

Milena Uzeda Garrão  
migarrao@terra.com.br

# Linguística de corpus: o lugar da fusão entre semântica e pragmática

## Corpus linguistics: where semantics and pragmatics are blended together

**RESUMO** - Este artigo apresenta a importância da faceta empírica do fenômeno linguístico, algo bastante discutido em virtude da crescente utilização da Linguística de Corpus como instrumento teórico. Traz algumas breves considerações filosóficas sobre essa linha de pensamento; apresenta o seu desdobramento dentro da Linguística e mais especificamente no domínio de Processamento de Linguagem Natural. Discute também a viabilidade e a acuidade de uma visão eminentemente pragmática para lidar com o fenômeno linguístico. Finalmente, conclui ser essa perspectiva a mais profícua para uma agenda de estudos da linguagem com objetivos aplicados (como por exemplo, a tradução e ensino de português como segunda língua) e para a lexicografia.

**Palavras-chave:** teorias linguísticas, empirismo, PLN, linguística de *corpus*.

**ABSTRACT** - This paper presents the importance of dealing with linguistic phenomenon from an empirical perspective, a rather controversial issue due to the increasingly use of Corpus Linguistics as a theoretical tool. It makes a brief philosophical account on this theoretical framework; then, it presents its development within Linguistics and, more precisely, within Natural Language Processing. It also discusses the feasibility and reliability of the conventionalist framework on linguistic phenomenon. Finally, it reasons that this perspective is the most eloquent one for coping with an agenda of language studies focused on applications (such as Translation and Portuguese as a Second Language) and lexicography.

**Key words:** linguistic theories, empirical accounts, NLP, corpus linguistics.

Valho-me deste artigo para demonstrar como o avanço tecnológico e, mais especificamente, como a utilização de dados linguísticos digitalizados em forma de corpus têm um papel determinante na descrição linguística de uma língua natural. Início minha argumentação trilhando um caminho que já havia sido sinalizado pelas considerações sofistas da Grécia antiga. Trata-se de um olhar sobre o fenômeno linguístico que privilegia o seu estatuto convencional. Dispensa, em contrapartida, qualquer tipo de teor simbólico ou representativo da linguagem em relação à realidade ou à mente.

Reservo, portanto, a primeira seção do artigo para breves considerações filosóficas sobre essa linha de pensamento; apresento na segunda seção o seu encaminhamento no âmbito linguístico e no âmbito de Processamento de Linguagem Natural (PLN, doravante); já na terceira seção, contraste esse olhar àqueles mais notoriamente empregados para lidar com o fenômeno linguístico, apresentando a vantagem ou, talvez seja melhor dizer, a conveniência desse ponto de vista se comparado a essas pers-

pectivas mais hegemônicas no domínio teórico dos Estudos da Linguagem. Finalmente, apresento os sub-domínios dos Estudos Linguísticos que se beneficiam mais diretamente dessa perspectiva teórica.

### Os primórdios do Pragmatismo Linguístico

Na história do pensamento filosófico, a visão sofista sobre o discurso é talvez a primeira manifestação documentada dessa concepção de linguagem. Muito do que sabemos sobre o pensamento sofista em relação à linguagem deve-se aos escritos platônicos<sup>1</sup>. O que, de certa forma, é sintomático, uma vez que Platão se opunha a uma visão convencionalista da linguagem.

De uma forma geral, os registros de Platão apresentam as considerações sofistas — em que a linguagem é vista à luz da célebre máxima de Protágoras, “o homem é a medida de todas as coisas” — como contraponto ao pensamento socrático. Este formato dialógico apresenta, de certa forma, um teor didático para reafirmação de sua

<sup>1</sup> O *Sofista* e *Crátilo* são exemplos de diálogos em que Platão tematiza a linguagem.

crença representacionista. Em outras palavras, o fato de Platão sustentar, em oposição violenta ao ideário sofista, a existência de verdades únicas e fixas, em alguma medida pode ter minimizado ou desqualificado a importância e teor das considerações sofistas ao longo da história do pensamento filosófico (cf. Souza Filho, 1997; Martins, 2003). Logo, esse investimento de Platão contra os sofistas deve ter contribuído para minimizar e desqualificar também o ângulo pragmático da linguagem.

Dentre as características privilegiadas por essa visão de linguagem está a mutabilidade do significado de uma expressão em virtude de esta depender das práticas humanas, como também a incapacidade de a expressão representar algo exterior à linguagem.

Esse viés eminentemente pragmático sobre a linguagem é retomado já na história mais recente da filosofia<sup>2</sup>, notadamente, por Ludwig Wittgenstein<sup>3</sup>. De forma ostensivamente resumida, pode-se dizer que o filósofo austríaco se recusa a enxergar o estudo da linguagem dentro dos mesmos moldes metafísicos propostos para as ciências, principalmente porque o “cientista”, neste caso, é uma das peças envolvidas no que chama de “jogos de linguagem” (cf. Wittgenstein, 1979). Ele desenvolve este conceito através de uma analogia com a idéia de “jogo”, e aponta certas características e semelhanças com a linguagem: jogos possuem regras, são práticas compartilhadas por uma comunidade, possuem peças, são autônomos, não requerem justificativas.

Sua proposta é a de que os jogos de linguagem “são a totalidade da linguagem e das atividades com as quais está interligada” (Wittgenstein, 1979, §§7 e 23). Os jogos de linguagem seriam atividades autônomas que prescindem de explicação; “são parte de nossa história natural, assim como andar, comer, jogar, etc.” (Wittgenstein, 1979, §25). E o fato de o próprio homem estar encarcerado nos rituais lingüísticos que regem toda e qualquer manifestação lingüística compromete o seu julgamento teórico ou sua tentativa de explicação acerca do assunto.

O estudo sobre a linguagem, portanto, seria ele mesmo mais um desses jogos. E como o significado de uma palavra encontra-se na execução da linguagem, não há nada a ser provado ou justificado. Não se pode explicar o que está explícito. A significação seria, então, a habilidade em lidar com as palavras na linguagem. E a explicação, o resultado da nossa “ânsia de generalidade”. Ela pode até ser utilizada, mas não deveria ser entendida como uma meta-regra; ou uma regra fora do jogo. Toda explicação é interessada (ver Scherer, 2002, sobre o conceito de *explicação*).

Sua visão nega, deste modo, a vocação representacionista da linguagem tão defendida por filósofos como Platão, Aristóteles e Locke. Não há, para ele, uma essência do significado que preceda o uso das palavras.

Ou seja, não haveria uma distinção clara entre Semântica e Pragmática. Para Wittgenstein, é somente na prática do uso que o significado se dá. E mesmo assim, ele não poderia ser “coisificado”. Por essa razão, é instável, contingente e fragmentado. Falar sobre a linguagem, ou fazer um relato sobre o relato, seria também um relato do jogo, e também faria parte dele. Portanto, nunca seria definitivo.

Se num primeiro momento sua visão parece amputar a Lingüística, e mais especificamente, eliminar a Semântica, uma vez que sugere uma total impossibilidade de tratamento do fenômeno da significação, concordo com Martins (1999) quando pondera que adotar uma visão wittgensteiniana na Lingüística “não corresponde à negação da possibilidade de qualquer estudo sistemático e empírico das línguas do mundo”. A autora sugere, portanto, que uma Lingüística sob a perspectiva wittgensteiniana “é viável e teria como propósito geral descrever as regularidades — parciais e contingentes — observáveis nos jogos de linguagem que constituem as línguas do mundo” (1999, p. 147). A descrição de uma língua, segundo o filósofo, estará sempre condicionada à parcialidade imposta pelas infundáveis possibilidades de lances dos jogos de linguagem.

## O pragmatismo radical na Lingüística e em PLN

A posição do lingüista Roy Harris (1996), por exemplo, em relação ao funcionamento da língua alinha-se às idéias convencionalistas solidificadas por Wittgenstein. Em Scherer (2002) é possível encontrar uma farta exposição da simpatia de Harris às idéias do filósofo austríaco. Um dos pontos em comum entre as duas posições é a impossibilidade de distinção entre Semântica e Pragmática, ou entre conhecimento lingüístico e enciclopédico.

Note-se que a visão de Pragmática exposta até agora se distancia de muitas abordagens pragmáticas da Lingüística. Estas conduzem suas explicações sobre o “uso da língua” com base na distinção *Semântica-Pragmática*, que com frequência se apóia na distinção *significado literal-significado de uso*; algo indissociável por uma visão wittgensteiniana. Scherer (2002, p. 28) avalia, portanto, que Wittgenstein e Harris compartilham uma visão pragmática radical, em que o uso lingüístico não é um dos componentes da linguagem, mas a única forma produtiva de se pensar os fenômenos lingüísticos.

Adam Kilgarriff, estudioso do léxico do ponto de vista computacional, ou seja, com interesse voltado para PLN, também advoga este pragmatismo radical. Em seu contundente artigo “*I don't believe in word senses*” (Kilgarriff, 2000), ele atribui ao corpus, ou a um conjunto de dados lingüísticos de uma língua, o poder de desambiguação de significados. Demonstra que as pala-

<sup>2</sup> Para uma boa apreciação sobre o assunto ver Martins (1999).

<sup>3</sup> Refiro-me ao segundo Wittgenstein e sua visão pragmática expressa em *Investigações Filosóficas*, 1979.

vras, em suas diversas acepções, estão desatreladas do seu alegado “sentido atômico” (*core meaning*).

Propõe, como alternativa, uma ontologia, ou um conjunto de palavras semanticamente relacionadas, com base em uma convergência (*cluster*) estatística dos seus vários usos em córpus, em detrimento do seu sentido atômico. Os extratos do córpus são agrupados em sentidos específicos de acordo com os objetivos de tarefas específicas. Acredita, portanto, que os significados só existem dependentes de propostas ou tarefas (*tasks*). Ele conclui que sem um fim ou uma tarefa, os sentidos das palavras não existem.

### Mais sobre o dilema Semântica versus Pragmática

Embora proposta antes do advento de córpus computadorizados, a Teoria Contextual do Significado proposta pelo lingüista britânico J. R. Firth (1957), subjaz a uma abordagem a partir de córpus. Ele propõe que o estudo do significado e do contexto devem ser centrais para a lingüística. Recusa-se a admitir qualquer tipo de distinção entre *langue e parole* (estabelecida antes dele por Saussure) ou entre *competência e desempenho* (estabelecida mais tarde por Chomsky), já que não enxerga a língua como uma entidade autônoma. Os *eventos lingüísticos*, nos seus termos, são recorrentes e repetidamente observáveis.

No mesmo ano em que Firth profetizou “*You shall know a word by the company it keeps*”, Chomsky tornou pública, em *Syntactic Structures*, a sua crença de que a língua deveria ser analisada em um nível muito mais profundo do que sua estatística superficial. Isto é, numa crítica à visão empirista, argumentou que uma abordagem com base em dados da língua consideraria como identicamente remotas as sentenças “*Colorless green ideas sleep furiously*” e “*Furiously sleep ideas green colorless*”. Ou seja, em qualquer modelo estatístico, as duas sentenças seriam igualmente excluídas, embora, como falantes, nós saibamos que a primeira obedeceria a algumas regras gramaticais. O que Chomsky defendia é que a primeira não poderia ser totalmente desconsiderada e equiparada à segunda simplesmente pelo fato de sua probabilidade de ocorrência ser remota<sup>4</sup>. Essa crítica, conhecida como *problema dos dados esparsos*, foi altamente influente para uma mudança de perspectiva teórica nos anos 1960.

Mas a evolução de técnicas estatísticas aliada à possibilidade de construção de córpus mais robustos de alguma forma equacionou aquilo que Chomsky utilizava como emblema de sua crítica. Além disso, a forma de Firth (1957) enxergar a língua não se presta à comparação com o modelo chomskyano. Para ele a língua é uma *evento*, uma forma de ação (*a way of “doing things”*); e é por isso que seria um ato legítimo por parte do lingüista se ater aos

eventos discursivos propriamente ditos. Firth acreditava que o evento era um fim em si mesmo e não uma forma de acessar o “verdadeiro” objeto de estudo: o sistema subjacente a ele, como Chomsky advoga.

### A tradição e a separação entre sentido literal e metafórico

Inspirada pelo pensamento filosófico de que a linguagem tem eminentemente uma função representativa, grande parte das teorias lingüísticas também exclui do seu campo de estudo qualquer modalidade de linguagem que não representa de fato “a coisa em si”. Ou seja, para se construir uma ciência da linguagem há de se excluir todo tipo de expressão lingüística que foge à representação da realidade. Chomsky assim como seus legatários edificaram suas teorias neste **pressuposto universalista**. Toda manifestação lingüística não-literal, ou seja, não-representativa, deixa de ter importância do ponto de vista científico. Bastaria, então, traçar o limite entre significado literal e figurado. E é nesta, aparentemente, infundável tentativa de divisão desses dois campos de significação que a maior parte dos estudos semânticos se situa. De fato, isto tem sido, se não a maior, uma das grandes interrogações da Lingüística.

É bem verdade que mesmo os defensores da chamada **visão literalista** admitem que a questão não é trivial: Ruth Kempson (1995, p. 74) pondera que, entre o literal e o metafórico, “há um *número grande* de casos duvidosos”. J. Sadock (1993, p. 42), por sua vez, afirma que os princípios subjacentes à metáfora são psicológicos e não estritamente lingüísticos e que, portanto, devem estar fora do escopo da lingüística sincrônica. Sadock, contudo, admite que “em *inúmeros casos* é difícil determinar onde começa o sentido figurado e onde termina o sentido literal” (1993, p. 48). No entanto, as incontáveis fronteiras nebulosas não parecem ter dado origem a questionamentos do pressuposto teórico, mas sim, como avalia criticamente Martins (1999, p. 55), apenas consideradas como “percalços naturais do fazer científico”.

Uma das grandes conseqüências sintomáticas do conflito literal versus figurado é a inevitável tensão entre homonímia e polissemia. Martins explica que quando há uma resistência para o estabelecimento da fronteira entre literal e figurado é comum entre os literalistas a utilização de duas estratégias: (a) redução do escopo da análise de modo a tentar excluir a variação polissêmica e (b) conversão dos casos de polissemia em casos de homonímia. Tomemos, a título de ilustração dessa tensão, a lista de frases abaixo:

- i) Ele *tirou* a camisa.
- ii) Ele *tirou* a camisa do armário.
- iii) Ele *tirou* 10 na prova.

<sup>4</sup> Ironicamente, a probabilidade de ambas as frases ocorrerem é grande em virtude de sua relevância ilustrativa na Lingüística.

- iv) Ele *tirou* aquela idéia da cabeça.
- v) Ele *tirou* o corpo fora.

Dentro de uma concepção literalista do significado, poderíamos supor que somente as frases i), ii) e iii) seriam consideradas instâncias de fragmentos lingüísticos passíveis de análise lingüística, onde i) e ii) apresentariam significados análogos do verbo *tirar* com transitividades distintas e iii) talvez fosse equacionada através da estratégia descrita em (b); isto é, seria um verbo distinto de i) e ii) apresentando apenas a mesma forma fonológica. Mas o que fazer, então, com iv) e v)? A rigor, a estratégia de homonímia também poderia ser aplicada em iv). Ou seja, embora iv) tenha a mesma estrutura sintática de ii), muito provavelmente deixaria de ser considerada analisável dentro da mesma entrada lexical, uma vez que o uso do verbo não é literal. Isto é, por uma perspectiva tradicional, não é possível tirar “literalmente” uma idéia da cabeça. Em v) teríamos ainda o agravante de haver uma aberração estrutural. Por ser uma frase sintaticamente anômala, tal estrutura só poderia ser analisável por uma perspectiva literalista como uma expressão fixa; com sentido unitário, indivisível. Assim sendo, conclui-se que uma frase corriqueira do PB como *pode tirar o seu cavalinho da chuva*, seria interpretada como a retirada de um mamífero quadrúpede da precipitação atmosférica ou como um segmento indivisível em que tem sentido análogo a *Desista!* Esta última interpretação, entretanto, estaria fora do domínio sintático-semântico e seria atribuída a um campo do estudo lingüístico fronteiro com a sociologia e a psicologia: a pragmática. Seria um exemplo de **conhecimento enciclopédico** do PB.

### Quando a metáfora é a regra

Segundo a visão literalista exposta na seção anterior, portanto, as metáforas são simples expressões lingüísticas fora do escopo da linguagem corriqueira, assim como as expressões idiomáticas (tradicionalmente consideradas um tipo de CMs) têm sentido arbitrário. Lakoff e Johnson (1980), em contrapartida, rebatem esse senso comum através da argumentação de que se a metáfora fosse uma mera expressão lingüística não haveria tanta evidência apontando para um mapeamento entre domínios específicos, o que demonstra uma restrição conceitual para a construção do fenômeno.

Os autores explicam que a construção metafórica seria o mapeamento entre um domínio-fonte e um domínio-alvo, que compartilhariam correspondências ontológicas sistemáticas (A VIDA É UMA JORNADA,

DISCUSSÕES SÃO GUERRAS, O CORPO É UM CONTAINER, dentre outros)<sup>5</sup>. Esses mapeamentos obedeceriam ao Princípio da Invariância, um princípio que caracteriza a classe de analogias possíveis. Trata-se de um mapeamento convencional, isto é, uma parte fixa do nosso sistema conceitual. Tal caracterização também se prestaria a dar conta da polissemia, um outro grande problema dentro de estudos semânticos literalistas.

Para explicar o mapeamento entre domínio-fonte e domínio-alvo é necessário considerar que esses compartilham uma “estrutura de nível geral”. E por ser convencional, a metáfora é usada constantemente e automaticamente, sem esforço ou consciência. Lakoff, de fato, responde a uma pergunta constante dentro de estudos sobre o significado. Por que nossas metáforas são como são? Porque, diria ele, seriam baseadas na experiência. E a experiência motiva a metáfora, não a determina. Por exemplo, o mapeamento conceitual MAIS-EM CIMA/ MENOS-EMBAIXO (que viabiliza expressões como *ter alto astral*, *estar meio cabisbaixo*) prediz que simplesmente será difícil numa língua haver a correspondência “menos-em cima” e que o falante de uma língua que não tenha nenhum desses mapeamentos irá apreender a correspondência MAIS-EM CIMA com muito mais facilidade do que aquela improvável (Lakoff, 1987).

Voltemos aos exemplos expostos na seção anterior:

- i) Ele *tirou* a camisa.
- ii) Ele *tirou* a camisa do armário.
- iii) Ele *tirou* 10 na prova.
- iv) Ele *tirou* aquela idéia da cabeça.
- v) Ele *tirou* o corpo fora.

O mapeamento conceitual O CORPO É UM CONTAINER justificaria o uso do verbo *tirar* em i), iv) e v). Já a explicação para ii) não se pautaria em nenhum mapeamento entre domínio-fonte e domínio-alvo por não se tratar de um exemplo metafórico. Como já mencionado na seção anterior, Lakoff reconhece que há “uma extensa gama de conceitos não metafóricos” (como “*o balão subiu*”). Na verdade, pode-se dizer que quando se trata somente do domínio-fonte, estaríamos diante de conceitos não-metafóricos.

De fato, há uma elegância explanatória nesses exemplos. Mas o que fazer com iii)? Teríamos que conceber *prova* como sendo CONTAINER? Até onde é possível estender essas analogias? Quando deixam de ser válidas? Ou seja, o que fazer com aquela parte da formação de metáfora que foge ao mapeamento; o que fazer com as irregularidades da formação da metáfora? Essas pergun-

<sup>1</sup> Em Lakoff e Johnson (2002), tal abordagem bidomínial cede espaço a uma teoria que envolve quatro domínios — a teoria da mesclagem —, segundo a qual nosso cérebro adquire informação do resto do nosso corpo. A relação do corpo com o mundo estrutura os conceitos que usamos para pensar. Não podemos pensar o que queremos, somente aquilo disponibilizado pelo nosso cérebro. Nessa obra, há um incremento dos mapeamentos, mas o conceito cognitivo lançado nos anos 80 parece persistir.

tas não ficam claramente respondida pelo modelo. O autor argumenta que a metáfora nova se constrói dentro de parâmetros impostos pelo sistema conceitual fixo, e “raramente ocorre independente dele” (1987: 228), mas não esclarece os casos em que isso acontece.

Por uma perspectiva radicalmente pragmática, contudo, algumas dessas grandes questões epistemológicas não são simplesmente deixadas de lado. Apenas não se prestam a uma verificação incontestável como em outros campos do saber. Tal escolha leva a uma perspectiva lingüística amplamente amparada pela faceta estatística do significado. Trata-se de uma forma de enxergar as recorrências lingüísticas abrindo mão de rotulações semânticas pré-concebidas ou de considerações experientialistas.

Essa visão nega, deste modo, a vocação representacionista da linguagem tão defendida por filósofos como Platão, Aristóteles e Locke. Não há, no Pragmatismo radical, uma essência do significado que preceda o uso das palavras. É somente na prática do uso que o significado se dá. É por essa razão que o único caminho legítimo para verificação da prática desse uso é o seu registro exaustivo; em outras palavras, é a lingüística de *corpús*.

### Uma abordagem com base em *corpús* é segura?

Como o leitor já pode supor a esta altura do artigo, a descrição de uma língua, segundo Wittgenstein, estará sempre condicionada à parcialidade imposta pelas infundáveis possibilidades de lances dos jogos de linguagem. A propósito, sobre a idéia de completude da linguagem, Wittgenstein (1979) confronta o leitor com o seguinte questionamento no §18 de *Investigações Filosóficas*:

[...](e com quantas casas ou ruas, uma cidade começa a ser cidade?) Nossa linguagem pode ser considerada como uma velha cidade: uma rede de ruelas e praças, casas novas e velhas, e casas construídas em diferentes épocas; e isto tudo cercado por uma quantidade de novos subúrbios com ruas retas e regulares e com casas uniformes.

Esse parágrafo de *Investigações Filosóficas* é particularmente relevante para legitimar a abordagem com base em *corpús*, uma escolha teórica e também metodológica comumente utilizada na Lingüística atual. Um exemplário da língua sempre pode ser maior, é verdade; mas nunca deixará de ser um extrato válido do seu uso. Portanto, a idéia de abarcar a língua de uma forma totalizante seria uma utopia. Conclui-se, daí, que a parcialidade é inevitável. Ao longo do tempo, algumas expressões são construídas, outras demolidas. Isto é parte do jogo. Todo *corpús* de uma língua seria, portanto, uma caracterização válida da mesma<sup>6</sup>.

Vejamos o caso do estudo lexicográfico para o francês de Verlinde e Selva (2001), que compararam a aborda-

gem tradicional da construção de um dicionário de aprendizes de segunda língua, baseada na intuição do lexicógrafo, a uma abordagem baseada em *corpús*. Os autores constataram que é atribuído ao número de entradas (macroestrutura) de um dicionário uma importância maior do que ao conteúdo de cada entrada (microestrutura). Portanto, a ênfase recai nas palavras simples em detrimento das combinações de palavras, o que vem a ser um contra-senso, já que, para fins comunicativos, os aprendizes de segunda língua necessitam muito mais de informações sobre combinações do que sobre palavras isoladas.

Os pesquisadores puderam detectar, por exemplo, que, mesmo com a difundida repressão do governo francês aos anglicismos, o *corpús* apresentava uma frequência bastante representativa desses estrangeirismos. A partir dessa constatação, podemos atribuir aos dados lingüísticos um poder maior do que geralmente lhes é conferido.

Por outro lado, os autores utilizaram como fontes periódicos jornalísticos, o que, segundo Sinclair (1991), pode ser questionável: “um *corpús* é uma coletânea de um material amplamente homogêneo, mas retirado de fontes diversas de forma que a individualidade de uma fonte se perde, a não ser que o pesquisador queira isolar um texto em particular. [...] A diversidade de fontes é uma garantia de segurança dos dados” (p. 17-18).

Corroborando as preocupações de Sinclair (1991), Verlinde e Selva (2001) destacam que a lexicografia de *corpús* é uma evidência empírica necessária à intuição do pesquisador, que serviria para preencher as lacunas de *corpús* que chamam de “não equilibrados”. Portanto, eles ressaltam a urgência da construção de um extrato de dados equilibrado para o francês e outras línguas.

Pode-se notar, portanto, que grande parte dos lexicógrafos que se dedicam ao estudo de *corpús* compartilham o sentimento de que a compilação de fontes diversas evita o chamado “paradoxo do *corpús*”. Isto é, evita a formação de um *corpús* tendencioso. Contudo, essa caracterização do *corpús* ideal ignora a imagem da “cidade” apresentada por Wittgenstein e revela uma ilusão corrente na Lingüística: a idéia de que é possível fugir do paradoxo do *corpús*.

A questão freqüentemente intocada é a de que a língua em si não é completa; sempre é possível acrescentar mais uma casa ou mesmo uma rua, o que torna um *corpús* um fragmento de algo já fragmentado. Portanto, não há como fugir desse paradoxo uma vez que a completude da língua também é algo inatingível.

Tal constatação, entretanto, não desabona um estudo descritivo da língua; o fato de não ser possível exaurir todas as possibilidades de jogos da linguagem não impede um lingüista ou um lexicógrafo de descrever o que geralmente é constatado nos jogos (Martins, 1999, p. 147). Por outro lado, é desejável que o pesquisador se cerque de

<sup>6</sup> Para uma melhor apreciação das idéias wittgensteinianas ver Glock (1996) e Martins (1999).

alguns critérios para aferir a adequação de um corpus em função dos objetivos pragmáticos que se deseja alcançar.

## Discussão

Note-se que as idéias pragmáticas de teóricos da linguagem como Firth, Kilgarriff e Harris fogem de certa forma do embate entre sentido literal e sentido metafórico, assim como da divisão entre conhecimento lingüístico e conhecimento enciclopédico. Tal escolha de um caminho minimamente comprometido com representação de significado, contudo, não se trata de uma fuga ou uma evitação do confronto entre Semântica e Pragmática, apenas de uma clareza de que tal confronto talvez não levaria a uma resposta definitiva.

De fato, não parece nenhuma heresia afirmar que a língua pode ser descrita como um fenômeno probabilístico, uma vez que há nitidamente padrões de combinações vocabulares recorrentes. Esses padrões, muitas vezes, extrapolam o nível lexical (como expressões idiomáticas mais triviais), culminando em estruturas sintagmáticas super-oracionais (Ver Jackendoff, 1997, cap. 7; Garrão e Dias, 2001, para exemplos em inglês americano e português do Brasil, respectivamente).

De certa forma, esta perspectiva atenua a visão chomskiana da linguagem, focada no poder criativo do falante ideal, e prioriza uma visão de língua inseparável da pragmática; isto é, enfatiza o teor eventivo do fenômeno lingüístico.

Hoje em dia já dispomos de bons recursos estatísticos para detecção de estruturas convencionais da língua. Existem alguns métodos estatísticos disponíveis para a detecção de combinações lingüísticas. Por isso, o avanço tecnológico teve um papel seminal na constatação de que uma visão probabilística da língua é viável e altamente profícua. E o corpus, além de servir como base de dados para identificação de tais convenções, também tem um papel preditivo ao fornecer os ambientes lingüísticos tipicamente relacionados a elas.

Como Wittgenstein define em *Da certeza* (§559), “o jogo de linguagem é, por assim dizer, imprevisível. Queiro dizer: não está fundamentado. Não é racional (ou irracional). Está aí - como a nossa vida”. Portanto, como jogadores, talvez a atitude mais prudente seja a constatação e descrição de partes dos jogos, sem tentar alçar vôos teóricos extremamente ambiciosos e totalizantes. Afinal de contas, o jogo em si já dá o que falar.

## Referências

- CHOMSKY, N. 1957. *Syntactic Structures*. The Hague, Mouton & Co., 120 p.
- FIRTH, J.R. 1957. A Synopsis of Linguistic Theory 1930-1955'. *Studies in Linguistic Analysis*, 51:1-31.
- GARRÃO, M. e DIAS, M.C. 2001. Um estudo de expressões cristalizadas e sua inclusão em um tradutor automático bilíngüe. *Cadernos de Tradução*, VIII:165-182.
- GLOCK, H.J. 1996. *Dicionário Wittgenstein*. Rio de Janeiro, J. Zahar, 400 p.
- HARRIS, R. 1996. *The Language Connection*. Bristol, Thoemmes Press, 213 p.
- JACKENDOFF, R. 1997. *The Architecture of the Language Faculty*. Cambridge, MIT Press, 340 p.
- KEMPSON, R. 1995. *Semantic Theory*. Cambridge, Cambridge University Press, 220 p.
- KILGARRIFF, A. 2000. *I don't believe in word senses*. Disponível em [http://www.kcl.ac.uk/humanities/cch/seminar/99-00/seminar\\_kilgarriff.html](http://www.kcl.ac.uk/humanities/cch/seminar/99-00/seminar_kilgarriff.html). Acesso em: 21/02/2005.
- LAKOFF, G. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago, University of Chicago Press, 642 p.
- LAKOFF, G. e JOHNSON, M. 1980. *Metaphors we live by*. Chicago, University of Chicago Press, 256 p.
- LAKOFF, G. e JOHNSON, M. 2002. *Philosophy in the flesh: the embodied mind and its challenge to Western Thought*. New York, Basic Books, 624 p.
- MARTINS, H. 1999. *Metáfora e Polissemia no estudo das línguas do mundo: uma aproximação não representacionista*. Rio de Janeiro, RJ. Tese de Doutorado inédita. UFRJ, 250 p.
- MARTINS, H. 2003. *Três Caminhos na Filosofia da Linguagem*. In: F. MUSSALIM e A.C. BENTES (orgs.), *Introdução à Lingüística*. Vol. III, São Paulo, Cortez Editora, p. 439-474.
- SADOCK, J. 1993. *Figurative Speech and Linguistics*. In: A. ORTONY (org.), *Metaphor and Thought*. Cambridge, Cambridge University Press, p. 42-57.
- SCHERER, M. 2002. *Uma questão de vocabulário: considerações sobre o campo lexical no ensino de português para estrangeiros*. Rio de Janeiro, RJ. Dissertação de Mestrado inédita. PUC-Rio, 230 p.
- SINCLAIR, J. 1991. *Corpus, Concordance, Collocation*. Oxford University Press, Oxford, 197 p.
- SOUZA FILHO, D. 1997. *Iniciação à história da Filosofia*. Rio de Janeiro, Jorge Zahar, 300 p.
- VERLINDE, S. e SELVA, J. 2001. *Corpus-based versus intuition-based lexicography: defining a word list for a French learners dictionary*. In: *Corpus Linguistics Conference*. Lancaster University, UK, 2001. *Proceedings...* Disponível em [www.kuleuven.ac.be/ilt/grelep/publicat/verlinde.pdf](http://www.kuleuven.ac.be/ilt/grelep/publicat/verlinde.pdf). Acesso em: 12/08/2004.
- WITTGENSTEIN, L. 1979[1953]. *Investigações Filosóficas*. São Paulo, Abril Cultural, 211 p. (Coleção Os Pensadores).
- WITTGENSTEIN, L. 1990. *Da Certeza*. Lisboa, Edições 70, 191 p.

Submetido em: 10/2006

Aceito em: 11/2006

Milena Uzeda Garrão

Doutora em Estudos da Linguagem pela PUC-RJ