

## SEMIOSE E GRAMATICALIDADE

Jorge de Albuquerque VIEIRA (Pontifícia Universidade Católica de São Paulo)

*ABSTRACT: We will discuss here the concept of natural language as the result of a internalization process of actions between human systems and the environment. A homomorphy between these two levels carrying systemic coherence it's the origem of natural languages and of the human communication systems.*

*KEYWORDS: natural languages; homomorphy; relational collapse.*

### 0 - Introdução

O objetivo deste trabalho é apresentar alguns fundamentos para um enfoque realista e semiótico da noção de *Linguagem Natural*. O argumento fundamental a ser desenvolvido é que linguagens e processos de comunicação decorrentes são o resultado de internalizações relacionais estabelecidas entre os sistemas vivos e seus ambientes naturais, o que implicaria uma forma de *homomorfia* entre as leis sígnicas desta realidade e as leis sígnicas que caracterizam o *Umwelt* da espécie viva considerada. Sendo assim, teremos os sistemas de informação e os sistemas de comunicação como uma decorrência natural de um processo evolutivo fundamental, que trabalha pelo estabelecimento de *conectividade* entre os vários níveis ontológicos do real.

### 1 - Linguagens naturais

Uma linguagem dita natural nasce de um processo não planejado pela espécie humana e, nesse sentido, apresenta um caráter bem próximo à objetividade, apesar de envolver aspectos humanos subjetivos. Falamos aqui de *objetividade* no sentido da emergência de um sistema que não foi previamente planejado por um sujeito humano, nascendo espontaneamente ao longo de um processo evolutivo e histórico. Do ponto de vista sistêmico, uma linguagem (o que inclui principalmente as naturais) é um *sistema de signos*, satisfazendo às características e parâmetros sistêmicos como propostos pela atual *Teoria Geral de Sistemas* (Vieira, 2000: 22).

O caráter natural deste tipo de sistema de signos é manifesto por sua *estocasticidade*, ou seja, constituído por estados de um sistema onde a ocorrência de cada um depende de uma certa probabilidade. Frisemos o que isso implica: dado um sistema *S*, o mesmo pode ser representável pelo par ordenado (Bunge, 1977: 125):

$$S = \langle M, P \rangle$$

onde *M* é uma variedade associada à *composição* sistêmica, sistemas de referência, itens ambientais, etc., e *P* é uma lista de características ou propriedades do sistema, tal que

$$P = \{ p_i(t) \}$$

ou seja, uma lista de propriedades que apresentam intensidades variáveis no tempo. Tal variabilidade é devida, principalmente, ao fato de sistemas sempre serem *abertos* em algum nível. Se fixarmos um determinado instante de tempo, teremos que a coleção de intensidades das propriedades vigentes nesse instante é o *estado* em que o sistema se encontra no instante considerado. No caso da linguagem natural, a ocorrência seqüenciada de signos ao longo do tempo do discurso caracteriza a evolução dos estados do sistema de signos. Um *processo estocástico* apresenta a ocorrência dos estados como dependente de probabilidades. Ou seja, cada signo *i* depende, para ocorrer, de uma probabilidade  $p_i$ , tal que  $\sum_i p_i = 1$  ou 100%.

Assim, as *linguagens naturais* constituem uma classe dos chamados *processos estocásticos*, ou seja, processos naturais onde a evolução dos seus estados no tempo ocorre regida por uma *distribuição de probabilidades*. Para o desenvolvimento deste conceito, recorreremos a elaborações formais como encontradas nos estudos em *Linguística Matemática* (Marcus, 1978: 561). Seja uma *gramática G* definida a partir do par ordenado

$$G = \langle A, R \rangle$$

onde *A* é um conjunto de signos geralmente tomado como finito, o *alfabeto* e *R* uma relação definida sobre ele, a *sintaxe*. A aplicação da sintaxe ao alfabeto engendra subsistemas de signos, chamados *mensagens, M*. Uma *linguagem* é então definida como sendo um novo par

ordenado formado pela gramática e pelas mensagens que esta possibilita, tal que

$$L = \langle G, M \rangle$$

As linguagens naturais apresentam as várias probabilidades variando de signo a signo, sempre com algum grau de *heterogeneidade* na distribuição destas probabilidades. Assim, na linguagem do Português, a probabilidade  $p_A$  da letra A vir a ocorrer é diferente daquela da letra B, etc. Este é o caso de uma *fonte de informação discreta* (Shannon e Weaver, 1975: 43) ou seja, quando esta gera mensagens por meio de signos sucessivos. A partir desta concepção, podemos gerar uma hierarquia de processos estocásticos: quando os signos do alfabeto são escolhidos com a mesma probabilidade; quando os signos são escolhidos com diversidade nas probabilidades; quando os signos tiverem suas probabilidades dependendo da ocorrência de signos anteriores, ou seja, o pressuposto de uma *sintaxe*; e, finalmente, quando começamos a considerar, além das duplas de signos, os trios, as quádruplas, etc., até chegar a “palavras” de comprimentos diversos, arranjos estes submetidos a *probabilidades condicionais*. Uma elaboração desta seqüência, para a língua inglesa, é apresentada por Shannon e Weaver, 1975, p. 46.

É importante agora ressaltar a diferença entre um *sistema de informação* (como é o caso das linguagens naturais) e um *autômato*. Um autômato construirá suas mensagens sempre usando os mesmos signos, sempre sem alterar o conjunto de seus estados finais, só alterando a posição relativa entre os signos. Quando o sistema *seleciona* um certo número de signos do alfabeto disponível e assim constrói mensagens na forma de “palavras” de comprimentos diversos, estamos diante de um *sistema de informação*. Consideremos um sistema com  $n$  estados finais possíveis e uma mensagem chegando ao sistema de modo a reduzir o número de alternativas de  $n$  para algum número  $m < n$ . A quantidade de informação transportada pela mensagem é considerada igual a  $\log_2 (n/m)$  *bits* (Bunge, 1979: 271). Nesta formulação, a quantidade de informação do *sinal* é

$$I = \log_2 (1/p)$$

Onde  $p$  é a probabilidade do sinal atingir o receptor. Ou seja, uma linguagem natural é um exemplo de um sistema de informação discreto e finito. Quando consideramos a distribuição de probabilidades, temos uma

associação entre signos e suas respectivas probabilidades. Se considerarmos a matriz

$X_1 \ X_2 \ \dots \ X_n$

$p_1 \ p_2 \ \dots \ p_n$

associando os estados  $X_i$  às probabilidades  $p_i$ , temos o chamado *esquema finito* para um *sistema completo de eventos* tal que um e somente um ocorre em cada tentativa. (Khinchin, A. I., (1957) *Mathematical Foundations of Information Theory*. New York: Dover Publ. Inc., pg. 2). Todo esquema finito descreve um estado de *incerteza*, que é maior quando as probabilidades envolvidas são iguais e decresce com a heterogeneidade destas. Sabemos, da Teoria da Informação, que uma grandeza que mede a quantidade de incerteza associada com um dado esquema finito é a *entropia*, ou seja,

$$H(p_1, p_2, \dots, p_n) = - \sum_{1..n} p_k \log p_k$$

Desta forma, uma linguagem natural como processo estocástico apresenta uma certa entropia que, de acordo com o já aprendido na Mecânica Estatística, é uma medida de desorganização. A idéia de incerteza tem uma conotação nitidamente subjetivista, mas é uma hipótese, de natureza ontológica e que estaremos adotando aqui, que *todo processo subjetivo é uma decorrência de algum processo objetivo internalizado na memória do sistema pela evolução*. Ou seja, nossos critérios de incerteza têm raízes objetivas na desorganização do sistema de informação.

O que é sempre importante notar é que a entropia definida acima corresponde, matematicamente, a uma *média* e nesse sentido ela é *informação média* - a *informação* pode ser definida, nos casos sem ruído, como (Goldman, 1968: 4),

$$I = - \log [\text{probabilidade no receptor do evento, antes da mensagem ser recebida}]$$

o que é coerente com o conceito apresentado por Bunge, anteriormente; esta grandeza, “pesada” pelas probabilidades, ganha uma forma de média, a *informação média por símbolo*, (Goldman, 1968: 25).

A informação média ou entropia é uma medida de desorganização na mensagem. Pela teoria da informação, temos que, para um determinado alfabeto, a entropia máxima é sempre igual ao logaritmo do tamanho deste alfabeto, para qualquer base. Podemos então calcular uma entropia relativa, dividindo a entropia que uma mensagem ou sinal possui realmente pela entropia máxima que o sinal poderia ter. Isto nos dá uma medida de *desorganização relativa*. Se quisermos calcular o teor de organização do sinal ou mensagem, basta calcular o complemento desta razão, ou seja,

$$R = 1 - S_r/S_{\max}$$

onde R é a chamada *redundância*,  $S_r$  é a entropia realmente existente na mensagem e  $S_{\max}$  é a entropia máxima, no caso considerado. Ou seja, o que uma redundância mede, efetivamente, é o grau de organização do sinal ou da mensagem, uma grandeza, portanto, associada ao *vigor gramatical* envolvido.

## 2 – Colapsos relacionais

Tem sido colocado na literatura que a Teoria da Informação tem um caráter puramente sintático; mais ainda, que nesta as *regras gramaticais* são expressas diretamente pela distribuição das probabilidades associadas aos vários subsistemas de signos produzidos pela gramática. Apesar da limitação quanto à dimensão semântica, a base formal e sintática permite-nos discutir a hipótese que temos adotado, quanto à origem objetiva de processos subjetivos nos seres humanos, como decorrência da evolução.

Como poderia ser possível, a partir da imersão de um sistema cognitivo em um ambiente objetivo, preenchido de processos naturais estocásticos (entre outras classes de processos), o surgimento da *incerteza* subjetiva característica na maioria das interpretações da Teoria da Informação? Ou ainda, como se estabelece o enlace entre uma entropia do sujeito e a entropia física do ambiente em que este sujeito foi forjado pela evolução?

O que faremos a seguir é tentar encaminhar uma primeira discussão acerca desse problema, frisando sempre que este sempre será bem mais complexo do que poderia ser elaborado pelo nosso conhecimento atual.

O intercâmbio co-evolutivo entre um sistema cognitivo e seu ambiente sugere a existência de um processo de *mapeamento* entre tais sistemas. Na literatura, uma das propostas mais ricas quanto a este processo é sugerida por Rosemblyeth (1970: 57) na forma de uma *isomorfia estrutural*, um conceito conhecido em matemática (por exemplo, ver Weyl, 1949: 25). Segundo o autor, quando sistemas interagem e comunicam, o que efetivamente passa de um a outro é o *mapa estrutural isomórfico*, definido não sobre os elementos dos sistemas diversos, mas sobre *relações* de mesma classe ou tipo destes sistemas. Ou seja, a comunicação ou mapa conectivo se torna efetivo quando os sistemas se tornam “emparelhados” de forma coerente neste nível relacional. Na verdade, só há o processo efetivo de comunicação quando ocorre uma *ação* entre os elementos emparelhados, o que, na ontologia sistêmica de Bunge (1979: 6) caracteriza uma *conexão*.

Os processos de comunicação, de maneira geral, implicam em ações efetivas entre os sistemas, onde fluxos de informação (*diferenças*, no ponto de vista objetivo) são possíveis graças ao citado mapeamento (o que, em Teoria do Conhecimento, embasa as formas de Pragmatismo). As ações efetivas entre os sistemas cognitivos e a realidade, apoiadas na isomorfia estrutural, parecem ter gerado as características básicas das nossas linguagens. Como sugerido por Uexkull (1992: 288):

A análise da mente humana como um receptor de signos e a maneira como trabalha na construção do nosso Universo empírico exibe uma estrutura ou “anatomia” que é mais impressionante em sua analogia com a estrutura da linguagem. Assim como a linguagem tem signos de níveis diversos integracionais (fonemas, palavras, sentenças, etc.) , assim a análise de nosso universo empírico traz à luz diferentes níveis, em cada um dos quais processos signícos de complexidade variável podem ser estudados. Em cada um desses níveis os signos também revelam surpreendentes analogias aos signos da linguagem humana.

Dessa maneira temos a *ação dos signos* ou, como definida na semiótica de Charles Sanders Peirce, *semiose*. A semiose e os processos intersemióticos envolvem, assim, uma noção relacional ou sintática mais básica e de origem objetiva, que chamaremos *gramaticalidade* (ver, por exemplo, Vieira, 2000: 22; para uma discussão ampla do conceito de semiose, ver Santaella, 1995).

O conceito de *ação* entre sistemas e, no caso, *semiose*, quando estudada pelo ponto de vista sistêmico, que contém o citado *mapeamento estrutural isomórfico*, abre espaço para um outro conceito sistêmico,

aquele do *colapso relacional*. Para o sistemista russo Avanir Uyemov (1975: 98), há uma assimetria fundamental na natureza, que pode direcionar o “eixo do tempo”, manifesta no domínio relacional pela existência de dois tipos de relações: as *internas* e as *externas*.

Uma relação entre objetos  $x_1, x_2, x_3, \dots, x_n$  será dita *interna* quando ela for indispensável para a existência destes objetos. Uma relação *externa* será determinada não somente pelos objetos relacionados, mas também por certas condições dependendo de objetos externos. Por exemplo, quando uma pessoa é mais velha do que outra, temos uma relação interna (não podemos abandonar nosso lugar no eixo do tempo); quando uma pessoa está ao lado de outra a relação é externa (podemos mudar de posição).

A partir destes conceitos, Uyemov desenvolve sua análise sobre mudanças em coisas e mudanças em relações entre coisas. Seu objetivo no texto é discutir o conceito de *irreversibilidade* e sua característica temporal, o que o leva ao conceito de *colapso relacional*: uma vez que coisas entram em uma certa relação, elas atingem um tal estado (dependendo do efeito desta relação) tal que elas não podem abandoná-la a não ser abolindo suas existências como as coisas dadas (Uyemov, 1975: 99). Do ponto de vista sistêmico, na busca da *permanência* (Vieira, 2000:xx), as coisas e os sistemas decorrentes internalizam progressivamente as relações, “ao longo do tempo”.

Teremos assim sistemas externos, onde as relações externas predominam, e os internos, para o outro caso. As conclusões do autor implicam que

... se um sistema externo muda para um sistema interno, então o sistema permanece um sistema interno (Uyemov, 1975: 100)

E no contexto trabalhado, a conclusão final é:

O tempo flui na direção que corresponde à transição de sistemas externos para sistemas internos (Uyemov, 1975: 101)

Propomos então que a progressiva internalização de relações entre os sistemas e seus ambientes, admitida a *gramaticalidade* da realidade, gerou nos sistemas cognitivos uma gramaticalidade que, mais tarde, no processo evolutivo, veio a manifestar-se por meio das nossas linguagens naturais e demais linguagens utilizadas pelos nossos corpos (vide o código genético ou ainda a semiose típica do sistema imunológico).

As linguagens naturais faladas pelos seres humanos são processos estocásticos, assim como o são outros processos que identificamos no mundo físico. Ou seja, o que falamos é um caso de processo estocástico internalizado para sistemas cognitivos com as nossas características humanas. Nós, os humanos, comunicamos uns com os outros e, em talvez um nível menor de complexidade, os subsistemas que compõem a realidade objetiva comunicam também entre si.

### 3 -Codificação e intersemiose

Gostaríamos agora de frisar algumas aspectos concernentes ao *Problema da Essência do Conhecimento*, como encontramos na Gnosiologia. Tanto nos trabalhos de Immanuel Kant quanto nos de Charles Sanders Peirce, encontramos a ênfase na *Fenomenologia*. Em sua formulação clássica, há coisas reais, mas não podemos conhecer a sua essência. Só podemos saber “como” as coisas são, mas não “o que” são (Vita, 1964: 102). No âmbito da Semiótica, o que acessamos são os *signos* que representam as coisas.

Seguindo nossa hipótese centrada na objetividade, argumentaremos agora que a construção de signos em nossa atividade mental tem uma base objetiva onde o conceito de *Umwelt* de Jakob von Uexkull (1992) executa papel fundamental. Ou seja, o “Universo à volta” de uma espécie viva não é somente o “Universo subjetivo” desta espécie, mas é uma interface com uma conexão com a objetividade da realidade e outra, interna ao sistema vivo, com sua subjetividade. Nós, os humanos, podemos tentar falar do nosso caso particular, tendo em vista a grande dificuldade em descrever, entender ou visualizar o *Umwelt* de uma espécie viva diferente.

O exemplo mais óbvio do que propomos em nossa hipótese é o do *Umwelt* em sua (seu) componente visual. Quando vemos o mundo, sofremos o que pode ser chamada uma *tradução intersemiótica*. Algo na realidade (um objeto) emite processos físicos, entre os quais aqueles de natureza eletromagnética, tal que tais processos se propagam no nosso ambiente e agem sobre nosso olho. Este, em sua parte mais “externa”, apresenta fortes características físicas, embora com uma materialidade biológica. Uma lente orgânica, com transparência e curvatura, que direciona o fluxo de fótons para células especiais (cones e bastonetes), onde encontramos uma substância (rodopsina) que flutua em concentração de maneira coerente com a dinâmica do fluxo eletromagnético. Ocorre assim uma *tradução* que envolve um código físico, eletromagnético, e um código bioquímico, a flutuação na

concentração da rodopsina. A partir desta última, teremos codificações de natureza neuronal, cada vez mais “aprofundadas” em nosso cérebro, até que finalmente atingimos a impressão fenomênica de que “vemos algo”.

Ou seja, há uma ponte entre a realidade objetiva e nossa subjetividade; vários códigos estão envolvidos nessa sucessão de traduções, daí o termo *intersemiose*. E o que nos parece o ponto mais forte: se não houvesse eficiência nessas traduções, no sentido de construção de sistemas de signos *coerentes* com a realidade, sistemas vivos como nós *não teriam conseguido sobreviver nesta realidade*.

#### 4 – Considerações finais

Podemos agora resumir os pontos principais da discussão feita:

Nossa subjetividade apresenta processos que são os resultados evolutivos de processos objetivos progressivamente internalizados no tempo. Isso é o que caracteriza, em uma Ontologia Sistemica, o conceito de *função memória* (Bunge, 1979: 161). Um sistema complexo, como é o caso de um sistema humano, apresenta vários tipos de memória e uma delas responde por nossa capacidade de elaborar discursos.

A principal característica do conhecimento discursivo é a sua temporalidade. Esta refletiria a natureza orientada do tempo objetivo como um índice do processo de internalização que gerou nossas linguagens naturais.

Admitida a isomorfia que permite os mapeamentos que acarretam a internalização das relações (ou ainda uma *homomorfia*, se levarmos em conta a natureza “ruidosa” dos mecanismos discutidos), estaríamos diante da possibilidade de uma realidade que apresenta, incluindo nossos corpos, vários níveis de textualidade.

E, finalmente, as linguagens naturais humanas evoluem de um mecanismo de geração de organização coerente para uma forma de conectividade que pode vir a garantir a emergência de um efetivo sistema social.

#### 5 – REFERÊNCIAS BIBLIOGRÁFICAS

BUNGE, Mario. *Treatise on Basic Philosophy* – Vol. 3. Dordrecht: D. Reidel Publ. Co., 1977.

\_\_\_\_\_. *Treatise on Basic Philosophy* - Vol. 4. Dordrecht: D. Reidel Publ. Co., 1979.

GOLDMAN, Stanford. *Information Theory*. New York: Dover Publ., 1968.

VIEIRA, J. A. Semiose e gramaticalidade. *Revista Intercâmbio*, volume XV. São Paulo: Lael/PUC-SP, ISSN 1806-275X, 2006.

- KHINCHIN, A. I. *Mathematical Foundations of Information Theory*. New York: Dover Publ. Inc., 1957.
- MARCUS, Solomon. Mathematical and Computational Linguistics and Poetics. *Revue Roumaine de Linguistique*, Tome XXIII, N. 1-4, 559-589., 1978.
- ROSEMBLUETH, Arturo. *Mind and Brain - A Philosophy of Science*. Massachusetts: The MIT Press, 1970.
- SANTAELLA, Lúcia. *A Teoria Geral dos Signos - Semiose e Autogeração*. São Paulo: Ed. Ática SA., 1995.
- SHANNON, Claude ; WEAVER, Warren. *A Teoria Matemática da Comunicação*. Rio de Janeiro: Difel/Difusão Editorial S.A., 1975.
- UEXKULL, Thure. A Stroll Through the Worlds of Animals and Men. *Semiotica (Special Issue)*, Berlin, 89-4., 1992.
- UYEMOV, Avanir. *Problem of Direction of Time and the Laws of System's Development*. In *Entropy and Information in Science and Philosophy*. KUBAT, Libor.; ZEMAN, Jiri. (Eds.), 93-102. Praga: Elsevier Sc. Publ. Co., 1975.
- VIEIRA, Jorge de Albuquerque. Organização e Sistemas, *Informática na Educação: Teoria e Prática / Programa de Pós-Graduação em Informática na Educação - vol. 3, n. 1*. Porto Alegre, UFRGS, 11-24., 2000.
- VITA, Luiz Washington. *Introdução à Filosofia*. São Paulo: Ed. Melhoramentos, 1964.
- WEYL, Hermman. *Philosophy of Mathematics and Natural Science*. Princeton: Princeton, 1949.T