos pela coleta de dados empreendida frente aos mais conceituados *softwares* existentes do mercado para apoio a tradução.

Entendemos que boa parte desta explanação centra-se na discussão de conceitos e apresentação dos detalhes técnicos envolvidos, procurando antes de impor conclusões, permitir que sejamos conduzidos até elas. Não se trata de projeto vulgar, mas sem dúvida de projeto audacioso, procurando por respostas ainda não encontradas, por soluções ainda não descobertas.

A sociedade procura uma solução que facilite a comunicação entre povos de diferente linguagem há séculos, tentativas foram feitas e muitas ainda continuam em curso, todo um mercado de transcrições se formou em redor do mundo, exatamente porque o resultado final desejado ainda não foi alcançado. Nos últimos anos, uma evolução tecnológica sem precedentes nos foi apresentada, a civilização hoje está muito mais próxima, os quatros cantos do mundo estão ao alcance de um botão.

Este progresso justifica a expectativa positiva de que podemos estar muito mais próximos de resultados surpreendentes do que possivelmente se imagina. *Não há nada mais poderoso do que uma idéia cujo momento chegou* (Victor Hugo). Utilizar o potencial dos



recursos eletrônicos, unido-os a capacidade ainda não igualada (ou sequer compreendida) do ente humano pode ser a chave da fórmula de obtenção de uma Base de Dados rica e de um Algoritmo de tradução com performance inédita.

Chegamos desta forma a conclusão deste estudo, certos de que os conceitos apresentados, embora de natureza conjectural, podem uma vez que aplicados com adequada metodologia, atribuindo-se pesos e recursos apropriados, gerar o produto final desejado: Desenvolvimento de Algoritmo para Aperfeiçoamento do Processo de Tradução Automática ou de forma muito mais audaciosa, da obtenção conseqüente de um Tradutor Universal.

## 6. ANEXO 1: Computing Machinery and Intelligence

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-435.

### 1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game.' It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.

We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

### 2. Critique of the New Problem

As well as asking, "What is the answer to this new form of the question," one may ask, "Is this new question a worthy one to investigate?" This latter question we investigate without further ado, thereby cutting short an infinite regress.

The new problem has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man. No engineer or chemist claims to be able to produce a material which is indistinguishable from the human skin. It is possible that at some time this might be done, but even supposing this invention available we should feel there was little point in trying to make a "thinking machine" more human by dressing it up in such artificial flesh. The form in which we have set the problem reflects this fact in the condition which prevents the interrogator from seeing or touching the other competitors, or hearing - their voices. Some other advantages of the proposed criterion may be shown up by specimen questions and answers. Thus:

Q: Please write me a sonnet on the subject of the Forth Bridge.

A : Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

A: (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A: Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R-R8 mate.

The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include. We do not wish to penalise the machine for its inability to shine in beauty competitions, nor to penalise a man for losing in a race against an aeroplane. The conditions of our game make these disabilities irrelevant. The "witnesses" can brag, if they consider it advisable, as much as they please about their charms, strength or heroism, but the interrogator cannot demand practical demonstrations.

The game may perhaps be criticised on the ground that the odds are weighted too heavily against the machine. If the man were to try and pretend to be the machine he would clearly make a very poor showing. He would be given away at once by slowness and inaccuracy in arithmetic. May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.

It might be urged that when playing the "imitation game" the best strategy for the machine may possibly be something other than imitation of the behaviour of a man. This may be, but I think it is unlikely that there is any great effect of this kind. In any case there is no intention to investigate here the theory of the game, and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man.

## 7. ANEXO 2: Soneto 17 na íntegra

Texto original:

Who will believe my verse in time to come,  
If it were fill'd with your most high deserts?  
Though yet, heaven knows, it is but as a tomb  
Which hides your life and shows not half your parts.  
If I could write the beauty of your eyes  
And in fresh numbers number all your graces,  
The age to come would say 'This poet lies:  
Such heavenly touches ne'er touch'd earthly faces.'  
So should my papers yellow'd with their age  
Be scorn'd like old men of less truth than tongue,  
And your true rights be term'd a poet's rage  
And stretched metre of an antique song:  
But were some child of yours alive that time,  
You should live twice; in it and in my rhyme.

William Shakespeare

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ALIS Technologies e a Internet Society, As Principais Línguas do Mundo, disponível em <http://alis.isoc.org/langues/grandes.pt.htm>, acesso em 03/11/2007

BRASIL, Senado Federal, Atividade Legislativa, disponível em <http://www.senado.gov.br/sf/atividade/Comissoes/consComPerm.asp?com=1363>, acessado em 13/12/2008

COMISSÃO DE TRABALHO ON-LINE, Câmara dos Deputados, disponível em <http://ctasponline.blogspot.com/2008/04/teletrabalho-ser-incluido-na-clt.html>, acesso em 13/12/2008

CONPET, Programa Nacional da Racionalização do Uso dos Derivados do Petróleo e do Gás Natural, disponível em <http://www.ed.conpet.gov.br/converse.php>, acesso em 29/11/2008

COSTA, Macário, Sistemas Tutores Inteligentes, disponível em <http://www.nce.ufrj.br/ginape/publicacoes/trabalhos/MacarioMaterial/Sti.htm>, acesso em 29/11/2008

FREIGANG, K.-H. "Machine-aided translation". In Baker, M. (Ed.), *Routledge encyclopedia of translation studies*. Londres: Routledge. 1998.

GOOGLE Tradução, disponível em [http://www.google.pt/intl/pt-PT/help/faq\\_translation.html#statmt](http://www.google.pt/intl/pt-PT/help/faq_translation.html#statmt), em 27/11/2007

GOOGLE Adwords, disponível em <https://adwords.google.com/select/Login>, acesso em 10/12/2007

IDGnow, revista eletrônica da IDGBrasil, disponível em <http://idgnow.uol.com.br/mercado/2007/04/19/idgnoticia.2007-04-19.3354414267/>, acesso em 10/11/2007

INTERNET WORLD STATS, Internet World Users by Language, disponível em <http://www.internetworldstats.com/stats7.htm>, acesso em 04/11/2007

LOEBNER, Hugh Gene, Homepage, disponível em <http://www.loebner.net/Prizef/TuringArticle.html>, acesso em 27/11/2008

MAGALHÃES, Francisco José (1996). *A tradução profissional em Portugal*. Lisboa: Colibri.

MEDIA WIKI, Portal, disponível em [http://www.jorwiki.usp.br/L3not07/index.php/Eliza\\_pode\\_ser\\_humana%3F](http://www.jorwiki.usp.br/L3not07/index.php/Eliza_pode_ser_humana%3F), acesso em 29/11/2008

NAMEDIDA, por Rodrigo Polacco, disponível em <http://blogs.predicta.com.br/namedida/index.php/author/rodrigo-polacco/>, acesso em em 10/11/2007

PEREIRA, Aisa, A história da Internet, disponível em <http://www.aisa.com.br/historia.html>, acesso em 03/11/2007

PESSOA NETO, Anselmo, Conhecimento, experiência e literatura, Disponível em

[http://www.vermelho.org.br/museu/principios/anteriores.asp?edicao=76&cod\\_not=529](http://www.vermelho.org.br/museu/principios/anteriores.asp?edicao=76&cod_not=529), acesso em 03/11/2007

PÓLYA, George, *How to solve it*, Universidade de Princeton, EUA, 1945

SARTORI PORTO, Leonardo Uma investigação filosófica sobre a Inteligência Artificial, disponível em <http://www.seer.ufrgs.br/index.php/InfEducTeoriaPratica/article/viewFile/2304/1005>, acessado em 29/11/2008

SCHÜTZ, Ricardo, *Diferenças Idiomáticas*, disponível em <http://www.sk.com.br/sk-idiom.html>, acesso em 13/12/2008

SEARLE, John. *A redescoberta da mente*. São Paulo: Martins Fontes, 1997b.

SIMÕES, Alberto Manuel ; ALMEIDA, José João ; GOMEZ GUINOVART, Xavier - Memórias de tradução distribuídas. [S.l. : s.n.], [c. 2004]. ISBN 972-99166-0-8

SIQUEIRA, Rodrigo de Almeida, Homepage, disponível em <http://www.comciencia.br/reportagens/2005/10/12.shtml>, acesso em 27/11/2008

SNELL-HORNBY, Mary (1995), *Translation Studies: An integrated approach*, Filadélfia, EUA, Edição revista [1ª edição 1988].

SOUZA, Fabrício, *Sonetos de Shakespeare*, Biblioteca da Unicamp, disponível em <http://www.unicamp.br/iel/site/alunos/publicacoes/textos/s00001.htm>, acessado em 03/12/2008

TAVARES, P. de Campos; *Algoritmo*, Enciclopédia Verbo Luso-Brasileira da Cultura, Edição Século XXI, Volume II, Editorial Verbo, Braga, Janeiro de 1998 ISBN 972-22-1864-6

TURING, Alan. "Computing machinery and intelligence". In: BODEN, Margaret (ed.). *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, 1990. P. 40-66.

TURING, Alan. "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society, Series 2, Volume 42*, 1936, Hewlett, NY

UNITED NATIONS, Internet users, disponível em <http://mdgs.un.org/unsd/mdg/SeriesDetail.aspx?srid=608&crid=>, acesso em 04/11/2007