

**FUNDAÇÃO GETULIO VARGAS
ESCOLA BRASILEIRA DE ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS
CURSO DE MESTRADO EM GESTÃO EMPRESARIAL**

**ANÁLISE COGNITIVA DO ESTUDO
COMPARATIVO ENTRE ESTRATÉGIAS
HUMANAS DE TRADUÇÃO E MECANISMOS
DE TRADUÇÕES AUTOMÁTICOS**

**DISSERTAÇÃO APRESENTADA À ESCOLA BRASILEIRA DE ADMINISTRAÇÃO
PÚBLICA E DE EMPRESAS PARA OBTENÇÃO DO GRAU DE MESTRE**

JOSÉ ARINO MACHADO
Rio de Janeiro - 2010

JOSÉ ARINO MACHADO

**ANÁLISE COGNITIVA DO ESTUDO COMPARATIVO ENTRE
ESTRATÉGIAS HUMANAS DE TRADUÇÃO E MECANISMOS DE
TRADUÇÕES AUTOMÁTICOS**

**Dissertação de Mestrado apresentado à Escola Brasileira de
Administração Pública e de Empresas da Fundação Getúlio
Vargas, como requisito parcial para a obtenção do título de
Mestre em Gestão Empresarial.**

Orientador: Prof. Dr. Alexandre Linhares

Rio de Janeiro
2010

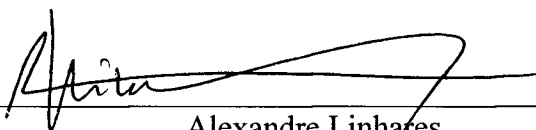
FUNDAÇÃO GETÚLIO VARGAS
ESCOLA BRASILEIRA DE ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS
CENTRO DE FORMAÇÃO ACADÊMICA E PESQUISA
CURSO DE MESTRADO PROFISSIONAL EXECUTIVO EM GESTÃO EMPRESARIAL

DISSERTAÇÃO DE MESTRADO


**EMPREENDEDORISMO E TECNOLOGIA: UM ESTUDO COMPARATIVO ENTRE AS
ESTRATÉGIAS HUMANAS DE TRADUÇÃO VERSUS MECANISMOS DE TRADUÇÃO
AUTOMÁTICOS.**

APRESENTADA POR: JOSÉ ARINO MACHADO

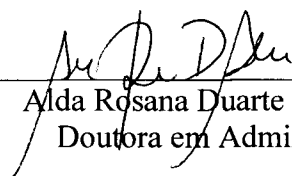
E
APROVADO EM:
PELA BANCA EXAMINADORA



Alexandre Linhares
Doutor em Computação Aplicada



Filipe João Bera de Azevedo Sobral
Doutor em Administração



Alda Rosana Duarte de Almeida
Doutora em Administração

AGRADECIMENTOS

Agradeço apenas ao Amor. Porque se mostra Pai quando faz nascer o Sol sobre todos sem distinção. Porque se mostra humilde quando se torna humano. Porque é luz que ilumina a todos mostrando-nos o caminho a seguir na ciência.

Agradeço ao Amor que se fez “minha família”. Muito Obrigado José Eduardo! Muito Obrigado Gabriel Lucas! Muito Obrigado Pedro Henrique! Vocês são a alegria do pai. A ti minha Amada esposa, Te Amo! Esse trabalho também é teu, tem tua marca, teu esforço, teu olhar, teu carinho.

Agradeço ao Amor que se fez família. Obrigado Pai! Minha inspiração enquanto administrador, meu autor preferido. Obrigado Mãe! Minha primeira professora nessa vida que sempre me exigiu persistência. Obrigados manos Júnior, Francelli (*in memoriam*), Tekinha, Sandalinha e Xilikinha pelo companheirismo, pelo amor. Vocês sempre foram, são e serão o cais que busco e me abrigo nas tormentas que por vezes tentam dominar meu coração. Muito Obrigado seu ManéZeca e Dona Enedina pela “adoção”! Muito Obrigado Valter e Taís porque os “adotei”!

Agradeço ao Amor que se fez amizade pelo incentivo e presença por vezes silente, mas sempre presença. Obrigado Marlon, Nilson, Pe. Sérgio (*in memoriam*), Paulo, Glória, Jucimar, Ângela, Luciano, Mara, João, Jéssica, Daniel, Keli, Duda, Clarisse e, de forma muito especial, a todos os que não citei aqui por falta de espaço (felizmente).

Agradeço ao Amor que se fez inteligência pela agudeza de espírito e busca constante do perfeito através da ciência. Muito Obrigado Hofstadter, Klein, Pinker, Habermas, Chomsky, Tenório, Horkheimer, Morente, Vergara e Linhares meu orientador, autores que me deram a base para meu desenvolvimento e me fizeram repensar a ciência.

Por fim, agradeço ao Amor que se fez apoio e organização através das pessoas que compõe a FGV, seus colaboradores, seu corpo docente de alto nível e em especial seu corpo discente da qual tive o privilégio de fazer parte. Muito Obrigado MEGE!

“Aquele que não ama não conhece a Deus, porque Deus é Amor”. I Jo 4, 8

RESUMO

Embora os progressos na área de informática sejam bastante significativos e velozes, na tradução automática há muito ainda o que ser feito. Desde meados dos anos 40 já havia um interesse, em especial pelos americanos e ingleses, numa tradução mais rápida e eficiente de documentos russos, porém até hoje o que se vê em termos de tradução automática está aquém daquilo que se possa chamar de uma boa tradução. Para buscar uma tradução automática eficiente os cientistas têm usado como fonte principal meios estatísticos de solução para tal problema. Esse trabalho visa dar um novo enfoque a tal questão, buscando na ciência cognitiva sua principal fonte de inspiração. O resultado a que se chega com o presente trabalho é que a estatística deve continuar sendo sim uma fonte de auxílio em especial na definição de padrões. Porém, o trabalho trás consigo o propósito de levantar a sobreposição semântica como via de possível solução que possa vir auxiliar, ou, até mesmo trazer maior rapidez a questão da tradução automática. No campo organizacional levanta uma questão interessante, o valor da experiência como meio inteligente de buscar melhores resultados para as empresas.

Palavras-Chave: Cognição. Inteligência organizacional. Inteligência artificial. Sobreposição semântica.

ABSTRACT

Although progress in computer science has been quite significant and rapid, as far as machine translation is concerned there is still much to be done. In the mid-40s, there was already an interest, particularly by the Americans and British, in a faster and more efficient translation of Russian documents. Yet, what we see today in terms of machine translation falls short of what one might call a good translation. To search for an efficient automatic translation, scientists have used, as a primary tool, statistical means to solve the problem. This paper aims to set out a new approach to the question, drawing on cognitive science as its major source of inspiration. The result achieved shows that statistics play an important role in the process, particularly for the definition of some patterns outlined by its means. This paper, however, is meant to inspire research that will draw on cognitive science as a valid means of improvement for greater efficiency in machine translation. With that it proposes semantic overlap as one of the possible ways for such improvements. On the organizational field, it highlights the importance and value of experience as an intelligent means of seeking better results for companies in the same semantic overlap that this experience brings.

Keywords: Artificial intelligence. Cognition. Organizational intelligence. Semantic overlap.

LISTA DE ILUSTRAÇÕES

Quadro 1 – Comparativo de Traduções. Período julho/janeiro..... 14

Quadro 2 – Traduções Possíveis 18

Figura 1 – Problemas de Bongard - Descrições potenciais múltiplas - 85 23

Figura 2 – Problemas de Bongard - Descrições potenciais múltiplas - 87 24

Figura 3 – Problemas de Bongard - Descrições potenciais múltiplas - 91 24

Figura 4 – Problemas de Bongard - Descrições potenciais múltiplas - 6 24

Figura 5 – Problemas de Bongard - Modo de Ver - 96 26

Figura 6 – Problemas de Bongard - Modo de Ver - 97 26

SUMÁRIO

1	INTRODUÇÃO	9
2	ANÁLISE E CLASSIFICAÇÃO DOS ERROS SEMÂNTICOS DE TRADUÇÃO AUTOMÁTICA	13
3	TRADUÇÃO HUMANA.....	16
3.1	A PESQUISA DE CAMPO	16
4	OBSERVAÇÕES TEÓRICAS À ANÁLISE COGNITIVA	20
5	A TESE - SOBREPOSIÇÃO SEMÂNTICA	29
5.1	SOBREPOSIÇÃO SEMÂNTICA – A INTELIGÊNCIA POR TRÁS DA TRADUÇÃO ..	29
6	CONCLUSÃO.....	34
	REFERÊNCIAS	35

1 INTRODUÇÃO

A importância do mundo dos negócios via internet é uma realidade que se sedimenta no dia a dia das organizações e também das pessoas. As relações via internet têm sido e serão cada vez mais de interesse comum a todos em qualquer parte do mundo. O mundo globalizado é fato inquestionável.

No campo empresarial, negócios são fechados a milhares de quilômetros de distância num simples *click*. São empresas internacionais que, ao invés de produzirem um determinado trabalho em três dias de serviços num mesmo local, em turnos de 8 horas diárias, fazem-no em 24 horas começando num ponto do planeta, dando a volta no globo em escritórios espalhados pelo mundo e terminando o mesmo serviço no mesmo escritório exatamente 24 horas depois de iniciado. Friedman (2007) cita como exemplos de planificação do mundo o caso da terceirização dos serviços de contabilidade dos americanos feito na Índia e o serviço de *call center* feito na China para os japoneses. Outro fator de relevância destacado neste mundo plano refere-se à mudança da concorrência que fora outrora regional e passa a ser uma concorrência global. O estudo feito por Whitaker et al. (2008) demonstra a necessidade de mudanças de posturas das empresas apontando para uma necessária atenção à tecnologia da informação e para a internet como fontes de melhoria das relações com o consumidor e também da redução de custos e integração da cadeia de suprimentos.

In response to the challenge from emerging-economy firms, firms in both developed and emerging economies have begun to reexamine their competitive strategies. While in the past firms could adopt a single strategy focused on either revenue growth or cost reduction to win and defend their market, 8 they must now simultaneously achieve both revenue growth and cost reduction to be successful in the global marketplace.⁹ Information technology and the Internet enable revenue growth through improved customer relationships and cost reduction through supply chain integration (WHITAKER et al., 2008, p. 3).

É a tecnologia da informação e a internet a serviço de um novo quadro de mudanças e de desafios.

Se considerarmos a importância destas relações via internet em termos financeiros chegamos a números bastante expressivos. Citando apenas os *advertisings*, Evans (2008, p. 1) assim destaca:

Online advertising began in 1994 when HotWire sold the first banner ads to several advertisers.¹ Revenue in the United States grew to an estimated \$7.1 billion in 2001 or about 3.1 percent of overall advertising spending. The dot-com bust destroyed or weakened many of the early online advertising industry players and reduced the demand for online advertising and related services.

The industry regained momentum by 2004 as the business model for “Web 2.0” came together.² A number of businesses such as Advertising.com, Google and Value Click emerged that facilitated the buying and selling of advertising space on web pages. Many web sites settled on the traditional “free-tv” model: generate traffic by giving away the content and sell that traffic to advertisers. Of the 20 top web sites, based on unique visitors, in the United States, 14 generate the preponderance of their revenues from the sale of advertising inventory—the eyeballs that view space allocated for promotions—to advertisers.³ In 2007, advertisers in the United States spent more than \$21 billion advertising on websites, about 7 percent of all U.S.

Para um negócio que começou em 1994 e que em 2001 já contabilizava um crescimento estimado em US\$ 7,1 bilhão, passando para mais de US\$ 21 bilhões em 2007 é sem dúvida uma expansão bastante expressiva.

Os serviços de buscas da google, por exemplo, possuem 71% das quotas de buscas nos Estados Unidos e 87% no Reino Unido, com 69% dos anúncios on-line, 24% das receitas de publicidade on-line. As páginas indexadas em 98 eram de 26 milhões, em meados de 2000, chegaram a mais de 1 bilhão e hoje podem estar próximas de um trilhão de endereços (JARVIS, 2009).

Com tamanho vulto de negócios acontecendo no mundo todo através da internet, nada mais lógico do que a necessidade de serviço de tradução para otimizar os negócios via internet.

Embora precisem de melhorias consideráveis, hoje os principais serviços de internet possuem também serviço de tradução instantânea de textos e web sites. As empresas Google, Alta Vista, AOL, Yahoo e MSN usam para tal serviço o mesmo software, a Systran possui uma abordagem de tradução estatística com o SMT - *Statistical Machine Translation*, conforme destacado pelo site da wikipedia.com¹:

O sistema de tradução estatística parte de um banco de dados de um novo par de línguas a partir do zero. Para compor esse banco de dados, a Google utiliza documentos das Nações Unidas.

¹ http://en.wikipedia.org/wiki/Google_Translate#cite_note-2

As formas de tradução automática são diversas. Uma das formas mais rudimentar é a tradução palavra por palavra, porém tais traduções mostram-se ao final, caóticas e terminam por definir textos desconexos, sem sentido. Por isso os pesquisadores buscam hoje outros meios de tradução para a construção de um software que seja capaz de melhorar tais percalços. A inclusão de modelos estatísticos aliados a análise sintática e semântica são apenas algumas das soluções que resultam em melhorias constantes, porém ainda não possuem um corpo de tradução satisfatório. Apenas para citar alguns, temos o *Lattice-based Minimum Error Rate Training*, um método estatístico que busca minimizar a probabilidade de sentenças de erros, Macherey et al. (2008); outro método traz a combinação de diversos sistemas de tradução de qualidade similar, Machery e Och (2007); ainda num outro modelo, o vínculo do idioma combinado com um sistema de alinhamento da palavra, Kumar, Macherey e Och (2007); há ainda estudos voltados para uma aplicação de modelos estatísticos a serem aplicados em larga escala na tradução automática numa busca de ampliar e tornar mais rápida a decodificação da tradução, Brants et al. (2007); o modelo *Minimum Bayes Risk* trabalha a estatística sob o risco mínimo, Tromble et al. (2008); e, por fim, o outro modelo é o de tradução de probabilística sincronizada em gramática de contexto livre ou PSCFG – *Probabilistic Synchronous Context-Free Grammar*. Dentro desse contexto livre, há a possibilidade de dois tipos de tradução, numa um modelo hierárquico e noutro a sintaxe aumentada.

Os modelos estatísticos apresentados são usados como fonte principal de tradução automática, porém, com o intuito de trabalhar alternativas não estatísticas de solução para o problema da tradução automática, o presente trabalho explora no campo da ciência cognitiva conceitos, visões que pretendem ser pontos de apoio de novas pesquisas em inteligência artificial e, principalmente, ser inspiração de novos rumos dentro da inteligência artificial, ou seja, o presente estudo se propõe a fazer uma análise cognitiva do problema, e não uma proposta de solução do mesmo. A metodologia deste estudo é moldada, portanto, em outras análises cognitivas, como Chalmers, French e Hofstadter (1992), French (2000a; 2000b), French e Labiouse (2001; 2002), Linhares (2000; 2005). A análise cognitiva se dará dentro de um comparativo entre as traduções humana e artificial bem como também aplicará comparativos com estudos já feitos no campo do xadrez, problemas de Bongard e resultados de neurociência. Para a concretização da análise cognitiva já no segundo capítulo, busca-se a classificação dos erros semânticos de tradução automática cometidos pelo sistema Systran. No terceiro capítulo, analisar-se-ão as traduções humanas feitas a partir da classificação

anteriormente definida, isto é, uma análise de erros semânticos na tradução humana. No quarto capítulo, levantar-se-ão as principais teorias que cercam o tema inteligência, enquanto que no quinto capítulo, a tese. E, por fim, a conclusão.

2 ANÁLISE E CLASSIFICAÇÃO DOS ERROS SEMÂNTICOS DE TRADUÇÃO AUTOMÁTICA

Modus Operandi – A análise das frases para posterior classificação dos erros foi feita a partir da tradução do texto *How to do what you Love* de Paul Graham. O tradutor automático usado é do sistema automático de traduções systran, o site usado foi o da Google. Foi considerada como núcleo de análise a frase. Considera-se aqui como frase todo e qualquer texto que teve seu início após alguma pontuação e que parou também ao encontrar uma pontuação. Com esse tipo de divisão, obteve-se do texto um total de 484 frases como núcleo de análise para a presente classificação. A classificação considerou tão somente os erros semânticos, não considerando, portanto, os erros de sintaxe.

A **Classificação** ficou definida em seis tipos distintos de erros que são os seguintes: (1) erro estético; (2) erro estético visível; (3) erro semântico visível; (4) erro semântico induzido, (5) erro de expressão idiomática visível e (6) erro de expressão idiomática induzido.

Dividiu-se ainda a classificação em dois grupos principais: grupos de erros estéticos e grupos de erros semânticos.

Grupo 1 - Erros Estéticos: nesse primeiro grupo, estão os casos de erros estéticos e os erros estéticos visíveis. Em ambos não ocorre a fuga de sentido, ou seja, o tradutor automático deixa clara e correta a tradução em termos semânticos, faltando porém, alguns detalhes que tornariam a tradução ainda mais correta ou exata.

No **erro estético**, como o próprio nome já define, fica claro que houve mudança de determinado vocábulo, porém foi mantida a mesma classificação gramatical. Com a correção do erro, a tradução fica ainda melhor, mais aceitável, mais bonita, porém trata-se apenas de uma questão estética.

No **erro estético visível**, falta ou sobra algum vocábulo, sem porém, perder o sentido correto da tradução.

Em ambos os casos não há perda de informação ou do sentido da frase, há, igualmente, uma compreensão completa e satisfatória do texto traduzido.

Grupo 2 - Erros Semânticos: no segundo grupo, os erros foram classificados em erros semânticos visíveis, induzidos e de expressão idiomática. Foram assim classificados por não traduzir corretamente o sentido da frase.

No **erro semântico visível**, fica evidente a falta de sentido. O leitor percebe claramente que a tradução não surtiu o efeito desejado, já que não consegue encontrar nexos ou coerência no conjunto das palavras traduzidas. A percepção no momento da leitura alerta o leitor quanto ao erro de tradução.

No **erro semântico induzido**, têm-se a impressão de que foi feita uma tradução correta, porém isso não ocorre. Pode ser considerado como o pior de todos os erros, já que passa a sensação de que a frase foi corretamente traduzida. Por isso foi definido como erro induzido, pois induz o leitor a certeza de uma tradução correta quando, na verdade, isso não ocorreu.

Por fim, tem-se uma terceira classificação que será chamada aqui de **erro de expressão idiomática** que também foi subdividida em **erro de expressão idiomática visível** e **erro de expressão idiomática induzido**. Tratam-se dos erros de tradução acontecidos por questões culturais, nos quais não há tradução do texto porque dizem respeito a expressões da cultura local. No português, poderíamos citar o “uai” mineiro, o “tchê” gaúcho ou ainda, expressões contemporâneas da internet como o “blz”, “rsrsrs” como exemplos de erro cultural máximo, enquanto que o “não tem” sul catarinense poderíamos citar como erro cultural induzido.

Um aspecto levantado pelo presente trabalho foi a tradução do mesmo texto em dois momentos distintos com um lapso de tempo de seis meses entre uma e outra tradução. A primeira tradução ocorreu em julho de 2009, enquanto a segunda em janeiro de 2010. O resultado foi uma melhoria bastante promissora, conforme o quadro abaixo:

Os números da 1ª Tradução							Os números da 2ª Tradução						
EE	EEV	ESV	ESI	TP	TNO	C	EE	EEV	ESV	ESI	TP	TNO	C
29%	26%	17%	20%	6%	1%	1%	43%	17%	10%	11%	3%	3%	13%
Grupo 1 (EE+EEV) □ 55%							Grupo 1 (EE+EEV) □ 60%						
Grupo 2 (ESV+ESI+EEIV+EEII) □ 37%							Grupo 2 (ESV+ESI+EEIV+EEII) □ 21%						

Quadro 1 – Comparativo de Traduções. Período julho/janeiro
Fonte: Adaptado de Wikipedia (2009).

EE: Erro Estético // EEV: Erros Estético Visível
ESV: Erro Semântico Visível // ESI: Erro Semântico induzido
EEIV: Erro de Expressão Idiomática Visível // EEII: Erro de Expressão Idiomática Induzido
TP: Traduções parciais // TNO: Traduções não ocorridas // C: Traduções corretas

Considerando-se apenas a compreensão semântica do texto traduzido, tem-se 56% de compreensão do texto no primeiro texto, passando-se posteriormente esse percentual para 73% de compreensão do texto traduzido, o que pode ser considerado um aumento significativo na eficiência de tradução em tempo relativamente breve.

3 TRADUÇÃO HUMANA

Há uma relação muito estreita entre a ciência computacional e a neurociência, na qual alguns pesquisadores se debruçam a entender o cérebro humano para empreender soluções inteligentes no campo computacional (HOFSTADTER, 2001; 2004; KANERVA, 1990; PINKER, 1997; 2002; CHANGEUX; CONNES, 1996).

Esta relação começa pelo fato de que a ciência computacional busca observar e copiar da neurociência, em especial no que diz respeito ao funcionamento do cérebro humano, características, funções, aspectos que possam trazer à máquina melhoria em excelência de seu funcionamento.

Com o mesmo propósito, este trabalho pretende conhecer melhor a atuação da inteligência humana. Para isso, definiu-se como frente de atuação o campo da tradução humana, com a intenção de conhecer o uso de suas estratégias de tradução para buscar meios distintos de solução para a tradução automática.

3.1 A PESQUISA DE CAMPO

A pesquisa de campo foi feita em corte longitudinal, com dois tradutores iniciantes a fim de analisar a tradução ao longo do tempo.

A Pesquisa – Com o propósito de compreender o aspecto inteligência, foram destacadas do texto de Paul Graham 164 frases para a tradução.

A pesquisa foi feita com dois adolescentes, iniciantes no idioma inglês e instruídos para que lessem os textos e os traduzissem usando o dicionário *on-line* da Google, eles foram alertados da importância de fazer a tradução “pensando alto”. Destacadas 164 frases para a tradução, toda a tarefa foi feita em vídeo para que fosse permitida uma análise mais pormenorizada posteriormente.

Observações – As seguintes observações puderam ser depreendidas da pesquisa:

Observação 1: a tradução de cada frase ocorreu de forma muito semelhante quanto à forma e método usados, valendo-se sempre da mesma sequência para empreender tal tarefa, que foi a seguinte:

- 1º - leitura da palavra no idioma estrangeiro;
- 2º - busca da tradução daquela palavra no dicionário;
- 3º - escrita do significado do vocábulo em inglês para o português;
- 4º - ao final da frase, verificação do sentido da frase;
- 5º - nos casos em que a frase não fazia sentido, faziam-se as modificações necessárias para torná-las coerentes.

Observação 2: a tradução feita por iniciantes assemelha-se muito a teoria P.O.S., de Simon e Gilmartin (1973), na qual a “tradução” das jogadas do xadrez se daria por uma sequência lógica a partir da análise quadro a quadro. Na tradução feita por iniciantes, essa estratégia acontece de forma semelhante, pois é feita palavra por palavra. A mesma estratégia de tradução automática foi tentada nos primórdios, porém, não raro, chegava-se a uma tradução caótica, sem sentido, da mesma forma como ocorreu com a tradução nessa pesquisa;

Observação 3: embora cada vocábulo possua uma série de significados, dando uma diversidade de opções de tradução, houve invariavelmente a escolha pela primeira palavra do dicionário, diminuindo-se consideravelmente a probabilidade de uma tradução mais eficiente;

Observação 4: sempre, ao final da tradução palavra por palavra, houve uma avaliação da tradução da frase com o propósito de verificar se ela fazia ou não sentido;

Observação 5: Uma Frase e Uma Imensa Diversidade de Traduções Possíveis – Como meio de analisar não apenas o sujeito da pesquisa, mas também o objeto pesquisado, buscou-se avaliar todas as possibilidades de tradução de uma frase simples como *“how to do what you Love”*. Essa frase possui apenas seis vocábulos e, usando-se o próprio tradutor automático da Google, obtiveram-se os seguintes significados para as palavras destacadas da frase em questão:

Classif.	Advérbio	Verbo no infinitivo		Adjetivo	Pronome pessoal	Substantivo	
T.O.	How	To	Do	What	you	Love	
T.L.P.	Advérbio	Preposição	Substantivo	Adjetivo	Pronome	Substantivo	
	Como	Para	Logro	o que	você	amor	
	de que maneira	em direção a	Mentira	a que	vocês	afeição	
	de que modo	a	Verbo	Que	tu	afeto	
	Quanto	a fim de	Fazer	Qual	te	paixão	
	Quão	a respeito de	Realizar	Quais	ti	ternura	
	por que	ao	Trabalhar	Advérbio	vós	dedicação	
	Substantivo	Até	Arranjar	de que maneira	senhor	cupido	
	modo de fazer	À	convir	em que	senhores	peessoa amada	
		Contra	cozer	Conjunção	senhoras	Verbo	
		De	cozinhar	Que	lhe	adorar	
		Do	estudar	tanto quanto	lhes	amar	
		Em	obsequiar	Pronome	o	querer	
		No	preparar	Quê	os	gostar de	
		Por	receber	o que	a	sentir afeto	
		Sobre	representar	aquilo que		sentir prazer	
		para si	visitar	Interjeição!			
			tratar	Como!			
			Verbo	quê!			
		Expressão					
		Fazer					
T.A.R.	Como fazer o que você ama?						
	Como fazer aquilo que você ama?						
	7	17	16	14	14	14	

Quadro 2 – Traduções Possíveis
Fonte: Adaptado de Wikipedia (2009).

- Legenda:
- T.O. □ Texto Original
 - T.G. □ Tradução Google
 - T.A.R. □ Tradução Anderson Ramos

Com o quadro acima exposto, pôde-se obter o número de todas as traduções possíveis com a multiplicação do número de possíveis traduções de cada vocábulo pelo vocábulo seguinte até a multiplicação do número de todos os vocábulos entre si.

Análise Combinatória = (N¹ x N² x N³ x N⁴ x N⁵ x N⁶)

AC = (7 x 17 x 16 x 14 x 14 x 14)

AC = 5.224.576

Esse é o número de possíveis combinações de tradução dos vocábulos apresentados nessa frase com seis palavras. Essa análise foi feita considerando-se o significado de cada palavra que compõe a frase, fazendo com que cada significado venha a se ligar a todos os outros significados de cada nova palavra. O objetivo é fazer uma simulação de tradução por tentativa de acerto ponto a ponto, testando cada possibilidade até encontrar a frase correta.

A tradução dessa frase “*How to do what you Love*” feita pelos dois iniciantes levou, em média, 1:30seg., o que poderia significar que levaram esse tempo para tomar mais de 5 milhões de decisões possíveis para traduzi-la. De acordo com Hofstadter, uma decisão tomada em nível de neurônio ficaria muito aquém disso “o tipo de decisão com que depara um neurônio – e isso pode ocorrer até mil vezes por segundo [...]” (HOFSTADTER, 2001, p. 370). Caso se considere que o texto foi lido e traduzido em 90 segundos, seriam disparadas 90.000 decisões nesse nível de decisão, um número muito aquém das 5 milhões de possíveis que precisariam ser tomadas caso se fizesse uma tradução pontual palavra por palavra. Duas observações podem ser depreendidas de tal fato: a primeira é de que não apenas um, mas uma rede de neurônios teve que fazer a realização e a segunda questão é que, ainda que aparentemente, a tradução tenha sido feita palavra por palavra, há meios pelos quais o cérebro usa para ordenar tamanho caos ou miríade de dados ao fazer a tradução, uma preocupação levantada por David Chalmers, Robert French e Douglas Hofstadter *apud* Hofstadter (1994, p. 169):

One of the deepest problems in cognitive science is that of understanding how people make sense of the vast amount of raw data constantly bombarding them from their environment. The essence of human perception lies in the ability of the mind to hew order from this chaos [...]

Essa habilidade se faz presente de forma bastante evidente naquele que é experto, já que faz a leitura dessa mesma frase na mesma velocidade com que a lê, ou seja, de dois a três segundos. Ainda que tenha uma gama de opções que equivaleria a mais de 5 milhões de decisões, ele o faz em menos de três segundos.

4 OBSERVAÇÕES TEÓRICAS À ANÁLISE COGNITIVA

1. A Gênese da Tradução – Para Vygotsky (1996; 2007), a verdadeira unidade da análise biológica é a célula viva, que possui todas as propriedades básicas do organismo vivo e não apenas os átomos que a compõe. Assim como a análise biológica parte de uma célula viva o significado da palavra é célula de apropriada investigação e não apenas uma letra. Dentro da gênese do conhecimento e da comunicação a comunicação real exige significado. Tolstoy afirma que as crianças experimentam amiúde certas dificuldades para aprenderem uma palavra nova, não pelo seu som, mas devido ao conceito a que a palavra se refere. Há quase sempre uma palavra disponível — quando o conceito se encontra maduro;

2. Traduzindo a Tradução – Uma tradução comumente conhecida nada mais é que a tradução de determinado idioma em outro. Porém, não obstante se esteja usando desse ambiente de pesquisa, a análise aqui não se restringe à tradução em si, mas à inteligência usada para tal intento, à percepção ou à estratégia usada. A tradução deve ser “lida”, portanto, dentro de um parâmetro conceitual como a transposição do objeto desconhecido para um objeto que passará a ser conhecido;

3. Uma Tradução de Memória Muito Antiga – Conforme se viu anteriormente, a tradução de uma palavra não se resume à tradução de um símbolo frio e calculista, mas, outrossim, à tradução de significados. Tais significados são construídos ou maturados por fatos e acontecimentos de uma vida. Como se não bastasse a multiplicidade de fatos e acontecimentos de uma vida, Changeux e Connes (1996) liga questões biológicas que ultrapassam uma vida e incluem gerações. Segundo o autor, o pensamento cristalizado de hoje num ser humano foi sedimentado ao longo de múltiplas gerações:

Existe, de um lado, uma memória genética. O organismo humano, tal como se apresenta hoje, resulta de múltiplas gerações de organismos que, antes, já viveram esse tipo de experiência. A resposta a um novo problema que se apresenta se inscreve na memória dos gens (CHANGEUX; CONNES, 1996, p. 187).

Isso dá indícios de que a maturação de um conceito pode estar alicerçada numa memória genética muito mais antiga e complexa do que se imagina. Ainda, tal sedimentação conceitual no tempo, por outro lado, não interrompe o conhecimento que sempre se atualiza com novos acontecimentos hoje. O cérebro se abre para a realidade exterior e, principalmente, pode servir-se da memória a longo prazo, que se depositou durante a experiência pós-natal, (CHANGEUX; CONNES, 1996).

4. A Não Linearidade do pensamento e a Inteligência Artificial – O pensamento linear ou cartesiano que fundamentou grande parte dos estudos em inteligência artificial deverá dar lugar à busca de uma maior flexibilidade dentro desse campo. A análise de uma tradução palavra por palavra com análise combinatória entre as possíveis traduções é uma possibilidade que precisa de revisão. Um aspecto da mente humana levantado por Klein (1998) indica uma série de fatores através de diversas experiências vividas por peritos em suas mais diversas formas de atuação que indicam haver mais do que simplesmente um raciocínio cartesiano nas tomadas de decisões. A simulação mental, pontos de influência, reconhecimento de padrão, visão do invisível, metáforas, analogia ou intuição são apenas alguns dos pontos levantados por Klein (1998) para compreender a tomada de decisão humana. O estudo de Klein foi baseado na experiência de peritos. Na apresentação do conceito de intuição, Klein (1998, p. 33-34) levanta aspectos como os que segue abaixo:

This is one basis for what we call intuition: recognizing things without knowing how we do the recognizing [...]. Rather than giving him specific facts from memory, the experience affected the way he saw the situation. Another reason that he could not describe his use of experience was that he was reacting to things that were not happening rather than to things that were. A third reason that he was unaware of his use of experience was that he was not drawing on his memory for any specific previous experience. A large set of similar incidents had all blended together.

Num outro exemplo, o autor destaca a tomada de decisão de um comandante com base apenas no reconhecimento de padrões que não aconteceu como era o esperado, ou seja, a falta de determinado reconhecimento de padrão alertou-o para a decisão correta. Um fator que pode ser considerado como fator de otimização chamado de ponto de influência

Leverage points are Just possibilities – pressure points that might lead to something useful, or might go nowhere. Expertise may be valuable in noticing these leverage points. Certainly in games like chess, the experts are more likely to see them. In interpreting situations, experts seem attuned to the leverage points – both the opportunities and the threats facing them – rather than just being aware of the physical and spatial arrangement of objects (KLEIN, 1998, p. 116).

Ainda dentro do aspecto não linear do pensamento e da busca de uma necessária flexibilização a inteligência artificial, Linhares e Brum (2007) demonstram em seus experimentos que os jogadores de xadrez são hábeis em perceber fortes similaridades entre diferentes posições quando as peças retêm o mesmo papel abstrato em ambas posições, indo ao encontro da idéia de que a analogia está no coração da cognição. O reconhecimento, por parte dos peritos, de posições com similaridades estratégicas não é reconhecida por principiantes. O experimento no campo do xadrez demonstra claramente que os peritos veem o que os iniciantes não conseguem ver. Há, segundo eles, papéis abstratos que os peritos identificam e que os iniciantes, fechados em questões irrelevantes, não conseguem enxergar (LINHARES; BRUM, 2009). Segundo os autores, o reconhecimento de padrões é mais importante do que um “olhar a frente” no jogo do xadrez. Klein (1998) identifica uma série fatores que os peritos veem e que os iniciantes não conseguem. Podem-se destacar os seguintes: padrões, anomalias, a visão global, a forma como tudo funciona, oportunidades e improvisações, acontecimentos, diferenças demasiado insignificantes para os principiantes e as próprias limitações.

Já a analogia é tida por Klein (1998) como fonte de orientação e ordenamento das idéias e pensamentos. De acordo com o autor, as metáforas e os casos análogos orientam o pensamento desenvolvendo a consciência da situação, identificando objetivos adequados e realçando informações relevantes, fornecendo uma estrutura que possibilite formar previsões quando existem muitos fatores desconhecidos. Ao ligarem conjuntos interativos de causas a resultados, são úteis para resolver problemas e criar expectativas. Hofstadter (1994) também ressalta a não linearidade como característica da inteligência humana e também como principal desafio à inteligência artificial e ressalta três características como foco de análise, que são a percepção de alto nível, a representação e a analogia. A percepção de alto nível segundo o autor é “a more global view of this information, extracting meaning from the raw material by accessing concepts, and making sense of situations at a conceptual level” (HOFSTADTER, 1994, p. 171). Uma das principais características da percepção de alto nível é a sua extrema flexibilidade

One of the most important properties of high-level perception is that it is extremely flexible. A given set of input data may be perceived in a number of different ways, depending on the context and the state of the perceiver [...] perception may be influenced by belief [...] goals [...] external context [...] perception of a situation can be radically reshaped where necessary (HOFSTADTER, 1994, p. 171-172).

A marca distintiva da percepção de alto nível é que ela é eminentemente semântica, ou seja, retira das situações seus significados intrínsecos.

5. Nos Problemas de Bongard O Fator Contexto – Os problemas de Bongard dizem muito das limitações da máquina em relação à compreensão humana, além de remeter ao que pode ser considerado, na tradução, um dos principais veículos de compreensão, o contexto no qual está inserido a frase ou a contextualização de um palavra dentro de uma frase. Tal aspecto pode mudar completamente o seu sentido, ora para um lado, ora para outro, conforme estiver no contexto. Linhares (2000) levanta a mesma questão da importância do contexto em que uma interpretação não deve de ser fechada ou acabada, mas, encontrada a partir de determinados métodos que consideram o todo, o conjunto para uma interpretação correta da unidade. Um dos destaques levantados pelo autor:

Proposition 3. The very same geometrical arrangement may need fundamentally different representations for each context on which it arises. Bongard provided a great number of examples where the same geometrical arrangement should be described in fundamentally distinct ways. A case in point: what is to be seen in BP#06 as a single object (i.e. a triangle) later appears in BP#91 (and in BP#85) as three distinct objects (i.e., three line segments that just happen to meet by coincidence). Another example: in BP#85, straight line segments must be seen as continuous, even when they intersect with another line segment. This does not happen, for instance, in BP#87, where the intersections happen to split the line segments. There is a persistent problem on the definition of what is to be regarded as a bounded entity in a problem, of where those boundaries reside, and of what should be the right level of description of the objects (LINHARES, 2000, p. 11).

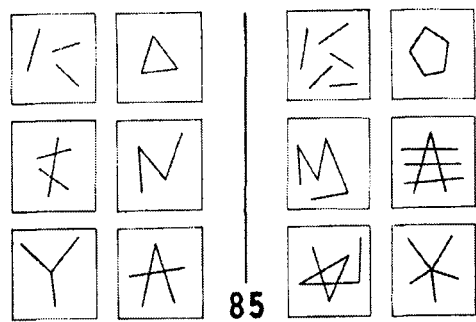


Figura 1 – Problemas de Bongard - Descrições potenciais múltiplas - 85
Fonte: Linhares (2000, p. 11).

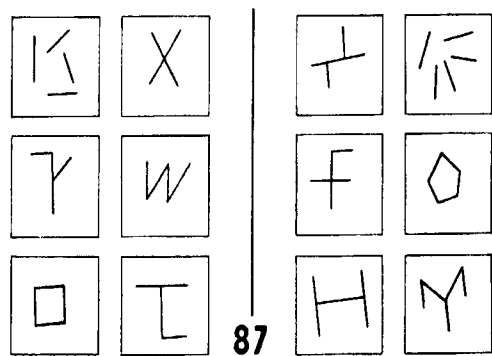


Figura 2 – Problemas de Bongard - Descrições potenciais múltiplas - 87
Fonte: Linhares (2000, p. 11).

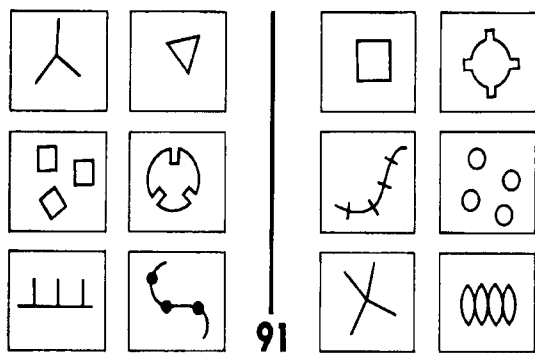


Figura 3 – Problemas de Bongard - Descrições potenciais múltiplas - 91
Fonte: Linhares (2000, p. 11).

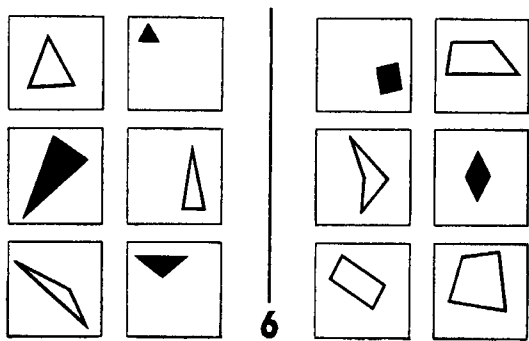


Figura 4 – Problemas de Bongard - Descrições potenciais múltiplas - 6
Fonte: Linhares (2000, p. 11).

Figure 4. Bongard problems require multiple potential descriptions – even for arrangements as simple as line segments and triangles.

As one examines these problems, it becomes apparent that a perception module could not, and this is in principle, create a description of their content without examining the overall context of the problem (i.e., looking at the other boxes), without interfering with the task of finding out what the solution might be like. It is impossible to separate the process of representation of one box from the consideration of the whole group (LINHARES, 2000, p. 11).

No campo da tradução, o mesmo fenômeno se repete e a necessidade de conhecimento de um contexto amplo para uma tradução coerente se faz indispensável.

6. O Fator Cultural - Como se “traduziria” para o inglês a seguinte frase tirada de um diálogo entre duas pessoas ao telefone?

Ator “A” _ [...] e então você vai me vender essa data?

Ator “B” _ Claro que sim, basta acertarmos os valores [...]

Antes mesmo de se ater à tradução português → inglês, deve-se observar que, caso se pergunte a um paranaense qual o significado de tal frase, ele não terá nenhuma dificuldade de reconhecer no diálogo uma negociação de um terreno. Porém, se a mesma pergunta for feita a um gaúcho, ele ficará se perguntando se acaso não se trata da negociação de uma data de trabalho para uma possível substituição de um colega de serviço por outro. Somente na continuação da conversa é que seria possível identificar a diferença entre uma tradução e outra. Bongard destaca a questão cultural nos jogos de figuras de “baleias”, “minas” e “submarinos”. Ainda, Hofstadter (2001, p. 414), ao citar na obra de Dostoiévski a história de uma tradução em que o autor nomeia uma rua em sua descrição, levanta a importância do nome da rua, que aspectos dessa rua são importantes? Ela tem um significado na história? Ou ela é importante por algum fator cultural, enfim, um contexto que muda o âmago, o valor e o significado da palavra.

7. A Analogia - Outra questão refere-se ao modo como se vê o mundo, ou a analogia com que costumamos usar para compreender determinados aspectos que desconhecemos. Nos problemas de Bongard, por exemplo, o padrão determinado pelas figuras mostra um aspecto que uma máquina não teria condições de “ver”. No dizer de Hofstadter: “we must see a thing *as another thing*”, ou seja, se é capaz de ver o que está por detrás da informação objetiva. Nas figuras 96 e 97 são vistas as formas geométricas não pelo que está exposto, mas pelo que falta, pelo estranho e contraditório “vazio ocupado” dando um significado pelo reconhecimento de um determinado padrão. Linhares (2000, p. 14) assim destaca:

It is important to understand precisely what is involved in the process of “seeing as”. In fact, this may be one of the most important problems in high-level perception, as it shows that the process of perception includes a kind of “analogy-making” – that representations can “slip” (this thesis has been fully developed elsewhere [12]).

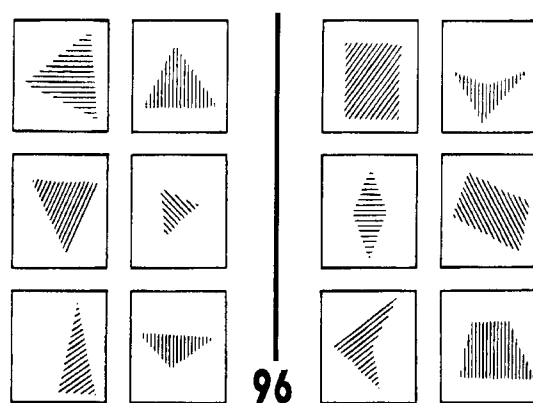


Figura 5 – Problemas de Bongard - Modo de Ver - 96

Fonte: Linhares (2000, p. 15).

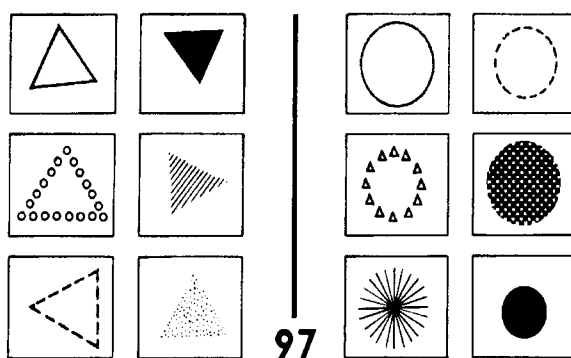


Figura 6 – Problemas de Bongard - Modo de Ver - 97

Fonte: Linhares (2000, p. 15).

Uma máquina ao “ver” as figuras acima diria que são apenas riscos contíguos ou paralelos, ou conjunto de bolinhas, ou conjunto de triângulos, quando na verdade a interpretação correta vem da identificação de um padrão e não da figura menor em si. A principal questão levantada pelos problemas de Bongard refere-se à intenção, cultura, significados ou à semântica enfim. É um problema que atinge diretamente aquele que busca encontrar uma semântica adequada à tradução. Finalizando-se, nada é mais limitador num tradutor do que uma tradução ponto a ponto, ou sem um mínimo de flexibilidade.

8. A Estratégia da “Tradução” dos Iniciantes – Embora uma tradução não tenha as características de um jogo de xadrez, na tradução de idiomas, os iniciantes lembram facilmente a teoria POS de Simon aplicada ao jogo de xadrez. Nota-se claramente que a tradução é feita por “pedaços”, ou, por palavras e que, não raro, ao final, mostra-se ineficiente. A falta de habilidade ou de conhecimento de um contexto maior faz com que o principiante busque a “tradução” do ambiente partindo de pequenos conhecimentos que lhes são mais próximos, de mais fácil adaptação e compreensão.

Linhares e Brum (2007) demonstram claramente a falta de conceitos mais abrangentes por parte dos neófitos do xadrez para que lhes possibilitassem melhores condições de fazer, ou no caso em questão, reconhecer jogadas estratégicas. A própria palavra “estratégia” denota uma visão mais ampla do jogo, ou no caso da tradução, da leitura do contexto em que a palavra está inserida;

A tradução de uma palavra ou de um texto de um idioma para outro requer o reconhecimento de um enfoque contextual e não apenas da palavra. A semântica de determinado texto está diretamente vinculada ao todo do texto e não pode ser traduzida de forma redutora ou particular. O nível estratégico numa tradução é, logicamente, pequeno, porém existe, afinal há que se tomar uma decisão, ainda que inconsciente, de como entender algo. Linhares e Brum (2007) remetem a uma compreensão da existência de fatores invisíveis, mas fundamentais que fazem com que experientes reconheçam princípios ou fundamentos nos quais se encontram as jogadas. Na tradução de idiomas, há regras ou aspectos que fazem parte do “jogo” de palavras e expressões que somente o experto tem conhecimento e consegue facilmente compreender o “não dito”, ou o “não escrito” do idioma a ser traduzido.

9. Organizando o Pensamento e o cérebro – A ligação entre pensamento e cérebro é fator de importante análise já que é no cérebro que Changeux e Connes (1996) mostra que a organização do pensamento pode se dar igualmente em níveis distintos dos quais se mostrará;

1º Nível de Organização - Situa-se no nível do próprio neurônio e da gênese dos impulsos, mas também no nível das sinapses, no ponto de articulação de células nervosas. É importante citar o fator químico que ajuda na transmissão das informações através dos impulsos elétricos (CHANGEUX; CONNES, 1996, p. 110). O mais frequente é que uma substância química, um neurotransmissor, a substitua;

2º Nível de Organização - É o nível dos circuitos de neurônios [...] os neurônios podem se reunir e se especializar em uma função particular, como os atos reflexos: caminhar, visão [...] um segundo nível plenamente realizado nos invertebrados como a minhoca ou a lesma, fornece aquilo que os etólogos chamam de “atos fixos”: bicar, voar, acasalar-se, capturar uma presa [...] sem saber se inato ou

adquirido diz-se que é um segundo nível de organização. São circuitos elementares da medula dorsal do tronco cerebral, dos “minicérebros”;

3º Nível de Organização - Constroem representações. Formam conjuntos distintos que serão aplicados em determinadas áreas de atuação. Juntam-se para fazer algo, é o neurônio a serviço de algo. A “codificação” desse algo coincide com a codificação dada pela direção e comprimento do “voto” do neurônio [...] cada neurônio é singular, único, que no conjunto estipula uma ação. É chamado de nível simbólico ou de entendimento;

4º Nível de Organização – É o nível da razão. Formam-se cadeias de representações. “assembléias de assembléias”, acontece no lobo central ou no córtex central, desenvolve-se no tempo. A temporalidade é de extrema importância [...] contribui de algum modo, ao que se chama de arquiteturas neurais da razão [...] ele separa funções “de rotina” das funções de “supervisão atenta”, que, segundo ele participam da detecção de erros, da formulação de novas hipóteses, da intervenção de novas estratégias. É razoável considerar, então, que existem áreas do encéfalo que participam da elaboração do pensamento racional [...] o lobo frontal tem um papel importante na gênese de hipóteses;

Concluindo – As condições acima expostas dão conta de nos informar que, para que uma máquina chegue às condições de um ser humano terá que transpor fases que hoje poderíamos, por motivo de coerência, reconhecer a completa impossibilidade de uma tradução cem por cento.

5 A TESE - SOBREPOSIÇÃO SEMÂNTICA

5.1 SOBREPOSIÇÃO SEMÂNTICA – A INTELIGÊNCIA POR TRÁS DA TRADUÇÃO

A Teoria P.O.S. – ao se analisar a teoria P.O.S. de Groot & Simon no campo do xadrez, percebe-se que os autores trabalham com a hipótese de codificação de combinações específicas de peças em determinados locais do tabuleiro formando um conjunto lógico de jogadas no xadrez na qual a decisão deverá vir obrigatoriamente pela análise rápida de todas as possibilidades contidas a cada passo. Uma teoria que para a tradução humana pode ser considerada como que de aplicação pouco provável. Porém, na fase inicial do conhecimento de um idioma estrangeiro, há algo que se assemelha a tal teoria, já que a tradução acontece palavra por palavra montando a frase a partir de uma prática que se aproxima bastante desse método. Entretanto, nos primórdios da tradução automática, a tradução ponto a ponto, palavra por palavra, demonstrou ser um desastre, tanto quanto o é a tradução feita pelos iniciantes quando empreendem uma tradução, provando ser a tradução humana, uma tradução de contextos e não apenas de núcleos separados, ou palavras.

Flexibilizando a Inteligência Artificial – Muito do que foi construído em inteligência artificial foi tomada como fonte e base a hipótese de um sistema de símbolos físicos postulada por Newell e Simon (1997) em que o pensamento ocorre pela manipulação de representações simbólicas que trouxeram consigo uma rigidez não condizente com a inteligência humana. David Marr foi um dos que percebeu essa necessidade de mudança: “the perception of an event or object must include the simultaneous computation of several different descriptions of it, that capture diverse aspects of the use, purpose or circumstances of the event or object” (MARR, 1977 *apud* HOFSTADTER, 1994, p. 175).

Mais recentemente McClelland e Rumelhart (1988) buscam a mesma flexibilização com modelos conexionistas sofisticados em que as representações são bastante dependentes do contexto.

Hofstadter (1994) busca a mesma flexibilidade com modelos de pensamento analógico dependente direto da percepção de alto nível e que proporciona alta flexibilidade ao pensamento. O pensamento analógico para o autor é visto como a percepção da ligação entre duas situações distintas pela identificação de estruturas essenciais idênticas que as ligam:

When people make analogies, they are perceiving some aspects of the structures of two situations – the essences of those situations, in some sense – as identical. Thus structures, of course, are a product of the process of high-level perception (HOFSTADTER, 1994, p. 179).

Linhares e Brum (2007) fazem um caminho semelhante quanto ao fator flexibilidade do pensamento quando descrevem um experimento com jogadores de xadrez no qual destacam a percepção das jogadas calcadas numa visão de papéis abstratos quando jogadores experientes fazem as jogadas. Contrastando com a teoria P.O.S. de Simon os autores levantam a questão flexibilidade da percepção em detrimento de uma mentalidade linear ou pontual como é o caso da teoria de Simon e Gilmarin (1973).

Kanerva (1990) dentro do mesmo aspecto flexibilização do campo da inteligência artificial observa que uma memória computacional organizada de forma linear não é a melhor maneira de imitar a memória humana. Assim assevera o autor:

A computer memory can be organized into arrays, records, trees, lists, stacks, and the like, and many different methods of storing and finding information in them are used [...] None of them is designed to find information when the retrieval cue is similar but not identical to an address used in storing information. In that sense, none of them is a good model of human memory [...] The aspect of the human (and animal) experience that I try to understand with my theory of memory is the retrieval of information about the past, both “at random” (i.e., directly from any part of the memory) and in sequence (KANERVA, 1990, p. 12-13).

A busca de tal flexibilidade por tais autores faz sentido já que nossa memória humana não busca dados, mas algo que faça sentido ou, que tenha algum significado para nós. Entretanto, sentido e significado são coisas que o mundo da computação não tem capacidade de “traduzir”. Como fazer então para buscarmos uma aproximação possível do que faz a mente humana? Criar formas lógicas que venham suprir tal dissociação.

Defende-se então a sobreposição semântica como a mudança de significado a partir de seu contexto. De maneira geral, a sobreposição semântica deverá estar alicerçada em um conhecimento profundo e amplo de determinado campo de atuação do elemento nuclear possibilitando relações diversas com áreas correlatas. Dentro de uma área restrita como a tradução automática, a análise da sobreposição semântica se fará em níveis conforme veremos abaixo:

1º nível - elemento nuclear □ o símbolo ou a letra;

2º nível - elemento nuclear semântico □ a palavra, o vocábulo;

3º nível - núcleo semântico contextual □ a frase;

4º nível - contexto semântico □ um texto, um livro, uma cultura, uma era.

Elemento Nuclear – Conhecer profundamente algo é conhecer o seu núcleo, sua consistência, sua composição, sua função. Se considerarmos a letra como elemento nuclear a palavra é o próximo passo desse contexto estudado. Uma letra por si só é um elemento distinto que tem sua característica, sua forma, sua diversidade de tipos e um único significado. Porém, a letra por si só em sua larga maioria termina por não ter significado. As letras “e”, “a” e “o” também não possuem um significado em si, mas se distinguem das demais por terem uma função importante, são elementos de ligação (a conjunção aditiva “e”) ou de definição do gênero das palavras (os artigos definidos “a” e “o”). Porém todas as letras só passam a ter algum sentido em si se vierem em composição com outras letras para formarem um novo sentido. A letra, portanto, é apenas um símbolo, sem um sentido específico. Surge aqui a necessidade de uma ligação, de uma sistematização de símbolos para que o sentido nasça.

Elemento Nuclear Semântico – A partir do conhecimento profundo de um elemento, passa-se a conhecer o meio de atuação desse elemento e, com ele, passa-se a dominar todo a sua base sistêmica de atuação. É a perfeita conexão entre as capacidades intrínseca do elemento e o seu meio de atuação. Buscando conhecer a capacidade intrínseca consideremos o elemento nuclear “letra”. No caso em questão o que sabemos de antemão é que “a” pode ser considerado como um artigo definido. Enquanto significado apenas indica algo, mas não é algo, tem uma função acessória, portanto. Porém, quando essa letra se relaciona com as demais 25 letras do alfabeto, ela fornece outros significados. Dá-se início a sobreposição semântica. Caso se considere a junção do “a” com o “p” nessa relação podemos extrair o primeiro significado “pá”. Se ainda se juntarem mais duas letras “to” temos agora “pato”, se se for adiante e se agregar “logia”, e agora se tem “patologia”. O “a”, por exemplo, passa a ser apenas uma parte do contexto, a unidade de sentido passa a ser a palavra e não mais a letra. Encontra-se no segundo estágio no qual o elemento nuclear é parte de um novo elemento nuclear, agora semântico, a palavra. Caso se considere a palavra a partir de agora como elemento nuclear semântico no próximo passo, o terceiro nível, ter-se-á um composto de palavras que passarão a dar sentido a esse novo núcleo semântico, agora contextual, a frase. Da mesma forma que os símbolos ou letras (p+a+t+o+l+o+g+i+a) formaram um núcleo semântico, a frase “como fazer o que você ama” também forma um novo nível semântico (como + fazer+o+que+você+ama+?) em que o contexto precisa ser levado em consideração.

Proposição 1 - Buscando fazer com que a máquina imite a percepção humana de alto nível propõe-se uma tradução através de núcleos semânticos. Segundo Hofstadter (1994, p. 171) “high-level perception begins at the level of processing where concepts begin to play an important role. Processes of high-level perception may be subdivided again into a spectrum, from the concrete to the abstract”.

Proposição 2 - A Neurobiologia a Serviço da Inteligência - Mudando de Estágio –
A mudança de estado ou de núcleo semântico pode ser abordada buscando nos conceitos da ciência biológica uma via de solução. Como passar de um nível semântico a outro? Segundo Changeux e Connes (1996) que, ao mencionar os níveis de organização no sistema nervoso, defende que a transição de um nível ao nível seguinte requer dois componentes fundamentais: um gerador de diversidade e um sistema de seleção. Novamente o que se observa é uma flexibilidade na passagem de um nível a outro, conforme nos diz:

Em um dado nível, os elementos se recombinaem entre si, variam de maneira aleatória e elaboram “formas” transitórias que se devem ao nível de organização imediatamente superior. Essas formas se engendram a partir de elementos já estruturados, logo, não necessariamente a partir de átomos. Há portanto uma produção de “variações” que podem transitoriamente ter acesso a um nível superior de organização. Um mecanismo de seleção estabiliza em seguida alguns de seus estados transitórios, e desse modo engendram um nível de organização mais elevado. A função age de maneira retroativa sobre a transição “matéria-forma”. O critério de seleção lega-se portanto a “nova” função determinada por uma forma transitória produzida pelo gerador de diversidade. Se essa nova função corresponde a uma ação sobre o mundo exterior que favorece a sobrevivência do organismo ela é selecionada (CHANGEUX; CONNES, 1996, p. 122-123).

As conexões entre as células nervosas, as sinapses, não se formam de uma só vez no ser humano. Evoluem de duas formas: uma evolução do número de conexões, que ocorre durante o desenvolvimento e uma evolução da eficácia das conexões entre os neurônios Changeux e Connes (1996). Interessante atentar-se para o fato de que há uma qualificação ou, no dizer do autor, uma evolução da eficácia em função da experiência.

Do Mundo Biológico à Tradução Automática - A frase “How to do what you Love?” possui 5.224.576 possíveis traduções se a tradução explorar o significado de cada palavra e fizer uma relação linear entre o significado de cada palavra. Se houvesse a “compreensão das palavras “to do” como um único núcleo semântico essa possibilidade ficaria restrita a 19.208. Se a expressão “How to do” fosse compreendida como um único núcleo semântico, então a possibilidade de tradução cairiam para 2.744.

Supondo-se que a expressão “*you Love*” fosse também considerada como núcleo semântico, então se teriam apenas 14 possíveis traduções. E, é claro, se a frase já tivesse sido traduzida, então teríamos um único núcleo semântico e a tradução seria completa. Entretanto como passar de palavras para núcleos semânticos? Faz-se por experiência. À medida que a expressão fosse sendo traduzida, formar-se-iam expressões que, aos poucos, iriam se sedimentando pela repetição através de um algoritmo com peso que viesse valorizar a repetição do termo traduzido, isto é, a memorização por reconhecimento de padrão (KLEIN, 1998).

6 CONCLUSÃO

Tal qual se deu no início dos estudos da inteligência artificial onde a base estava formada sob estudos da inteligência humana, hoje isso não mudou. O que necessariamente precisa mudar é a visão ainda bastante fechada com relação a linearidade do pensamento. A dissertação com os diversos conceitos levantados nos mais diversos autores em campos distintos de atuação busca dar novos subsídios a novos estudos com um enfoque distinto aos estudiosos da inteligência artificial.

O próprio crescimento, ainda que pequeno, no campo da tradução automática tem como fator de sucesso a flexibilidade engendrada já que os processos de descoberta de novos padrões por meios estatísticos, traz consigo uma natural flexibilidade e, por conseguinte, um novo crescimento a cada tempo.

Novos estudos com base na inteligência humana que tenham como foco sua não linearidade poderá dar um salto não apenas à tradução automática, mas a própria inteligência artificial. Quanto ao resultado do presente estudo, pode se inferir que a sobreposição semântica é apenas uma das diversas vias que podem ser usadas para a flexibilização e celeridade necessárias à inteligência artificial. Os resultados demonstram claramente que há mais do que resultados lógicos e passíveis de manipulação matemática básica, ou seja, há mais do que se possa ver ou conceber, há muito ainda a ser explorado na ciência cognitiva como força propulsora de novas descobertas a ciência artificial. O hiato de tempo encontrado entre as possíveis traduções e as que o ser humano faz de fato é um sinal claro do que ainda separa o conhecido do desconhecido em termos cognitivos. A sobreposição semântica poderá significar um salto pela otimização e celeridade que poderá dar ao processo de construção aos padrões levantados pelos meio estatísticos como uma forma de imitar essa habilidade humana de tradução do meio que vive.

REFERÊNCIAS

BRANTS, T. et al. "Large Language Models in Machine Translation". In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING AND COMPUTATIONAL NATURAL LANGUAGE LEARNING, 2007, Praguepp. **Anais...** Prague: Joint Meeting following, 2007. p. 858-867.

CHALMERS, D.; FRENCH, R.; HOFSTADTER, D. R. High-level perception, representation, and analogy: A critique of artificial intelligence methodology. **Journal of Experimental and Theoretical Artificial Intelligence**, v. 4, n. 3, p. 185-211, 1992. Disponível em: <<http://www.informaworld.com/index/778787585.pdf>>. Acesso em: 05 jun. 2009.

CHANGEUX, J. P.; CONNES, A. **Matéria e pensamento**. São Paulo: Editora da Universidade Estadual Paulista, 1996.

EVANS, D. S. The economics of the online advertising industry. **Review of Network Economics**, v. 7, n. 3, Sept. 2008. Disponível em: <<http://ssrn.com/abstract=1086473>>. Acesso em: 11 dez. 2009.

FRENCH, R. Peeking behind the screen: the unsuspected power of the standard turing test. **Journal of Experimental & Theoretical Artificial Intelligence**, v. 12, p. 331-340, 2000a. Disponível em: <<http://www.informaworld.com/index/24PHYLU4NX08NHQF.pdf>>. Acesso em: 05 jun. 2009.

FRENCH, R. When coffee cups are like old elephants, or why representation modules don't make sense. **Understanding Representation in the Cognitive Sciences**, parte 3. p. 158-163. 2000b. Disponível em: <<http://www.springerlink.com/index/m5p84u507h163530.pdf>>. Acesso em: 05 jun. 2009.

FRENCH, R.; LABIOUSE, C. Four problems with extracting human semantics from large text corpora. In: ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY, 24., 2002, Texa. **Proceedings...** Texa: Cognitive Science Society, 2002. p. 316-322. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.3395&rep=rep1&type=pdf>>. Acesso em: 05 jun. 2009.

FRENCH, R.; LABIOUSE, C. Why co-occurrence information alone is not sufficient to answer subcognitive questions. **Journal of Experimental & Theoretical Artificial Intelligence**, v. 13, n. 4, p. 421-429, 2001.

FRIEDMAN, T. L. **O mundo é plano**: uma breve história do século XXI. Rio de Janeiro: Objetiva, 2007.

FRIEDMAN, T. L. **The world is flat**: a brief history of the twenty-first century. New York: Farrar Straus & Goroux, 2007.

HOFSTADTER, D. R. **Fluid concepts and creative analogies**: computer models of the fundamental mechanism of thought. New York: Basic Books, 1994.

HOFSTADTER, D. R. **Godel escher, bach**: um entrelaçamento de gênios gigantes. 18. ed. São Paulo: Editora da Universidade de Brasília, 2001.

JARVIS, J. **What would google do?** 2009. Disponível em: <<http://online.wsj.com/article/SB123386274048053567.html>>. Acesso em: 15 jan. 2009.

KANERVA, P. **Sparse distributed memory**. 2. ed. Massachusetts: Massachusetts Institute of Technology, 1990.

KLEIN, G. **Sources of power**: how people make decisions. Massachusetts: Massachusetts Institute of Technology, 1998.

KUMAR, S.; OCH, F.; MACHEREY, W. Improving word alignment with bridge languages. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING AND COMPUTATIONAL NATURAL LANGUAGE LEARNING, 2007, Praguepp. **Anais...** Prague: Joint Meeting following, 2007. p. 42-50.

LINHARES, A. A glimpse at the metaphysics of bongard problems. **Artificial Intelligence**, v. 121, p. 251-270, 2000.

LINHARES, A. An active symbols theory of chess intuition. **Minds and Machines**, v. 15, n. 2, p. 131-181, 2005. Disponível em: <<http://www.springerlink.com/index/10.1007/s11023-005-5045-7>>. Acesso em: 05 jun. 2009.

LINHARES, A.; BRUM, P. R. da S. How can experts see the invisible? reply to bilalíc and gobet. **Cognitive Science Society**, v. 33, p. 748-751, 2009.

LINHARES, A.; BRUM, P. R. da S. Understanding our understanding of strategic scenarios: what role do chunks play? **Cognitive Science**, v. 31, p. 989-1007, 2007.

MACHEREY, W. et al. Lattice-based minimum error rate training for statistical machine translation. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2008, Honolulu. **Anais...** Honolulu: STANFORD UNIVERSITY, 2008. p. 725-734.

MACHEREY, W.; OCH, F. J. An empirical study on computing consensus translations from multiple machine translation systems. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING AND COMPUTATIONAL NATURAL LANGUAGE LEARNING, 2007, Praguepp. **Anais...** Prague: Joint Meeting following, 2007. p. 986-995.

MARR, D. Artificial intelligence: a personal view. **Artificial Intelligence**, v. 9, p. 37-48, 1977.

MCCLELLAND, J. L.; RUMELHART, D. E. **Explorations in paralell distributed processing**: a handbook of models, programs, and exercises. Cambridge: MIT, 1988.

NEWELL, A.; SIMON, H. A. **Computer science as empirical inquiry**: symbols and search. ind Design II. Cambridge: MIT, 1997.

PINKER, S. **How the mind works**. New York: W.W.Norton & Co Inc, 1997.

PINKER, S. **O instinto da linguagem**: como a mente cria a linguagem. São Paulo: Martins Fontes, 2002.

SIMON, H. A.; GILMARTIN, K. A simulation of memory for chess positions. **Cognitive Psychology**, v. 5, n. 1, p. 29-46, July 1973. Disponível em: <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6WCR-4D5XCKP-6Y&_user=10&_coverDate=07%2F31%2F1973&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&_view=c&_searchStrId=1482745828&_rerunOrigin=scholar.google&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=469234119396f00f9e6938e0c8fad12a&searchtype=a>. Acesso em: 05 jun. 2009.

TROMBLE, R. W. et al. Lattice minimum bayes-risk decoding for statistical machine translation. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2008, Honolulu. **Anais...** Honolulu: STANFORD UNIVERSITY, 2008. p. 620-629.

VYGOTSKY, L. S. **A formação social da mente**: o desenvolvimento dos processos psicológicos superiores. 7. ed. São Paulo: Martins Fontes, 2007.

VYGOTSKY, L. S. **Estudos sobre a história do comportamento**: o macaco, o primitivo e a criança. Porto Alegre: Artes Médicas, 1996.

WHITAKER, J. et al. Internet business practices across the globe: lessons from emerging economies. **Ross School of Business Paper**, n. 1113, 2008. Disponível em: <<http://ssrn.com/abstract=1017995>>. Acesso em: 05 jul. 2009.

WIKIPEDIA. The Free Encyclopedia. **Google translate**. Disponível em: <http://en.wikipedia.org/wiki/Google_Translate#cite_note-2>. Acesso em: 05 jun. 2009.