

O PECADO ORIGINAL DA LINGÜÍSTICA COMPUTACIONAL

Ronaldo MARTINS*

- RESUMO: Este artigo explora algumas das questões epistemológicas envolvidas na demarcação do território da Linguística Computacional, mais especificamente do Processamento Automático das Línguas Naturais, quando concebido como região de interseção entre a Linguística Tradicional, de um lado, e as Ciências da Computação, de outro. O texto procura mapear a natureza dessa interseção e constata que os objetivos, os objetos e os métodos da Linguística Computacional não coincidem com os da Linguística Tradicional, razão pela qual o novo domínio constituiria não exatamente um subdomínio ou área de aplicação desta, mas um novo campo de exploração científica, colonizado ainda pelas práticas de análise e de investigação que caracterizam a Linguística Tradicional, mas cujas contradições internas conduziram progressivamente à formação de uma nova razão linguística, que envolveria a constituição de categorias menos antropocêntricas e antropomórficas de análise e de descrição linguística, mas que se revelariam mais adequadas para a emulação, pela máquina, do comportamento linguístico humano.
- PALAVRAS-CHAVE: Linguística Computacional. Processamento Automático das Línguas Naturais. Tradução Automática.

Introdução

A Tradução Automática (*Machine Translation*), ou Tradução Mecânica (*Mechanical Translation*), ostenta a dupla condição de ser, entre os aplicativos do processamento automático das línguas naturais, a mais antiga veleidade e a mais reiterada frustração. Nas mais recentes competições entre sistemas promovidas pelo *National Institute of Standards and Technology* (NIST, 2008), o melhor sistema de tradução automática (Google) não chegava a nem mesmo 50% da referência humana. Se considerarmos que as primeiras tentativas de construção de um sistema que traduzisse automaticamente os enunciados entre duas línguas naturais coincidem com o desenvolvimento da técnica de criptografia, logo após a Segunda Guerra Mundial, seremos forçados a admitir que todo o capital simbólico e científico acumulado pelos estudos da Tradução Automática nos últimos 50 anos não foi capaz, ainda, de produzir sistemas inteligentes que pudessem substituir o tradutor humano. Nenhum dos sistemas disponíveis, sejam

* UNIVÁS – Universidade do Vale do Sapucaí. Faculdade de Filosofia, Ciências e Letras Eugênio Pacelli. Pouso Alegre - MG – Brasil. 37550-000 – ronaldo.martins@uol.com.br

provenientes das iniciativas de mercado, sejam derivados da pesquisa acadêmica, produziu, até hoje, resultados que pudessem prescindir de edição humana. Pelo contrário: mesmo para línguas que têm concentrado uma soma considerável de investimentos, não foram observados resultados que permitissem, qual em um teste de Turing, confundir o desempenho do tradutor mecânico com a expectativa de desempenho de um tradutor humano.

O sempre renovado fracasso das inúmeras tentativas de automação vem conduzindo a uma revisão do processo de tradução, no sentido de serem redefinidos, principalmente, seus objetivos e seu escopo de atuação. Martin Kay (1995, p.247-248) sugere, em um texto sintomaticamente intitulado “*Machine Translation: the disappointing past and present*”, três possibilidades de encaminhamento da discussão:

First, in the long run, there is no alternative to continuing to build more faithful models of human behavior. The second alternative is to design systems involving both people and machines, assigning to each those parts of the task to which they are best suited. The third is to seek ways of modifying the task so that the machine will naturally have greater control over the context.

Essas sugestões parecem-me francamente equivocadas e inócuas. Não creio que o acúmulo gradual de informação proveniente de redemarcações da tarefa original conduzirá à revolução qualitativa que, espera-se, virá redimir o campo, oferecendo alguma real expectativa de produção de um sistema escalável, robusto, eficaz e eficiente, capaz de gerar resultados completamente automáticos e de boa qualidade. Em nenhuma das direções propostas, o problema da Tradução Automática é realmente enfrentado, procedendo-se a um recorte metodológico que, se tem, evidentemente, flagrante legitimidade científica, faz, porém, supor que o processo de produção do conhecimento se dá antes de forma gradual, pelo acúmulo de informações que, incorporadas a um sistema dotado de alguma elasticidade, permitiriam que fosse ampliada, contínua, progressiva e indefinidamente, sua capacidade de atuação. Não creio que seja o caso. O que se percebe, na prática, é que a complexidade do processo não está apenas ou diretamente relacionada à quantidade ou à qualidade de informação necessária, mas, sobretudo, às premissas sobre as quais toda a ciência do Processamento Automático das Línguas Naturais (PLN), de que a Tradução Automática é parte, está fundada.

O PLN parece (re)viver hoje o cenário de crise que, segundo Thomas Kuhn (1992, p.24-25, grifo nosso), também teria caracterizado, por exemplo, a história das ciências físicas:

A ciência normal, atividade na qual a maioria dos cientistas emprega inevitavelmente quase todo seu tempo, é baseada no pressuposto de que a comunidade científica sabe como é o mundo. Grande parte do sucesso do empreendimento deriva da disposição da comunidade para defender esse pressuposto – com custos consideráveis, se necessário. Por exemplo, a ciência normal frequentemente suprime novidades fundamentais, porque estas subvertem necessariamente seus compromissos básicos. Não obstante, na medida em que esses compromissos retêm um elemento de arbitrariedade, a própria natureza da pesquisa normal assegura que a novidade não será suprimida por muito tempo. Algumas vezes um problema comum, que deveria ser resolvido por meio de regras e procedimentos conhecidos, resiste ao ataque violento e reiterado dos membros mais hábeis do grupo em cuja área de competência ele ocorre. **Em outras ocasiões, uma peça de equipamento, projetada e construída para fins de pesquisa normal, não funciona segundo a maneira antecipada, revelando uma anomalia que não pode ser ajustada às expectativas profissionais, não obstante esforços repetidos.** Desta e de outras maneiras, a ciência normal desorienta-se seguidamente. E quando isso ocorre – isto é, quando os membros da profissão não podem mais esquivar-se das anomalias que subvertem a tradição existente da prática científica – então começam as investigações extraordinárias que finalmente conduzem a profissão a um novo conjunto de compromissos, a uma nova base para a prática da ciência.

Não se verificou ainda o episódio extraordinário – a “revolução científica” – em que se observaria a “alteração dos compromissos profissionais”, o “complemento desintegrador da tradição à qual a atividade de ciência normal está ligada”, a “[...] alteração nos problemas à disposição do escrutínio científico e nos padrões pelos quais a profissão determinava [determina] o que deveria ser considerado como um problema ou como uma solução de problema legítimo.” (KUHN, 1992, p.25). Não houve ainda a transformação da imaginação científica.

Mas a anomalia se faz presente em toda parte: as peças linguístico-computacionais, sejam recursos (o chamado *lingware*, como gramáticas e dicionários), sejam instrumentos (como *taggers*, *stemmers*, *splitters*, *parsers* etc.), sejam aplicativos de uso final (como aconselhadores ortográficos, revisores gramaticais e de estilo, tradutores, alinhadores de texto, sumarizadores, reconhecedores e sintetizadores de fala etc.), vêm teimando em não funcionar da maneira adequada. Seus resultados, ainda que possam ser extremamente úteis, especialmente quando envolvidas habilidades linguísticas mecânicas e repetitivas, longe estão de poder ser considerados verdadeiramente adequados.

Pode-se, evidentemente, e em muitos casos, alegar a relativa juventude do domínio, cuja exploração sistemática é realmente recente. Mas o argumento não vale para a Tradução Automática, cuja prospecção metódica teve início em 1949,

e que vem concentrando, ao longo desses mais de 50 anos, numerosos grupos de pesquisa e desenvolvimento. Parece mais razoável supor que o PLN nunca chegou a constituir propriamente uma “ciência normal”, assim entendida “[...] a pesquisa firmemente baseada em uma ou mais realizações científicas passadas [...] reconhecidas durante algum tempo pela comunidade científica específica como proporcionando os fundamentos para sua prática posterior.” (KUHN, 1992, p.29). As realizações que poderiam nortear o desenvolvimento de novos projetos ou constituem tratados de Linguística pura, que tematizam a linguagem fora do escopo computacional, ou constituem trabalhos de Matemática Aplicada, nomeadamente de Inteligência Artificial, que abordam a linguagem de maneira muito superficial.

Não houve ainda um trabalho fundacional que definisse “implicitamente os problemas e métodos legítimos de um campo de pesquisa para as gerações posteriores de praticantes da ciência”, e cujas realizações tivessem sido: a) “suficientemente sem precedentes para atrair um grupo duradouro de partidários, afastando-os de outras formas de atividade científica dissimilares”; b) “suficientemente abertas para deixar toda a espécie de problemas para serem resolvidos pelo grupo definido de praticantes da ciência.” (KUHN, 1992, p.30).

Uma evidência de que o PLN não se teria ainda “normalizado” é o fato de que parece não caber, dentro do campo, a noção de paradigma, definido como “[conjunto de] exemplos aceitos na prática científica real – exemplos que incluem, ao mesmo tempo, lei, teoria, aplicação e instrumentação – [e que] proporcionam modelos dos quais brotam as tradições coerentes e específicas da pesquisa científica.” (KUHN, 1992, p.30). A comparação com a Óptica Física pré-newtoniana é, a essa altura, extremamente ilustrativa:

[...] qualquer um que examine uma amostra da Óptica Física anterior a Newton poderá perfeitamente concluir que, embora os estudiosos dessa área fossem cientistas, o resultado líquido de suas atividades foi algo menos do que ciência. Por não ser obrigado a assumir um corpo qualquer de crenças comuns, cada autor de Óptica Física sentia-se forçado a construir novamente seu campo de estudos desde os fundamentos. A escolha das observações e experiências que sustentavam tal reconstrução era relativamente livre. Não havia qualquer conjunto-padrão de métodos ou de fenômenos que todos os estudiosos da Óptica se sentissem forçados a empregar e explicar. Nestas circunstâncias o diálogo dos livros resultantes era frequentemente dirigido aos membros das outras escolas tanto como à natureza. Hoje em dia esse padrão é familiar a numerosos campos de estudos criadores e não é incompatível com invenções e descobertas significativas. Contudo, este não é o padrão de desenvolvimento que a Óptica Física adquiriu depois de Newton e nem aquele que outras ciências da natureza tornaram familiar hoje em dia. (KUHN, 1992, p.33).

O “campo de estudos” do PLN, embora evidentemente profícuo na produção de aplicativos de utilidade incontestável, constitui principalmente uma dispersão, sem que possa ser observada, nitidamente, a hegemonia de um corpo teórico sobre os demais. Trata-se, na verdade, de uma coleção de posturas difusas e fragmentárias (e inconsúteis) que orbitam um objetivo comum: ensinar a máquina a falar.

Convivem, por exemplo, no mesmo campo de investigação, abordagens simbólicas e subsimbólicas (estatísticas). As primeiras – chamadas “lógicas” ou “lógico-indutivas” –, organizadas em torno da ideia de “regra”, admitem a possibilidade de identificarem-se, explicita, declarativa, e introspectivamente, sob a aparente diversidade das ocorrências linguísticas, padrões regulares e computacionalizáveis de comportamento linguístico. As segundas – entre as quais as “conexionistas” – estão escoradas na chamada “aprendizagem de máquina” (*machine learning*), em algoritmos genéticos, no processamento de extensos *corpora*, na consideração de numerosos exemplários de ocorrências linguísticas, para a recuperação (induzida automaticamente) de regularidades subjacentes que, explicitadas, talvez não façam nenhum sentido para o observador, mas, adotadas, provocariam, do ponto de vista procedimental, os mesmos (?) resultados produzidos, em situações de uso da língua, pelo falante humano.

Este movimento pendular, do tráfego de teorias que avançam e recuam, nesta ou naquela direção, a depender das condições de produção da ciência e do discurso científico, e que repelem e imantam, de forma não necessariamente articulada – porque haverá sempre os “eccléticos” e os “híbridos” –, os vários grupos de pesquisa e desenvolvimento, parece confirmar que o PLN não se vem constituindo por meio do desenvolvimento linear do conhecimento – como podem supor, por exemplo, Popper (1963) e Hempel (1970) –, mas que também não caracteriza, pelo menos não ainda, uma “sucessão de descontinuidades relativamente bruscas”, como prevê Kuhn (1992).

Mais do que propriamente superposições de modelos de explicação científica, a Tradução Automática, e também o PLN, parecem confirmar a existência de justaposições (proliferações) de teorias que, competindo entre si, acabam prevalecendo antes por fatores extracientíficos (tenacidade, por exemplo) do que pelo caráter proibitivo e falseável das hipóteses que as compõem, sempre passíveis de contornos e adaptações auxiliares. Muitos dos aspectos relacionados ao domínio têm se revelado muito mais uma questão de propaganda, de poder de persuasão (e, seguramente, de estratégias de cooptação), de políticas de fomento que privilegiam esta ou aquela abordagem, este ou aquele objeto de estudo, do que o mero acúmulo de aproximações graduais (e desinteressadas) à Verdade.

Trata-se, portanto, de um terreno pantanoso, de limites movediços, cuja imaturidade é flagrante, mas que deve ser ocupado e desbravado, se estivermos

efetivamente dispostos a fazer do PLN e da Tradução Automática, mais do que uma técnica, mais do que uma coleção de relatos de experiência, mais do que um repertório de práticas desconcertadas, uma ciência normal no sentido kuhnyano do termo.

Dos objetivos e das tarefas do PLN

Para que se possa normalizar o PLN, impõe-se, de partida, o imperativo de definir qual seria o seu objeto e quais seriam os seus objetivos. E parece-me pertinente tentar retrilhar o percurso argumentativo percorrido por Saussure (1973) para a demarcação científica da Linguística. Aqui como lá, o objeto de investigação não se revela, pelo menos não à primeira vista, autoevidente:

Outras ciências trabalham com objetos dados previamente e que se podem considerar, em seguida, de vários pontos de vista; em nosso campo, nada de semelhante ocorre. [...] Bem longe de dizer que o objeto precede o ponto de vista, diríamos que é o ponto de vista que cria o objeto; aliás, nada nos diz de antemão que uma dessas maneiras de considerar o fato em questão seja anterior ou superior às outras. (SAUSSURE, 1973, p.15).

Para Saussure (1973, p.13), as tarefas da Linguística seriam três:

- a) fazer a descrição e a história de todas as línguas que puder abranger, o que quer dizer: fazer a história das famílias de línguas e reconstituir, na medida do possível, as línguas-mães de cada família;
- b) procurar as forças que estão em jogo, de modo permanente e universal, em todas as línguas e deduzir as leis gerais às quais se possam referir todos os fenômenos peculiares da história;
- c) delimitar-se e definir-se a si própria.

Sua utilidade, no entanto, seria confusa: “bem poucas pessoas têm a respeito ideias claras: não cabe fixá-las aqui” (SAUSSURE, 1973, p.14).

Situação inversa pode ser observada no PLN, cuja utilidade parece não ser matéria de controvérsia, mas cujas tarefas são passíveis de discussão. A confiarmos em Grishman (1986, p. 4), “[...] embora os objetivos da pesquisa em Linguística Computacional sejam muito variados, uma motivação primária tem sido sempre o desenvolvimento de sistemas práticos específicos que envolvem línguas naturais.” O autor identifica três grandes classes de aplicações que têm sido centrais no desenvolvimento do domínio: a Tradução Automática (*machine translation*), a recuperação de informação (*information retrieval*) e as interfaces homem-máquina (*man-machine interfaces*). Mas ressalta que, “além dessas ‘engenharias’, desses

temas orientados para aplicações, muitos investigadores têm objetivos de pesquisa 'científica' que são independentes de uma aplicação em particular". Cita, como exemplo, o uso da Linguística Computacional para testar as gramáticas propostas pela Linguística teórica. No entanto o próprio autor observa que isto não é algo que venha efetivamente ocorrendo.

Admitamos, por ora, que o PLN tenha efetivamente essa vocação prática, engenheira, de produção de *softwares* linguísticos, de desenvolvimento de sistemas especialistas que, ao fim e ao cabo, invejam e perseguem o desempenho observado para um falante humano, em suas várias formas de uso e manipulação da linguagem.

Se os objetivos são razoavelmente convergentes, o mesmo não se pode dizer das tarefas do PLN. Diferenciá-lo da Linguística me parece que seja, por exemplo, uma tarefa urgente, para que se possa justificar a necessidade e a pertinência de uma disciplina autônoma, com métodos e objeto próprios. No entanto não saberia dizer em que medida esta angústia é compartilhada pela comunidade de pesquisadores, que parece muitas vezes nutrir a esperança de um diálogo profícuo e interdisciplinar entre Linguística e Inteligência Artificial. Nessa linha, o que se observa, muitas vezes, é o desejo de que o PLN venha a constituir um domínio ponte, ou uma região de interseção, entre práticas linguísticas e computacionais já supostamente consolidadas.

Exemplo dessa expectativa pode ser encontrado em Reynolds (1954 apud HUTCHINS, 1997, p.222) que, ao reportar a sessão final do primeiro dia da primeira conferência sobre Tradução Automática realizada no MIT, observava:

There was general agreement on the part of both the panel and the audience that mechanical translation was feasible. It was interesting to note that the computer engineers present presented all of the difficulties standing in the way of producing a mechanical translator from the engineering standpoint; the linguist, from his standpoint; and the psychologists and philosophers from the standpoint of their respective disciplines. Each agreed, however, that, if the other two groups did their work, we could in the near future produce adequate and intelligible machine programmed translations.

O desejo deste inter-relacionamento pode ser justificado pelo fato de que a Linguística e o PLN têm realmente muitas tarefas comuns, como a de desenvolver um sistema formal de representação das línguas naturais, já que a matematização da linguagem é, com efeito, uma possibilidade teórica que afeta ambas as disciplinas. No entanto, em que pese a matriz muitas vezes compartilhada, os sistemas derivados desses mesmos compromissos em uma e outra disciplina são hoje razoavelmente distintos, principalmente porque o PLN tem negligenciado

atores extralinguísticos (como o falante, o ouvinte e o contexto, por exemplo) que, na Linguística, vêm participando dos modelos descritivos. Essa parece ser, aliás, a razão pela qual se diz, com frequência, que o PLN tem envolvido, à primeira vista, uma concepção muito ingênua e esquemática do funcionamento da linguagem.

No entanto – em defesa do PLN, é importante que o diga –, quase 100 anos após a realização do *Curso de linguística geral*, o conhecimento científico que os falantes têm a respeito da língua não atingiu ainda o estágio de permitir que pudessem ser replicadas, mesmo em modelos estritamente teóricos, sem vinculação imediata com a computação, atividades linguísticas mais sofisticadas, principalmente as de natureza semântica, como a tradução. A língua permanece uma esfinge a ser decifrada, e o processo de multiestratificação em diferentes níveis de análise, se permitiu, realmente, que um conjunto relativamente expressivo de informações fosse produzido, não gerou ainda a síntese necessária para que se tivesse, do funcionamento da linguagem, uma perspectiva mais geral e consolidada. Pelo contrário, a fragmentação da descrição linguística impediu que se constituísse uma teoria geral da linguagem, sem a qual qualquer tentativa de automação do comportamento linguístico se vê reduzida, principalmente, à experimentação cega de estratégias de validade limitada.

O fracionamento (ou a compartimentalização) da Linguística não pode constituir, porém, salvo-conduto para abordagens fragmentárias, e claro está que limitações operacionais não podem servir de álibi para que conceitos que se têm revelado pertinentes sejam rebaixados à condição de dispensáveis na descrição da linguagem. Observam-se, assim, entre linguistas e engenheiros da linguagem, acusações de parte a parte. De um lado, denuncia-se a ignorância de categorias descritivas da Linguística e rechaça-se o conseqüente subdimensionamento da complexidade dos problemas da linguagem; de outro, registra-se a dispersão e a volatilidade da Linguística e criticam-se categorias que não acompanham as exigências de formalização requeridas pela máquina. Em ambas as partes, a vívida consciência de uma desejável interdependência.

Os ataques recíprocos, da Linguística ao PLN, e do PLN à Linguística, podem ser particularmente ilustrados em fragmentos como o apresentado abaixo, extraído de um artigo intitulado “*Natural language parsing and Linguistic Theories: can the marriage be saved?*”, publicado, em 1991, em um periódico da área de Linguística (*Studies in Language*):

The fragmentation of the field of linguistics and the fuzzy philosophizing that passes for 'linguistic theory' among large segments of the linguistic population don't inspire much confidence among the language engineers, and the blissful ignorance about elementary facts of natural language that the engineers flaunt smugly in their publications does little to convince serious linguists that there is anybody out there

among mainframes who has any interest in applying whatever linguistic scientists may have found out (STAROSTA, 1991, p.178).

O texto de Stanley Starosta (1991) é emblemático não pela retomada dos argumentos que são cotidianamente esgrimidos em cada uma das duas trincheiras, mas principalmente pela indicação, implícita, de que a verdadeira fundação do PLN somente ocorreria quando fosse observada a desfragmentação (homogeneização?) da Linguística, de tal forma que se criassem as condições materiais para que pudesse ser superado o estado de ignorância sobre fatos elementares das línguas naturais que hoje acometeria os engenheiros da linguagem. À parte os problemas políticos e científicos associados à consignação de uma (utópica) unidade disciplinar à Linguística, entendo que se pode extrair do texto, como um de seus corolários principais, a proposição de que a distância que separa a Linguística do PLN seria consideravelmente menor do que a que a distinguiria da Psicologia, da Antropologia, da Gramática Normativa, da Filologia, “que separamos claramente da Linguística, mas que, por culpa de um método incorreto, poderiam reivindicar a linguagem como um de seus objetos” (SAUSSURE, 1973, p.16). A acompanhar essa linha de pensamento, o PLN, longe de constituir um domínio autônomo, seria uma instância de aplicação da Linguística, com a qual deveria “casar-se” ao invés de dela se separar.

Percebe-se, portanto, que a indicação das tarefas do PLN não é exatamente isenta de polêmica e passa, necessariamente, pela definição de seu estatuto como ciência, particularmente em relação à Linguística. Resta saber – e esta a principal questão aqui – se essa tão sonhada interdisciplinaridade tem efetiva validade epistemológica, principalmente se considerados os diferentes compromissos que Linguística e PLN estabelecem com a linguagem. Em última instância, poder-se-ia afirmar que a crítica que a Linguística faz às abordagens computacionais, acusando-as de excessiva simplificação, talvez não passem de “linguisticismo”, de “reducionismo linguístico”, da tentativa de apagamento de diferenças que podem ser constitutivas do novo domínio¹. Mas pode ser também que esses esquematismos sejam exatamente o que deles se diz: que representam simplificação grosseira e contraproducente dos fenômenos da linguagem.

¹ O termo “linguisticismo” faz aqui evidente alusão a “fiscalismo”, termo cunhado por Rudolf Carnap, em *Conceituação fiscalista*, de 1926, como expressão da ideia de que a física constituiria um paradigma a ser utilizado por todas as ciências, fossem elas naturais ou humanas: “Uma das tarefas mais importantes, relativas à lógica da ciência, será o desenvolvimento das operações que o fiscalismo sustenta que são possíveis: indicar as regras sintáticas para a inserção dos diferentes conceitos biológicos, psicológicos e sociológicos na linguagem física. Essa análise dos conceitos de linguagem parciais conduz à concepção de uma linguagem unitária que suprimiria o estado de dispersão que reina atualmente na ciência.” (CARNAP, 1937). Uma hipótese que mereceria ser considerada com muito mais profundidade do que este texto permite é a de que a Linguística, que já foi considerada “ciência-piloto” e emprestou o paradigma estruturalista a muitas outras ciências humanas, talvez guardasse a ilusão (totalitária) de que todas as ciências que tomam a linguagem por objeto (ainda que indireto) de exploração deveriam obrigatoriamente compartilhar de uma mesma metalinguagem.

Da matéria do PLN

As indefinições que marcam o programa teórico do PLN não derivam apenas da ausência de um núcleo de tarefas comuns em torno das quais possam se organizar pesquisadores e desenvolvedores. O objeto do PLN é igualmente proteiforme. Como o era, de resto, o próprio objeto da Linguística no início do século XX.

Para Saussure (1973, p.17), o objeto de estudo da Linguística seria a língua, o “produto social da faculdade de linguagem e um conjunto de convenções necessárias, adotadas pelo corpo social para permitir o exercício dessa faculdade nos indivíduos”.

A opção de Saussure (1973) pela *langue* reveste-se de uma necessidade metodológica, mais do que de uma ingenuidade teórica acerca dos fatos da linguagem. Para ele, seja qual for a perspectiva que se adote para a consideração das manifestações da linguagem, “o fenômeno linguístico apresenta perpetuamente duas faces que se correspondem e das quais uma não vale senão pela outra.” (SAUSSURE, 1973, p.15). Haveria, assim, a face acústica e a face articulatória; a face fisiológica (acústico-vocal) e a face psicológica (mental); a face individual e a face social; a face estática e a face dinâmica:

Tomada em seu todo, a linguagem é multiforme e heteróclita; a cavaleiro de diferentes domínios, ao mesmo tempo física, fisiológica e psíquica, ela pertence além disso ao domínio individual e ao domínio social; não se deixa classificar em nenhuma categoria de fatos humanos, pois não se sabe como inferir sua unidade. (SAUSSURE, 1973, p.17).

O problema da ideia de “linguagem”, segundo Saussure (1973), seria justamente esta dispersão. A linguagem, do ponto de vista do seu todo, seria “incognoscível”:

Dessarte, qualquer que seja o lado por que se aborda a questão, em nenhuma se nos oferece integral o objeto da Linguística. Sempre encontramos o dilema: ou nos aplicamos a uma lado apenas de cada problema e nos arriscamos a não perceber as dualidades assinaladas acima, ou, se estudarmos a linguagem sob vários aspectos ao mesmo tempo, o objeto da Linguística nos aparecerá como um aglomerado confuso de coisas heteróclitas, sem liame entre si. (SAUSSURE, 1973, p.16).

A solução encontrada por Saussure (1973, p.16-17, grifo do autor) para evitar tamanha volatilidade e que se transformaria no divisor de águas da Linguística é bem conhecida:

Há, segundo nos parece, uma solução para todas essas dificuldades: *é necessário colocar-se primeiramente no terreno da língua e tomá-la como norma de todas as outras manifestações da linguagem*. De fato, entre tantas dualidades, somente a língua parece suscetível duma definição autônoma e fornece um ponto de apoio satisfatório para o espírito.

Saussure (1973, p.28) não ignora a irrecusável interdependência entre as várias faces da linguagem, mas não pode fugir à evidência de que a Linguística, se se pretende autônoma (diferente da Psicologia, da Antropologia, da Gramática Normativa, da Filologia), deve abstrair do todo heteróclito, inclassificável e intratável em função de sua parte mais sistemática, a língua:

Por todas essas razões, seria ilusório reunir, sob o mesmo ponto de vista, a língua e a fala. [...] Cumpre escolher entre dois caminhos [língua e fala] impossíveis de trilhar ao mesmo tempo; devem ser seguidos separadamente.

No PLN, parece ocorrer problema semelhante, mas a solução de Saussure (1973) talvez não represente a melhor alternativa.

Se admitirmos que o objetivo último do PLN é desenvolver sistemas especialistas que repliquem habilidades linguísticas determinadas (como a de traduzir, a de revisar textos, a de produzir resumos etc.), seu objeto de investigação seria exatamente cada uma dessas habilidades que se pretende sintetizar: a tradução, a revisão, a sumarização e assim por diante². Embora todas essas habilidades sejam obviamente linguísticas, no sentido de envolverem conhecimento da linguagem, é importante salientar que: a) não são apenas linguísticas; b) não envolvem apenas o conhecimento, mas também o uso da linguagem.

O caráter extralinguístico das habilidades linguísticas pode ser particularmente ilustrado pela atividade de tradução, em que intervêm, de forma determinante, o contexto e os propósitos da tradução, sem o que, por exemplo, não fariam sentido as distinções normalmente estabelecidas entre “tradução literal” e “tradução

² A legitimidade desse repertório de “habilidades linguísticas” não será aqui posta em discussão, embora talvez merecesse uma consideração mais detalhada: será possível efetivamente pensar nessa especialização de habilidades linguísticas, diferenciando-as em “tradução”, “revisão”, “sumarização” e tantas outras, ou constituirão todas elas diferentes faces de uma mesma competência linguística subjacente e que não caberia segmentar? Acompanho aqui o senso comum, minha intuição de falante e minha experiência como professor, que tem percebido, com frequência, que há de fato diferenças (linguísticas?) importantes em cada uma dessas atividades. No entanto, devo confessar que escapa ao objeto da discussão tentar declinar quais seriam as habilidades linguísticas e investigar se estariam todas elas efetivamente, e da mesma forma, relacionadas a uma competência linguística mais geral. O fato é que, ainda que as diferenças entre elas possam ser de ordem apenas cosmiética e accidental, tem sido observada uma divisão social do trabalho linguístico que, principalmente por meio de políticas de formação e de remuneração, tem confirmado a pertinência e a possibilidade dessa subclassificação.

livre”³. Da mesma forma, a habilidade linguística de tradução não supõe apenas o conhecimento passivo de duas línguas naturais, mas o movimento ativo de inter-relacionamento de estruturas linguísticas. Se pressupusermos que não haverá duas línguas completamente homólogas, ainda que historicamente aparentadas, seremos forçados a admitir que traduzir é, principalmente, operar escolhas léxicas e sintáticas que, em última análise, serão principalmente voluntárias.

Na medida em que considera, não apenas a linguagem, mas também o contexto em que é produzida, o mundo que é por ela falado (ou que é por meio dela constituído), e o uso que dela se faz, o PLN se vê novamente diante de um objeto multiforme e heteróclito, que conviria ser simplificado, sob o risco de permanecer incognoscível.

Consideremos, a título de ilustração, o desenvolvimento de um sistema especialista de Tradução Automática, cujo objetivo seria produzir em português um enunciado (1) *A neve é branca*, correspondente a (2) *Snow is white*, originariamente produzido em inglês. O objeto sobre o que nos debruçamos, o processo de transformação de (2) para (1), como a palavra *nu*, analisada por Saussure (1973), comporta múltiplas perspectivas de análise, o que faz que o fenômeno linguístico em questão – e admitamos que se trate de um fenômeno linguístico o que se está aqui observando – também se apresente dividido em faces contraditórias, a mais saliente das quais talvez seja o fato de que o processo de transformação é um processo de escrita (geração) do português mas é, ao mesmo tempo, um processo de leitura (análise) do inglês.

O dilema, novamente, se reinstalaria: o objeto do PLN nunca se nos ofereceria de forma integral. E a solução para essa nova dificuldade não poderia ser, agora, colocarmo-nos primeiramente no terreno da língua: trata-se de duas línguas diferentes (o inglês e o português) que devem ser consideradas simultaneamente. O objeto de investigação seria, na verdade, as relações que se estabelecem entre as duas línguas comparadas.

³ Num texto clássico, “On Linguistic Aspects of Translation”, Roman Jakobson (2000) deixa claro que a) há mais de um aspecto linguístico envolvido na tradução (o que é, aliás, confirmado pelo próprio título do texto) e b) há mais na tradução do que simplesmente Linguística. O mesmo ocorre em relação a Nida (1964), em que a tradução é concebida não apenas como equivalência formal (fidelidade à mensagem, i.e., à forma e à matéria do texto de partida), mas – talvez principalmente – como equivalência dinâmica (fidelidade à resposta do receptor do texto de partida). O autor considera, claramente, além dos aspectos linguísticos da tradução, os elementos contextuais e culturais, que seriam, em sua opinião, mais determinantes que os próprios elementos linguísticos: “*differences between cultures cause many more severe complications for the translator than do differences in language structure*” (NIDA, 2000, p.130). Mounin (1975, p.26-27), por fim, é categórico: “A tradução (sobretudo nas áreas do teatro, do cinema, da interpretação) comporta sem dúvida aspectos francamente não-linguísticos, extralinguísticos”. O autor afirma, com seus próprios grifos, que “os problemas teóricos suscitados pela legitimidade ou ilegitimidade da operação de traduzir, e por sua possibilidade ou impossibilidade, só podem ser esclarecidos em primeiro lugar no quadro da ciência linguística”. Mas é importante observar que “em primeiro lugar” não significa “exclusivamente” e que exerce, na sentença, o duplo papel de: a) afirmar a importância dos aspectos linguísticos da tradução; b) indicar que há outros aspectos a serem considerados.

Essas duas línguas compartilham, evidentemente, os mesmos atributos: ambas constituiriam sistemas de signos, entidades psíquicas de duas faces (o significante e o significado), vinculadas uma a outra por laços arbitrários, e que se organizariam por sobre o eixo das simultaneidades (a partir do qual poderiam ser observadas, entre os signos, relações associativas e sintagmáticas) e por sobre o eixo das sucessividades. No entanto, se considerarmos que “na língua só existem diferenças” (SAUSSURE, 1973, p.139, p.133, p.141 e p.131); que a língua é “um sistema em que todos os termos são solidários e o valor de um resulta tão somente da presença simultânea de outros”; que “o que distingue um signo é tudo o que o constitui”; que a língua, enfim, é “uma forma, não uma substância”, seremos forçados a admitir que a única matéria efetivamente comparável entre duas línguas (ou dois estados de língua) seriam as relações negativas que opõem, dentro de cada uma delas, seus signos.

Trata-se de semelhanças que não poderiam ser expressas por meio do recurso consagrado a um estado comum das duas línguas cotejadas, associado à confecção de um dicionário bilíngue e de uma gramática contrastiva. A associação – eminentemente substantiva – de um signo linguístico do português a um signo linguístico do inglês é espúria e carente de fundamento, porque, em cada uma das duas línguas, os dois signos associados (positivamente) estabelecem, com os outros signos do mesmo sistema, outras relações, e assumem, por isso, cada um deles, diferentes valores. É importante assinalar que, historicamente, a Linguística, que “*tem por único e verdadeiro objetivo a língua considerada em si mesma e por si mesma.*” (SAUSSURE, 1973, p.271, grifo do autor), surge exatamente da recusa do método comparativo empreendido pelos neogramáticos, que buscavam encontrar, nas línguas românicas e germânicas, substâncias comuns, sem observar que o efetivo parentesco, para além de histórico, que as unia a todas, vinculava o conjunto de relações entre os signos que as definia, cada uma delas, e cada uma à sua maneira⁴. Desta forma, a topologia das duas línguas será principalmente dessemelhante, impedindo que seja observada a biunivocidade (ainda que parcial) entre os vocabulários e a intercambialidade entre construções sintáticas.

O que torna possível a tradução do inglês para o português não pode ser, portanto, a isotopia entre os dois sistemas semiológicos, mas exatamente o movimento pelo qual o caráter endovalorativo e autológico de cada uma dessas estruturas é exovalorizado e heterologicizado. Um movimento que, obviamente, não concerne à *langue*, mas à *parole*. Por esse motivo, uma descrição que privilegie apenas o aspecto social da linguagem, sem a consideração dos movimentos

⁴ “[...] acredita-se, quando se trata da técnica interna de comparação das línguas estar lidando com a própria língua, mas isto é ilusão. Não é a própria língua, mas seus *disiecta membra*, que não permitem apreender a totalidade que é a língua; um tal método alcança as contribuições físicas e fisiológicas, psicológicas e lógicas, sociológicas e históricas, mas não a própria língua.” (HJELMSLEV, 1978, p. 180-181).

realizados pelos falantes, se necessária, não será suficiente para a reprodução do comportamento esperado. Em algum momento, outro tipo de informação, de natureza não estritamente interna, deverá ser mobilizado e incorporado ao sistema⁵.

Percebe-se, nesse caso, que o objeto de investigação não é (apenas) o estado da língua, mas um evento linguístico específico, a língua posta em funcionamento. A linguagem, aqui, interessa principalmente como ato, não unicamente como potência. Mais do que o conhecimento da linguagem, importa, no exemplo considerado, o seu uso. Mais do que a estrutura linguística, importa ali o acontecimento, o fato linguístico.

O objeto do PLN não pode ser, pois, (apenas) a *langue*, que “não constitui [...] uma função do falante”, na qual a reflexão intervém “somente para a atividade de classificação” (SAUSSURE, 1973, p.22). O objeto do PLN deve ser (também) a *parole*, “o ato individual de vontade e inteligência, no qual convém distinguir: 1) as combinações pelas quais o falante realiza o código da língua no propósito de exprimir seu pensamento pessoal; 2) o mecanismo psicofísico que lhe permite exteriorizar essas combinações.”

A questão que se instala é, pois, se a língua é realmente “necessária para que a fala seja inteligível e produza todos os seus efeitos” (SAUSSURE, 1973, p.27). Se o PLN passa a ter, por objeto, “a) combinações individuais, dependentes da vontade dos que falam; b) atos de fonação igualmente voluntários, necessários para a execução dessas combinações” (SAUSSURE, 1973, p.28), resta saber se, assim como foi possível desenvolver uma teoria da língua sem que se fizesse referência à fala, se poderia igualmente postular uma teoria da fala que não fizesse referência à língua.

⁵ Considere-se, a este propósito, apenas um exemplo: a tradução, para o português, da sentença do inglês (1) *The police refused the students a permit because they feared violence*, de forma que a tradução portuguesa da sentença corresponda a (2) *A polícia recusou uma autorização para os estudantes porque e_i temia violência*. Instala-se um problema se se admite que (1) é ambígua do ponto de vista sintático e/ou semântico, e não ambígua do ponto de vista pragmático. Ou seja, se se admite que, embora sejam possíveis duas indexações para o pronome *they*:

(1a) *The police_i refused the students a permit because they_j feared violence.*

(1b) *The police refused the students_i a permit because they_j feared violence.*

apenas a primeira (1a) é autorizada no contexto do estado “normal” das coisas no mundo, em que (a) a polícia (inglesa) tem o poder de conceder ou negar autorizações para manifestações públicas (como passeatas estudantis, por exemplo); (b) os estudantes (na Inglaterra) são obrigados a solicitar permissão à polícia para a realização de manifestações; (c) as manifestações estudantis não raro (?) acabam em protestos violentos; e (d) a polícia, como agente estatal de manutenção da ordem, não deve medir esforços para a prevenção da violência. É importante observar que, no caso, para a resolução da anáfora endofórica representada por *they*, não apenas não são suficientes as relações léxico-gramaticais (a desambiguação envolve, inapelavelmente, a representação, para a máquina, de a+b+c+d indicados acima), mas talvez essas relações não sejam nem mesmo necessárias, já que a indexação refere-se antes ao compartilhamento de tópico (tema) do que ao compartilhamento do número gramatical, que é diferente para o pronome (plural) e seu antecedente (singular).

Do método do PLN

Tomar (também) a fala como objeto de investigação do PLN implica redefinir, em larga medida, o referencial teórico que vinha sendo até aqui considerado: não existe, na matriz saussuriana, uma teoria da *parole*:

Pode-se, a rigor, conservar o nome da Linguística para cada uma dessas duas disciplinas e falar duma Linguística da fala. Será, porém, necessário não confundi-la com a Linguística propriamente dita, **aquela cujo único objeto é a língua. Unicamente desta última é que cuidaremos**, e se por acaso, no decurso de nossas demonstrações, pedirmos luzes ao estudo da fala, esforçar-nos-emos para jamais transpor os limites que separam os dois domínios. (SAUSSURE, 1973, p.28, grifo nosso).

A matéria do PLN seria, portanto, mais afeita àquilo que Saussure (1973) entendia por Linguística da Fala, e que aqui associarei à Psicolinguística, embora a Psicolinguística não represente, no caso, senão parte de uma Linguística (externa) da Língua.

A Psicolinguística pode ser definida, grosseiramente, como um domínio ponte entre a Linguística, cujos conceitos ressignifica, e a Psicologia, cujos procedimentos refuncionaliza. Trata-se de uma área do conhecimento que, principalmente a partir da década de 1950, procura programaticamente realizar a síntese entre fenômenos linguísticos e fenômenos psicológicos.

A possibilidade dessa interdisciplinaridade entre Linguística e Psicologia emerge, no contexto do desenvolvimento das chamadas Ciências da Cognição, de um reposicionamento epistemológico que postula que a linguagem, mais do que um sistema exterior e anterior ao falante, é uma sua faculdade mental, é uma sua competência, em parte congênita, determinada (ou suportada) por uma estrutura biológica específica, de natureza neuroanatomofisiológica, que projeta uma organização psicológica universal, compartilhada, inevitavelmente, por todos os falantes humanos. Recusa-se, portanto, a autonomia da língua, e afirma-se que as estruturas linguísticas, em lugar de funcionarem à revelia do falante, só existem na medida em que são por ele materializadas e dele derivadas; que as categorias linguísticas, longe de representarem uma nomenclatura intersubjetiva da realidade, longe de servirem à simples transmissão do conhecimento, são, elas mesmas, parte desse conhecimento – na verdade, o resultado da própria organização psicológica do falante; que a linguagem, enfim, é sobretudo ação.

As implicações dessa concepção dinâmica da linguagem são principalmente duas: o privilegiamento, como objeto de estudo, das restrições de desempenho do falante, as quais invariavelmente condicionam a produção e a compreensão

dos enunciados linguísticos; e a rejeição da abordagem despsicologizadora dos fenômenos da linguagem empreendida pela definição (estática) de língua como sistema exógeno – rejeição que, em última instância, desautoriza a exclusão da *parole* do objeto de investigação da Linguística tradicional.

Esse redirecionamento epistemológico é resultado, principalmente, dos problemas relacionados à aquisição da linguagem. A solução (neo)behaviorista – mais afeita à concepção de linguagem como uma habilidade que se adquiriria socialmente, por meio de condicionamento operante, sem recurso a processos mentais internos – postulava que o comportamento verbal do falante adulto seria produto histórico de estímulos ambientais (externos, portanto) a que teria sido exposto desde a infância. No entanto vários dos fatos observados na fala de crianças durante o processo de aquisição de linguagem parecem não consistir, pelo menos não diretamente, de respostas a estímulos externos. O modelo não explica, por exemplo, como as crianças projetam qualitativa e quantitativamente os dados primários a que têm acesso (necessariamente finitos e relativamente simples) para produzir e compreender o conjunto infinito de expressões da língua. Assim como também não explica como, a partir de informação unicamente positiva, a criança desenvolve toda uma série de conhecimentos negativos sobre a língua (os chamados juízos de aceitabilidade).

O argumento da pobreza dos estímulos primários – como o define Chomsky (1965) – inviabilizaria, assim, uma descrição exclusivamente empirista do fenômeno da aquisição da linguagem e acusaria a existência de outros mecanismos mentais inatos, mais sofisticados do que os princípios indutivos simples repertoriados pelo comportamentalismo: se a aprendizagem de uma língua não se resume à repetição e à memorização, se não existe uma relação direta entre os dados de entrada e os dados de saída, a linguagem já não pode ser definida como produto de um condicionamento operante: os processos mentais internos tornam-se ativos e essenciais.

A ruptura com a tradição empirista e, por extensão, a recusa de um determinismo *tout-court* entre linguagem e sociedade (postulado, por exemplo, pela Sociolinguística), fazem reintroduzir os fenômenos linguísticos na discussão sobre o funcionamento interno da mente, velho problema filosófico que, a partir da incorporação da metodologia experimental, acabará recebendo novos contornos e novas cores. As relações entre linguagem e pensamento, na esteira dessa nova orientação teórica, passam novamente a ocupar um foco privilegiado de atenção nos estudos da cognição humana.

No entanto, a interdisciplinaridade que então se pronuncia não é propriamente equipolente: a Psicologia acaba muitas vezes importando acriticamente os conceitos da Linguística, sem se preocupar muito com sua realidade (e pertinência) psicológica. Ainda que muito frequentemente ocorram, no âmbito

da Psicolinguística, ressemantizações de categorias linguísticas, a ontologia dos elementos da linguagem não é exatamente o móvel e a preocupação central dos psicolinguistas, que herdaram muitos dos pressupostos das abordagens linguísticas tradicionais.

Um exemplo notável desta situação é o que Cláudia De Lemos (1982, p.98) referia como o pecado original das teorias sobre aquisição de linguagem:

Há, porém, a meu ver, na área de aquisição de linguagem, um dilema de base – seu dilema ou “pecado” original – que, por não ter sido até agora plenamente reconhecido, poucas destas questões ou riquezas têm proiciado. Trata-se da incompatibilidade entre os dois compromissos que o psicolinguista que se dispõe a investigar como as crianças adquirem sua primeira língua, tem que assumir – ou julga ter que assumir – para dar conta de sua tarefa.

Um desses compromissos é com a diacronia, a saber, com a identificação e a explicação das mudanças qualitativas que definiriam o processo de aquisição de linguagem, ou, em outras palavras, seu compromisso com a gênese das estruturas e categorias. O segundo compromisso parece ser o que ele assume com a sincronia e pelo qual se obriga a descrever, em termos de categorias e estruturas definidas no interior das teorias linguísticas vigentes, os enunciados representativos de cada momento do período que isola como objeto de estudo.

Na prática da pesquisa psicolinguística este dilema se concretiza na impossibilidade do investigador ser fiel a esses dois compromissos. Não há dúvida, porém, de que é pelo cumprimento do segundo que ele tem optado.

Embora recuse a perspectiva (estática) de língua, a Psicolinguística não prescinde, portanto, paradoxalmente, dos conceitos da Linguística, que frequentemente toma por primitivos. Embora se constitua como novo domínio, voltado para nova classe de problemas, a partir de novos princípios teóricos e metodológicos, a Psicolinguística importa, com os conceitos e as categorias da Linguística, toda uma concepção de linguagem que, em última análise, ainda não é psicológica. A rigor, não é incomum que a Psicolinguística seja encarada apenas como instância de teste de teorias linguísticas.

O exemplo da Psicolinguística é particularmente simbólico porque os pontos de coincidência com o PLN são flagrantes. Historicamente constituído por cientistas da computação que passam a se preocupar com a formalização da linguagem – bem mais do que por linguistas que procuram testar computacionalmente os formalismos desenhados –, o PLN tem frequentemente pecado por sua excessiva porosidade: quero crer que o domínio, como no caso da Psicolinguística, padece de um comportamento esponjoso e exageradamente permeável, que recolhe e mantém, de forma muitas vezes acrítica, os conceitos da Linguística teórica,

como se possuíssem validade intrínseca, que caberia apenas confirmar. Parece ser ainda ponto de vista hegemônico, no PLN, que as descrições da linguagem, se válidas para o homem, também o seriam para os computadores. Se conceitos como “morfema lexical”, “sintagma verbal” e “sinonímia” revelam-se, em algum momento, úteis para a descrição do conhecimento que os falantes têm da língua, então úteis também seriam para a máquina e ali deveriam ser introduzidos e representados.

Ocorre, porém, que os homens têm se revelado bem mais inexatos do que querem supor os matemáticos, e as semelhanças entre mente e máquina não podem ser admitidas senão em relação às similaridades de *input* e *output*. À luz do paradigma das ciências humanas – indisputavelmente desvinculado dos métodos das ciências naturais, ainda que não de seus objetivos –, as descrições linguísticas, via de regra, não alimentam qualquer preocupação formal com sua implementação computacional, tendo sido desenhadas para que os homens, e não as máquinas, entendessem e representassem o que sabem sobre a linguagem. Em se confirmando a não equivalência estrutural entre mente e máquina, torna-se extremamente frágil a pertinência (e mesmo a utilidade) para o PLN de boa parte dos constructos da Linguística teórica.

Em parte, poder-se-ia dizer, porque a linguagem humana não é lógica e, conseqüentemente, pouco passível de ser replicada a partir de modelos artificiais invariavelmente amparados na ideia (binária) do tudo ou nada. A univocidade, limitação física pressuposta por qualquer implementação computacional da língua, impediria o tratamento da irrecusável equivocidade emergente dos dispositivos linguísticos, reduzidos, na máquina, à mera caricatura do real.

No entanto, a confortável recusa da possibilidade de mecanização da linguagem é tão legítima quanto estéril. Se é indisputável que haverá, sempre, um intervalo intransponível entre homem e máquina, é também indiscutível que inúmeros dos comportamentos humanos ditos “inteligentes” têm sido satisfatoriamente emulados – embora não mimetizados – pela prática computacional.

Obviamente, há larga distância entre os resultados já alcançados e o processamento de alto nível da linguagem. A complexidade, que já não é pequena em níveis mais elementares, como a revisão ortográfica, revela-se inextricável quando se pensa, por exemplo, na construção de sumarizadores e tradutores automáticos. O que não pode significar que a máquina seja intrinsecamente incapaz de processar a linguagem humana. Quero crer que a frustração da agenda da Tradução Automática e, em certa medida, também do PLN, aponta para a impossibilidade, que será apenas procedimental, de a máquina falar (ou pensar) tal qual fazem os homens; não compreende, porém, sua capacidade de atingir alguns dos mesmos resultados de forma completamente diferente da realizada pelos humanos.

Como produto desta perspectiva – que será ainda uma profissão de fé, mais do que uma hipótese que tenha sido empiricamente (com)provada –, este texto deve propor uma fronteira semântica imaginária entre duas diferentes classes verbais do português: de um lado, “emular”, “sintetizar”, “replicar”; de outro, “copiar”, “imitar”, “mimetizar”, “simular”. No primeiro caso, o compromisso único com a produção dos mesmos resultados; no segundo, a fidelidade também ao processo. Em ambas as classes, a adesão ao paradigma do dualismo psicofísico, pedra angular da Inteligência Artificial: o pensamento humano – porque independente de sua extensão material, o corpo – poderia ser imitado ou sintetizado pela máquina.

É interessante ressaltar que a metáfora pouco tem a ver aqui com a concepção mecanicista da mente que inaugura a Inteligência Artificial. Não se trata, absolutamente, da mente como máquina, do cérebro como *hardware*, mas da mente e da máquina como meios diferentes para se atingirem os mesmos resultados. A ausência de qualquer especularidade entre a arquitetura da mente humana e a do computador, e a incomensurabilidade dos meios de expressão – bioquímico, de um lado; físico, de outro –, repelem, já de início, a possibilidade de qualquer paralelismo e aproximação entre desempenho humano e desempenho computacional.

É preciso, pois, pensar a formalização de um modelo de processamento não supervisionado da linguagem capaz de fundar suas próprias orientações de pesquisa e universos categoriais. Interessa não dotar a máquina de um aparato teórico pronto e acabado, mas investigar a possibilidade de investi-la de uma autonomia que a torne capaz de replicar o dinamismo e a instabilidade que são próprios da linguagem. Em última instância, interessa pesquisar a formulação de estratégias computacionalmente implementáveis de aquisição (e não de representação) da linguagem.

Será preciso, então, discutir a aplicabilidade dos conceitos e métodos da Linguística teórica para o desenvolvimento de processadores automáticos de línguas naturais. É a hipótese que se espreita é justamente a da negação de que os instrumentos normalmente desenvolvidos pelos linguistas para entender o funcionamento da linguagem possam ser eficientemente tra(ns)duzidos para a máquina. Eles seriam pertinentes apenas a um paradigma absolutamente alheio à teoria computacional. Não seria autorizado, portanto, o movimento epistemológico que leva cientistas da computação a querer buscar, na Linguística, modelos de representação da linguagem. O deslocamento necessário seria outro: as máquinas redescrivendo a linguagem a partir de suas próprias experiências linguísticas.

A justificativa desta abordagem se constitui em torno desta aparentemente absurda possibilidade que se desdobra em uma plethora de outras interrogações a serem adicionadas àquelas que inauguram este texto: será possível à máquina

adquirir a linguagem humana? Será possível uma descrição não humana para a linguagem? Será possível a criação de uma metalinguagem diferente da metalinguagem desenvolvida pelo homem? Será possível à máquina criar, sozinha, alguma metalinguagem? Será possível um conhecimento da realidade não intermediado pelo homem mas por uma sua criação? Até que ponto convergem as descrições, se possíveis, do criador e da criatura?

Acredita-se, aqui, em que o aprofundamento dessas interrogações, ainda que não sejam alcançadas respostas, pode ser extremamente elucidativo dos fenômenos da Linguística e da linguagem. A possibilidade de uma descrição não antropológica da linguagem (e de um tratamento não antropomórfico para a máquina), funda uma ótica particularmente interessante (se não ininterpretável) de abordagem dos fenômenos linguísticos. Ainda que não se possam jamais comparar os resultados da Linguística com os resultados do PLN, é forçoso considerar que apenas a possibilidade de uma descrição não mentalista (como será inapelavelmente o caso da descrição computacional) para os fenômenos da linguagem já é, por si mesma, merecedora de atenção.

MARTINS, R. The original sin of Computational Linguistics. *Alfa*, Araraquara, v.55, n.1, p.287-307, 2011.

- *ABSTRACT: This paper explores some of the epistemological issues concerning the scope of Computational Linguistics, more specifically the scope of Natural Language Processing (by computers) when conceived of as the area of study where Traditional Linguistics and Computer Sciences overlap. By discussing the nature of this overlap, it concludes that the goals, objects and methods of Computational Linguistics do not match those of Traditional Linguistics, and that the former would not constitute a subdomain or application area of the latter, but a new scientific exploration field, where, although Traditional Linguistics research practices are still dominant, its internal contradictions gradually would lead to a new linguistic "reason" that would postulate less anthropocentric and anthropomorphic categories of analysis and description of languages, which would prove to be more suitable for the human linguistic behavior emulation by computers.*
- *KEYWORDS: Computational Linguistics. Natural Language Processing. Machine Translation.*

REFERÊNCIAS

CARNAP, R. *The logical syntax of language*. London: K. Paul Trench, 1937.

_____. *Physikalische begriffsbildung*. Karlsruhe: Braun, 1926.

CHOMSKY, N. *Aspects of the theory of syntax*. Cambridge: MIT Press, 1965.

DE LEMOS, C. T. G. Sobre a aquisição da linguagem e seu dilema (pecado) original. *Boletim da Abralin*, Recife, n.3, p.97-136, 1982.

GRISHMAN, R. *Computational linguistics: an introduction*. Cambridge: Cambridge University Press, 1986.

HJELMSLEV, L. T. *Prolegômenos a uma teoria da linguagem*. São Paulo: Perspectiva, 1978.

HEMPEL, C. G. Aspects of scientific explanation. In: _____. *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press, 1970.

HUTCHINS, J. From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology. *Machine Translation*, Alemanha, v.12, n.3, p.195-252, 1997.

JAKOBSON, R. On linguistic aspects of translation. In: VENUTI, L. (Ed.). *The translation studies reader*. London: Routledge, 2000. p.113-118.

KAY, M. Machine translation: the disappointing past and present. In: COLE, R.A. et al. (Ed.). *Survey of the state of the art in human language technology*. Oregon: NSF/CEC/CSLU; Oregon Graduate Institute, nov. 1995. Disponível em: <<http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>>. Acesso em: 4 set. 2009.

KUHN, T. S. *A estrutura das revoluções científicas*. São Paulo: Perspectiva, 1992.

MILNER, J. C. *O amor da língua*. Porto Alegre: Artes Médicas, 1987.

MOUNIN, G. *Os problemas teóricos da tradução*. São Paulo: Cultrix, 1975.

NIDA, E. Principles of Correspondence. In: VENUTI, L. (Ed.). *The translation studies reader*. London, New York: Routledge, 2000. p.126-140.

_____. *Toward a science of translating, with special reference to principles and procedures involved in bible translating*. Leiden: E. J. Brill, 1964.

NIST. *Open machine translation evaluation (MT08): official evaluation results*. 2008. Disponível em: <http://www.itl.nist.gov/iad/mig//tests/mt/2008/doc/mt08_official_results_v0.html>. Acesso em: 13 set. 2010.

POPPER, K. *Conjectures and feutations: the growth of scientific knowledge*. London: Routledge, 1963.

SAUSSURE, F. de. *Curso de linguística geral*. Tradução de Antônio Chelini, José Paulo Paes e Izidoro Blikstein. São Paulo: Cultrix, 1973.

STAROSTA, S. Natural language parsing and linguistic theories: can the marriage be saved? *Studies in Language*, Amsterdam, n.15, p.175-197, 1991.

Recebido em setembro de 2010.

Aprovado em fevereiro de 2011.

