

ASPECTOS FORMAIS DA LINGÜÍSTICA TEXTUAL: PROPOSTA DE UM MODELO DE ANÁLISE SINTÁTICA TRANSFRÁSTICA

ALDO LUIZ BIZZOCCHI¹

Resumo: O presente trabalho procura examinar sob que condições uma dada seqüência de frases constitui um texto e, a partir daí, formula algumas hipóteses sobre as relações sintáticas existentes ao nível transfrástico do texto, para, a seguir, propor um pequeno esboço do que seria um modelo formal (lógico-matemático) de análise e descrição de textos lingüísticos com base em tais relações.

Palavras-chave: Lingüística textual; sintaxe transfrástica; modelos matemáticos de análise lingüística.

Abstract: This paper tries to examine under what conditions a given sequence of sentences constitutes a text and, subsequently, formulates some hypotheses about the syntactic relationships existing at the textual level, in order to, next on, propose a small sketch of what a formal (logical-mathematical) model of analysis and description of linguistic texts on the basis of such relationships would be.

Keywords: Textual linguistics; textual syntax; mathematical models of linguistic analysis.

Introdução

O interesse de cientistas, filósofos e homens de cultura em geral pelo objeto *texto* não é novo. Entretanto, somente nas últimas décadas tem sido possível empreender um estudo mais sistemático desse objeto na medida em que a própria ciência lingüística evoluiu, fornecendo as chaves fundamentais de uma nova abordagem desse complexo problema. Para tanto, a lingüística estruturalista trouxe decisiva contribuição ao descrever as estruturas que subjazem à linguagem, o que permitiu o assentamento dos estudos lingüísticos em bases formais. A possibilidade de formalização das estruturas lingüísticas, por sua vez, conduziu à enunciação das regras operatórias que regem a formação e transformação de tais estruturas: foi essa a grande contribuição da gramática gerativo-transformacional. Contudo, as tentativas de descrição dos princípios estruturais da língua não haviam, até algum tempo atrás, conseguido passar além do limite da frase.

Com o desenvolvimento da semiótica lingüística e da análise do discurso, o objeto de estudo da lingüística se expandiu até o nível textual. No entanto, tais abordagens

consideram sobretudo o texto no seu todo muito mais do que na constituição de suas partes e assume um ponto de vista muito mais semântico do que sintático.

Sabe-se, por outro lado, que as frases também se organizam de forma estruturada na constituição do texto. O grande desafio que se impõe ainda hoje, visto que o problema não está totalmente resolvido, é enunciar as regras de uma sintaxe transfrástica com a mesma precisão e rigor formal com que se pode hoje tratar a sintaxe frástica.

Diversos estudos vêm sendo empreendidos nesse sentido, em especial na Alemanha, França e em países do Leste europeu. Tais estudos, longe de ser conclusivos, apontam, no entanto, em direções convergentes, permitindo vislumbrar a “gramática do texto”, capaz de descrever qualquer texto, independentemente de sua natureza, como uma rede de inter-relações funcionais, isto é, como conjunto de elementos que se articulam entre si mediante relações semânticas e lógicas, formando uma estrutura compacta, uma unidade de comunicação. Em outras palavras, tal gramática reduz o texto a uma expressão “algébrica”.

Este trabalho pretende formular algumas hipóteses sobre as relações sintáticas existentes no nível transfrástico, bem como propor um esboço de modelo formal de descrição de textos com base em tais relações.

Coesão e coerência

Os teóricos da lingüística textual têm perseguido um modelo de estruturação do texto ao mesmo tempo sintática e semântica. Buscam regras que permitam não apenas distinguir entre textos gramaticais e agramaticais mas também entre textos semanticamente aceitáveis e inaceitáveis. Os conceitos de gramaticalidade/aceitabilidade semântica estão intimamente ligados aos de coesão/coerência textual, que passamos a discutir.

Dentre os autores que fazem distinção entre *coesão* e *coerência* estão Mateus et alii (1983:186), que falam em *conectividade seqüencial* (coesão) e *conectividade conceptual* (coerência). A coesão resulta então dos “processos de seqüencialização que asseguram (ou tornam recuperável uma ligação lingüística significativa entre os elementos que ocorrem na superfície textual”. Já a coerência “é um fator de textualidade

¹ Doutor em Semiótica e Lingüística Geral pela Universidade de São Paulo e pesquisador do Círculo Fluminense de Estudos Filológicos e Lingüísticos – CiFEFiL. Website: www.aldobizzocchi.com.br.

que resulta da interação entre os elementos cognitivos apresentados pelas ocorrências textuais e o nosso conhecimento do mundo”.

Para Van Dijk (1977), a coerência de um texto está ligada à noção de macroestrutura textual, também definível como a estrutura profunda do texto, ao passo que a coesão diz respeito ao encadeamento linear das frases ao nível da estrutura de superfície do texto.

Marcus (*apud* Bernárdez, 1982:157) define coerência como “uma certa capacidade de atuar como unidade, enquanto coesão se refere à existência de conexão entre as diferentes partes. A coerência é antes de natureza semântica, nos remete a um certo significado global do texto; a coesão parece dominada por aspectos sintáticos e relacionais entre os componentes. Entretanto, seria imprudente e simplista considerar a coerência como fenômeno exclusivamente semântico, e a coesão como exclusivamente sintático”.

De forma geral, pode-se entender a coesão textual como o conjunto de relações sintático-semânticas dos elementos constitutivos do texto entre si, enquanto a coerência textual é a relação entre o texto e o contexto extralingüístico, os sujeitos da enunciação, os dados da experiência acerca do mundo disponíveis aos falantes da língua e as condições objetivas e subjetivas de produção do ato lingüístico.

Todas as definições de coerência textual implicam a noção de *mundo*: a análise da coerência de um texto é função de sua compatibilidade ao conhecimento que temos do mundo. Postula-se assim a existência de um *mundo real* (normal) e de *mundos alternativos* (anormais) em relação aos quais o esquema cognitivo proposto pelo texto deve ser interpretado. A decisão sobre a coerência ou não do texto deve então ser tomada a partir de um conhecimento extratextual – e por vezes extralingüístico – que possuem os sujeitos da enunciação acerca do tema proposto. É essa dimensão pragmática, característica do nível textual, o que, em última instância, permite distinguir o discurso “em grau zero” do discurso irônico e do ficcional. Por isso mesmo, as marcas da coerência textual nem sempre são encontráveis no próprio texto: um enunciado incoerente quando tomado de forma isolada pode recuperar sua plena coerência quando religado à sua situação original de enunciação.

Discussão semelhante se deu, ainda em termos da gramática frástica, quando da proposição da célebre dicotomia chomskyana *gramaticalidade/aceitabilidade*. Desde logo constatou-se que frases agramaticais (do ponto de vista gerativo-transformacional) podem ser perfeitamente aceitáveis semanticamente, assim como uma frase gramaticalmente bem construída pode conduzir à contradição lógica e ao *nonsense*. Uma frase como

O menino sexagenário fará oitenta anos amanhã.

será perfeitamente normal e coerente numa peça de teatro do absurdo, por exemplo. A coerência, portanto, depende tanto do contexto do enunciado (por vezes chamado de *contexto*) quanto da situação de enunciação, isto é, da macroestrutura lingüística na qual se dá o ato de comunicação. Já a coesão textual, embora também apresente uma dependência semântica, é de natureza essencialmente gramatical. Eis por que, sendo impossível enunciar regras de coerência textual independentes do aspecto pragmático, a busca de um modelo formal de descrição sintática de textos deverá recair na questão da enunciação de regras de coesão textual. O problema se reduzirá assim à descrição das relações estritamente lingüísticas existentes entre as frases de um texto dado, não importando a interpretação subjetiva que dele se faça em razão dos diferentes contextos em que possa ocorrer.

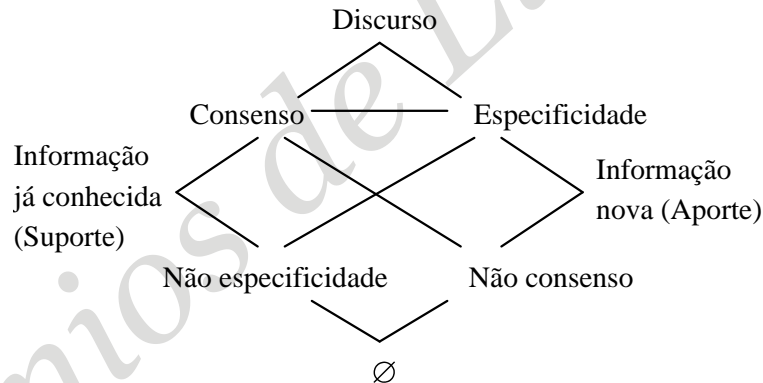
Progressão temática e núcleos temáticos

Todo discurso parte de um consenso em direção a uma especificidade, vale dizer, parte de uma informação já conhecida por ambos os sujeitos da enunciação, quer por sua ocorrência em discursos anteriores, quer por sua presença no contexto extralingüístico, sobre a qual se faz o aporte de informação nova, inédita. À informação já conhecida dá-se o nome de *tema* ou *tópico*, e à informação nova dá-se o nome de *rema* ou *comentário*. Evidentemente, toda informação nova torna-se de domínio comum dos sujeitos da enunciação no momento de sua veiculação no texto. Transforma-se assim em suporte de uma nova informação. Dito de outro modo, todo rema pode tornar-se tema em relação a novos remas. Por essa razão, distinguiremos entre *tema principal* (aquele que não deriva de outro tema dentro do mesmo texto, podendo entretanto

derivar do contexto pré e extratextual) e *tema secundário* ou *subtema* (aquele que é rema de temas anteriores pertencentes ao mesmo texto). O constante movimento de tema a rema é chamado de *progressão temática*. Ela é condição *sine qua non* da existência do texto. Com efeito, um discurso circular em torno de um tema sobre o qual nada se diz de novo só pode ser encarado do ponto de vista patológico ou como pura abstração teórica, fugindo ao interesse da descrição científica. Entre os limites da informação nula e da informação infinita é que se localiza o ato comunicativo, isto é, o texto.

Ao nível da superfície textual, os sucessivos temas e remas podem ser detectados através de unidades léxicas que sintetizam e concentram a carga semântica do texto: podemos chamá-las de *núcleos temáticos*. A progressão temática poderá então ser entendida como o encadeamento de tais núcleos ao longo do texto.

Para Charaudeau (1972), todo discurso resulta da tensão dialética entre o consenso e a especificidade, que, como vimos, articula o não sabido e o já sabido num mecanismo de suporte/aporte. Ou esquematicamente:



Assim, é natural que todo texto seja parcialmente profluente e parcialmente redundante. A profluência do texto garante sua informatividade, ao passo que a redundância assegura a fixação do tema e a integração dos constituintes textuais no seu desenvolvimento seqüencial.

A esse respeito, Charolles (1978) propõe a existência de quatro meta-regras básicas de coerência textual: a *repetição*, a *progressão*, a *não contradição* e a *relação texto/contexto*. Destas, particularmente as meta-regras de repetição e de progressão

dizem respeito mais diretamente à construção gramatical do texto, contemplando a citada tensão redundância/profluência. Essa tensão se reflete no surgimento progressivo de novos núcleos temáticos ao longo do texto, núcleos estes que, no entanto, tendem, com maior ou menor frequência, a se repetir à medida que o texto avança.² De modo geral, todos os elementos lexicais de um texto que se repetem de forma significativa constituem núcleos temáticos. Contudo, Charolles distingue três tipos diferentes de repetição, que agruparemos em duas categorias: a repetição direta, que o autor chama de *definitivização* (reiteração da mesma palavra ou expressão), e a repetição indireta, que se subdivide em *pronominalização* (uso de pronomes e proformas em lugar de um termo), a qual pode ser anafórica ou catafórica, e *substituição* (quer por sinonímia quer por paráfrase). Numa possível representação esquemática do texto, poderíamos convencionar que os núcleos temáticos fossem representados por algarismos romanos, a cada repetição direta de um núcleo correspondendo a repetição do algarismo. Os núcleos substituídos por pronomes e proformas ou por sinônimos e paráfrases (substituição indireta) seriam indicados por um algarismo romano precedido do prefixo “sub”, ao passo que os núcleos substituídos por \emptyset (fenômeno a que damos o nome de *elipse*) seriam representados por um algarismo romano entre parênteses. Caso dois núcleos temáticos sejam antônimos ou apresentem no texto algum tipo de oposição funcional exclusiva, indicaremos ambos com o mesmo algarismo, antepondo a um dos dois um sinal de menos (-).

As unidades sintáticas do texto: frasemas e strings

A concepção do texto como uma estrutura implica a existência de unidades constitutivas que mantêm vínculos funcionais entre si. Tal concepção nos conduz à idéia de uma hierarquia entre essas unidades e, conseqüentemente, à possibilidade de explicitá-la na forma de um esquema ou diagrama. Essa hierarquia subjacente aos elementos constitutivos do texto identifica-se bastante bem ao que Van Dijk (1977; 1983) denomina *macroestrutura textual*. Todavia, para poder estudar as relações

² Quanto maior a frequência de repetição de um núcleo temático, maior a sua relevância semântica para o texto e maior a sua pertinência ao tema a que está subordinado. Este é inclusive o princípio de conceituação de palavra-tema e palavra-chave, de que, no entanto, o núcleo temático cumpre distinguir-se.

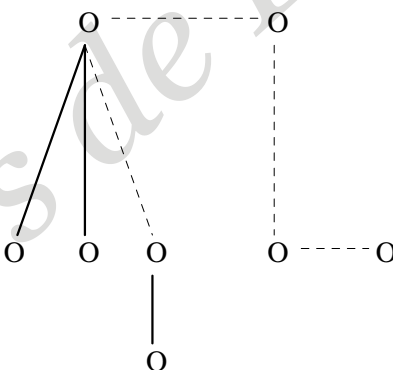
funcionais entre as unidades constituintes do texto, é necessário primeiramente definir quais são essas unidades.

Tradicionalmente, costuma-se considerar o texto como um conjunto de frases, e, portanto, seria a frase a unidade textual por excelência. Esse ponto de vista acarreta inúmeros problemas, a começar da própria conceituação de frase (veja-se a esse respeito Câmara Jr., 1977:162). Mesmo que se adote uma definição consistente de frase, ainda resta o problema da possibilidade de reescritura de uma frase em várias ou vice-versa. Por essa razão, alguns teóricos preferem adotar como unidade textual não a frase mas sim a oração, em parte devido à expectativa de que a gramática textual contenha a gramática frasal (Van Dijk, 1977:204-5).

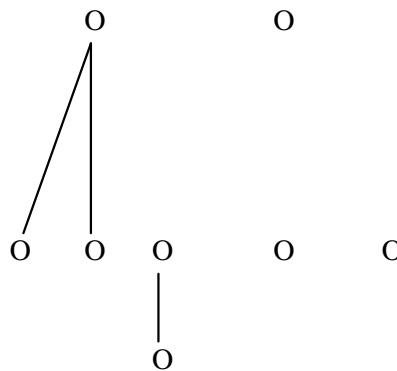
Entretanto, o estudo sintático dos períodos compostos revela diferentes tipos de relação entre as orações, e o problema da reescritura se recoloca: há orações subordinadas que podem transformar-se em coordenadas e vice-versa; podem, por vezes, constituir frases autônomas. Há outras, no entanto, que permanecem sempre subordinadas a uma oração principal, como elemento sintático obrigatório daquela, não podendo ser reescritas sem que o significado geral do período seja drasticamente modificado. Além disso, embora toda oração subordinada desempenhe uma função sintática no seio da oração principal, alguns tipos de subordinadas podem ser suprimidas *sem prejuízo sintático* da oração principal, embora com evidente prejuízo semântico, ao passo que outras, quando suprimidas, simplesmente destroem a oração principal. Isso revela que, mesmo na subordinação, há diferentes graus de dependência entre as orações, o que sugere adotar como unidade de texto todo segmento sintático que se enquadre em uma das situações abaixo:

- (i) período simples;
- (ii) oração coordenada;
- (iii) oração subordinada passível de supressão sem prejuízo sintático da oração principal;
- (iv) oração principal cujas subordinadas se enquadrem em (iii);
- (v) período composto por subordinação, não decomponível nos termos de (iii) e (iv).

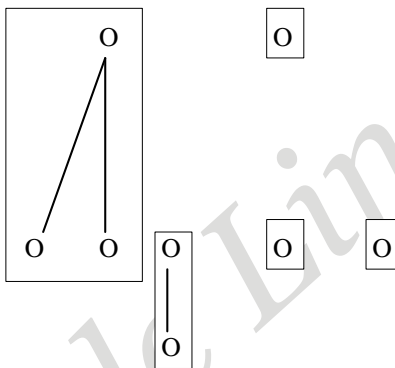
A essa unidade assim definida daremos o nome de *frasema*, e estabeleceremos sua definição em bases mais rigorosas. Para tanto, partiremos da constatação de que há períodos formados por uma única oração, períodos formados por orações coordenadas e períodos formados por uma oração principal e uma ou mais subordinadas (além, é claro, das várias combinações dessas possibilidades). Especificamente no que tange ao período composto por subordinação, observa-se a existência de dois diferentes tipos de relação subordinativa, segundo a oração subordinada exerça uma função sintática fundamental dentro da oração principal ou apenas uma função acessória. Ao primeiro tipo de relação daremos o nome de *subordinação forte* e ao segundo, de *subordinação fraca*. Pode-se dizer que a subordinação forte é uma relação indissociável, ao passo que a subordinação fraca e a coordenação são relações dissociáveis. Se explodirmos a frase de modo que se rompam todas as relações dissociáveis, permanecendo apenas as indissociáveis, cada uma das subdivisões resultantes dessa explosão será uma oração simples ou um conjunto formado por uma oração principal e uma ou mais orações subordinadas fortes. São exatamente essas unidades o que chamamos de *frasema*. Seja o seguinte diagrama:



onde O representam orações, as linhas pontilhadas horizontais indicam coordenação, as linhas descendentes contínuas indicam subordinação forte e as linhas descendentes pontilhadas representam subordinação fraca. Trata-se, portanto, da representação esquemática de um período composto. Explodindo esse período, isto é, apagando do diagrama as linhas referentes a relações dissociáveis, teremos o novo diagrama a seguir.



Emolduremos agora cada conjunto isolado assim obtido num quadrilátero.



Cada um dos quadriláteros acima representa um frasema.

Da definição dada de subordinação forte e subordinação fraca resulta que são subordinadas fortes basicamente – mas não exclusivamente – as orações substantivas, com grande acerto chamadas também de orações integrantes. As demais subordinadas (adjetivas e adverbiais) são, via de regra, subordinadas fracas.

A partir de agora, representaremos o frasema sempre entre barras verticais (|).
Vejam os exemplos a seguir:

- (1) |Comprei um carro novo.|
- (2) |Vim,| |vi,| |venci.|
- (3) |Despediu-se de todos| |e partiu.|
- (4) |Como estivesse doente,| |João não foi à escola.|

(5) |Se for eleito,| |ele será empossado. |

(6) |Quero que você me faça um favor. |

(7) |É importante que eu vá à reunião. |

A frase (1) constitui um período simples, e portanto, um frasema. Em (2) e (3) temos frasemas coordenados sindética ou assindeticamente. Já as frases (4) a (7) representam períodos compostos por subordinação. Em (4), por exemplo, a oração subordinada “Como estivesse doente”, ao ser suprimida, produz a oração simples “João não foi à escola”, a qual se mantém perfeitamente gramatical e de sentido completo, o mesmo ocorrendo em (5). Em ambos os casos, temos dois frasemas. Em (6) e (7), ao contrário, a supressão da subordinada conduz a orações incompletas como “Quero” ou “É importante”. Neste caso, as frases não são decomponíveis em frasemas menores.

Há casos, entretanto, em que a decomposição de um período composto em frasemas menores, embora possível, não é necessária. É o caso das orações subordinadas adjetivas, que normalmente vêm encaixadas no meio da principal. Por exemplo, o frasema

(8) |O homem que trabalha vence na vida. |

é perfeitamente decomponível em dois frasemas menores, |O homem vence na vida. | e |que trabalha|; contudo, a posição tática da oração adjetiva no interior da principal desaconselha, por razões meramente práticas, sua decomposição. Aliás, a decomposição de um frasema em frasemas menores é, em geral, facultativa e depende do tipo de relação existente entre os frasemas constituintes e de sua importância dentro da macroestrutura textual em que se encontram.

Enquanto unidades básicas da sintaxe do texto, os frasemas estabelecem entre si relações lógico-funcionais, a que chamaremos de *conexões frásticas*, formando assim cadeias de frasemas. Estas, por sua vez, também apresentam conexões em relação a outros frasemas e/ou cadeias de frasemas, de modo que tais cadeias se comportam de forma análoga a um frasema simples. Vamos chamá-las de *strings* por analogia às cadeias lineares de mesmo nome, propostas por Harris (*apud* Borba, 1979). Os *strings* podem combinar-se progressivamente, formando unidades cada vez maiores. Por essa

razão, todo *string* pode ser chamado de *substring* em relação ao *string* mais extenso que o contém. Da definição de *string* resulta que todo texto é um grande *string*, subdivisível em *substrings*; por outro lado, todo frasema pode também ser considerado como um *string* unitário. Cumpre lembrar ainda que o conceito de *string* não coincide com o de parágrafo, embora em geral todo parágrafo constitua um *string*.

Núcleo temático, frasema, *string* e conexão são, portanto, os conceitos básicos do modelo proposto. Na notação esquemática, o *string* será indicado por colchetes ([]) envolvendo os frasemas e/ou *substrings* que o compõem. Isso permite uma hierarquização das conexões do texto, de forma análoga ao procedimento da parentetização em álgebra. Com efeito, uma sentença matemática do tipo

$$3 \times 4 + 2 = 12 + 2 = 14$$

é diferente de

$$3 \times (4 + 2) = 3 \times 6 = 18$$

Os parênteses no segundo caso indicam qual operação aritmética deve ser efetuada em primeiro lugar. Conseqüentemente, nessa sentença o fator 3 multiplica todo o conteúdo dos parênteses, ou seja, $4 + 2 = 6$, enquanto no primeiro caso multiplica apenas o 4. Esquemas semelhantes são largamente utilizados em lógica matemática e seus corolários (teoria da informação, computação, lingüística gerativa, etc.). A resolução de esquemas parentetizados parte sempre dos parênteses mais internos da sentença em direção àqueles mais exteriores. A esse procedimento dão os lógicos o nome de *análise centrípeta*. Por exemplo, na sentença

$$((4 + 6) \times (5 - 3)) + 5$$

resolvem-se primeiro os parênteses internos, isto é, $(4 + 6)$ e $(5 - 3)$, obtendo assim a nova expressão

$$(10 \times 2) + 5 = 20 + 5 = 25$$

Em nosso caso específico, a parentetização dos frasemas, criando *strings*, resulta da própria hierarquia funcional dos frasemas entre si, de forma totalmente análoga a uma sentença matemática. Porém, em lugar de números e operações aritméticas, teremos respectivamente frasemas e conexões frásticas. A exemplo da lógica e da aritmética, cabe lembrar que todo colchete aberto deve ser fechado, de sorte que toda sentença completa (em nosso caso, todo diagrama de texto) deve ter tantos sinais “]” quantos forem os sinais “[”. Por conseguinte, os colchetes serão sempre em número par.

As conexões frásticas

Conforme anteriormente definido, *conexão frástica*³, ou simplesmente *conexão*, é a relação lógico-funcional existente entre duas unidades sintáticas transfrásticas (frasemas ou *strings*), constituindo portanto o liame fundamental de todos os elementos integrantes da armação do texto. Sendo uma relação de caráter essencialmente lógico, a conexão pode ser encarada como uma função análoga à relação entre proposições na lógica matemática. Sobretudo, isso permite representar graficamente as conexões por meio de sinais semelhantes aos dos operadores lógicos. É preciso, contudo, lembrar que nem sempre o tipo de conexão existente entre os frasemas ou *strings* é detectável na superfície textual por índices sintáticos (advérbios, conjunções, pontuação, etc.). Na verdade, o estatuto das conexões é de natureza fundamentalmente semântica, conforme assinala Van Dijk; por essa razão, o modelo aqui proposto, sendo essencialmente descritivo, não discute quais índices, quer sintáticos quer semânticos quer pragmáticos, determinam tal ou qual tipo de conexão, mas procura apenas enunciar uma operatória dessas conexões, deixando aos semanticistas a tarefa de pesquisar os mecanismos intrínsecos de estruturação do texto.

Isso posto, passemos a relacionar os principais tipos de conexões frásticas e sua simbologia⁴:

³ O conceito de *conexão* aqui formulado concorda com o de Van Dijk (1977) e equivale aos termos *junção* (Beaugrande; Dressler, 1988), *conjunção* (Halliday; Hasan, 1975) e *coesão interfrástica* (Mateus et alii, 1983).

⁴ A lista de conexões aqui apresentada evidentemente não esgota todas as possibilidades de conexão entre frasemas. Certamente muitos outros tipos de conexão poderão ser detectados à medida que mais e mais amostras forem estudadas. Além disso, esta classificação é bastante sumária e simplista, estando sujeita a

Desenvolvimento (\rightarrow): é a conexão que liga um frasema temático ou subtemático, representados graficamente entre barras verticais duplas ($\| \|$), ao seu rema. O desenvolvimento é o elemento básico do processo de textualização dissertativa e também por vezes narrativa, visto que tal tipo de texto consiste fundamentalmente na apresentação do tema ou assunto a ser tratado (suporte) seguido da informação que sobre ele o texto traz (aporte). Um exemplo típico de desenvolvimento se dá entre o tópico frasal e o restante do parágrafo.

Síntese (\leftarrow): trata-se do inverso do desenvolvimento. Neste caso, as diversas informações aportadas pelo texto confluem para um frasema que as sintetiza e lhes serve de suporte. Na síntese, o tema ou tópico frasal vem posposto ao rema. Por vezes, encontramos na superfície textual índices de síntese que poderíamos chamar de *conectores sintéticos* (expressões tais como *em suma, em resumo, trata-se, portanto, de, etc.*).

Conjunção (+): parafraseando a definição de Mateus et alii (1983:192), a conjunção articula seqüencialmente frasesmas cujos conteúdos proposicionais se verificam de forma compatível entre si, quer simultânea, quer sucessivamente, no mundo proposto pelo texto. A conjunção é uma das formas mais freqüentes de conexão frástica. Dentre os diversos conectores conjuntivos encontráveis na superfície textual estão *e, a seguir, também, assim como, ao mesmo tempo, da mesma forma, etc.*

Disjunção (/): articula seqüencialmente frasesmas cujos conteúdos proposicionais estão em relação alternativa, o que equivale a dizer que a ocorrência de um deles pode (e eventualmente deve) excluir a ocorrência de outro. Temos disjunção nas proposições alternativas do tipo *ou...ou*, por exemplo (Mateus et alii, 1983:193).

Contração (//): conecta frasesmas cujos conteúdos proposicionais apresentam oposição contrastiva ou adversativa, sendo, entretanto, e à diferença do que acontece na disjunção, co-ocorrentes no mesmo mundo dado (*idem*, p. 194-5). Índices freqüentes de contração são os conectivos *mas, porém, entretanto, todavia, etc.*

Implicação (\Rightarrow): é a conexão que estabelece entre dois frasesmas uma relação de causa e efeito. Verifica-se quando o conteúdo proposicional do frasema antecedente é condição necessária, suficiente ou possível para a ocorrência do conseqüente. Sua

revisões e reformulações. De qualquer maneira, o objetivo principal é demonstrar a possibilidade de descrever a sintaxe transfrástica por meio de um modelo formal.

definição é, portanto, semelhante à da implicação lógica e seus conectores principais são *portanto, conseqüentemente, se...então, por isso*, etc.

Contra-implicação (\Leftarrow): é o inverso da implicação. Neste caso, o frasea antecedente vem posposto ao conseqüente. O segundo membro da conexão é, pois, a causa ou justificativa do primeiro. Seus índices mais comuns são *porque, dado que, visto que*, etc.

Explicação (:): introduz um frasea ou *string* cuja função é a de explicar, detalhar, ampliar a informação contida no frasea precedente. A essa função podemos chamar de *expansão semântica*, visto que a explicação opera uma adição de novos semas ao semema da proposição a ser explicada. Por vezes, a explicação desempenha papel redundante, parafraseando metalingüisticamente o conteúdo do frasea antecessor ou introduzindo uma exemplificação do mesmo. Cumpre não confundir a explicação com a implicação ou o desenvolvimento: a explicação não encerra uma relação determinística de causa e efeito nem introduz comentário acerca do tópico. Nem sempre a explicação apresenta conectores sintáticos explícitos; entretanto, poderíamos citar os dois pontos (:), além de expressões como *isto é, ou seja, com efeito*, etc.

Pressuposição (\Rightarrow): mais comum entre *strings*, a pressuposição serve para apresentar um argumento ou raciocínio que dá sustentação ideológica ao tema a ser desenvolvido. Pode-se dizer que a pressuposição é a “justificativa do texto”, ao passo que a implicação introduz a justificativa de um frasea ou *string*. As famosas fórmulas jurídicas do tipo *considerando que...o Presidente resolve...* são exemplos característicos de conexão pressupositiva. Embora em geral não haja nexos semânticos de causalidade (ao menos em nível explícito) na pressuposição, esta também expressa uma forma de motivação, porém trata-se aí de uma motivação lógico-psicológica por parte do locutor em relação ao seu texto. A pressuposição é, portanto, a mais pragmática das conexões frásticas.

A aplicação dos conceitos acima à análise de um texto a título de ilustração tornará mais claro o funcionamento das conexões e, de modo geral, da sintaxe transfrástica.

O modelo descritivo da sintaxe transfrástica – aplicação a um texto real

O texto descrito a seguir segundo o modelo sintático proposto é o resumo do artigo científico “Preservação parcial da concha de *Tentaculites crotalinus*, da Formação

Ponta Grossa (Devoniano)”⁵, de autoria de José Henrique Godoy Ciguel, Oscar Rösler e Robin M. Hofmeister, a quem agradecemos pela gentil cessão do mesmo para nosso uso. Para permitir a comparação entre o texto e seu diagrama analítico, este último é apresentado logo acima de cada linha do texto. Nele estão notados os núcleos temáticos em algarismos romanos, segundo a convenção adotada, e sublinhados os mesmos no texto. Para maior facilidade de esquematização, os frasemas foram numerados com algarismos arábicos. Os conectores frásticos, quando explícitos, aparecem no texto em negrito. Para melhor visualização, o texto foi transcrito de forma seqüencial, isto é, sem espaço entre parágrafos. Em substituição a estes, usou-se na transcrição o sinal §.

⁵ Trabalho publicado no *Boletim IG-USP*, série científica, v. 18, 1987, p. 17-26.

I II III
 O presente trabalho trata da análise de um espécime de *Tentaculites crotalinus* SALTER emend. IV
 CIGUEL et al. (1984), coletado em afloramentos da Formação Ponta Grossa (Devoniano) nas V subII VI
 proximidades da cidade de Jaguariaíva, Estado do Paraná, e da discussão sobre sua possível estrutura VII || → [1]2 -subVI VIII
 e composição esquelética. § O relativo desconhecimento da microestrutura esquelética da concha dos subIII -IX
 Tentaculitoidea tem dificultado o reconhecimento das relações filogenéticas desses subIII |⇒3 subIII -subIX [1] + 4
 invertebrados; conseqüentemente, o seu posicionamento sistemático tem sido problemático. § Os subIII IV V
 tentaculitídeos são bastante comuns no Devoniano da Bacia do Paraná e conhecidos através de seus X [1]5 II XI
 moldes externos e internos. O espécime em questão apresentou provável fragmento da parede da VIII X [1]6 subI XII XIII
 concha entre os moldes. § Esse estudo teve por objetivo reconhecer se essa camada representaria uma subX XI VIII || → [1]7 subXII
 contramoldagem ou fragmento da parede da concha. § Com essa finalidade, analisou-se a VII XIII X
 composição química da camada e do molde externo, através de espectrometria de Raios X X VI XIII
 e difratometria de Raios X apenas no molde externo, além da observação da estrutura da camada ||⇒[1]8 XIII
 através de lupa binocular e de microscópio eletrônico de varredura. § Verificou-se que essa camada subVI XIV
 apresenta uma microestrutura foliada, formada pela superposição de lamelas orientadas subparalela XI VIII |9 XIII
 a paralelamente à parede da concha. Verificou-se também que a camada apresenta uma VII
 composição química distinta da rocha circundante, mantendo elementos possivelmente oriundos da VII VIII [1] ⇒ [1]10
 composição original da concha. § Torna-se, portanto, bastante provável tratar-se de um caso de XI VIII |11 VIII subVII
 preservação parcial da parede da concha. A concha seria composta por carbonato de cálcio com

XIV subVI |+12 XIII
lamelas sobrepostas possivelmente com microestrutura foliada. A camada, durante os processos
 diagenéticos da matriz, sofreu uma impregnação superficial de compostos de ferro e manganês. § A
 subVI subIII
microestrutura esquelética foliada sugere que a ordem Tentaculitida LJASCHENKO (1955) pode ser mais
 IX]
 propriamente incluída no Filo Mollusca.

Vamos agora reproduzir o diagrama do texto com os frasemas preenchidos apenas pelos núcleos temáticos, para maior facilidade de visualização.

1 2 3
 [| I II III IV V subII VI VII ||->[[[[] -subVI VIII subIII -IX subIII |=>| subIII -subIX|]
 4 5 6 7
 + | subIII IV V X]//| II XI III X]>[[| subI XII XIII subX XI VIII ||->[| subXII VII
 8 9 10
 XIII X X VI VII |=>[| XIII subVI XIV XI VIII |+| XIII VII VII VIII]]=>[| XI VIII | +
 11 12 13
 | VIII subVII XIV subVI | + | XIII]]=>| subVII subIII IX]]

Como se vê, obtivemos uma expressão “algébrica”, a qual pode ser “resolvida” de modo análogo a uma sentença matemática. Para tanto, vamos primeiramente resolver os *strings* mais internos (análise centrípeta). Observemos, por exemplo, que os frasemas 2 e 3 conectam-se por implicação, formando o *string* 2-3; este, por sua vez, conecta-se por conjunção ao frasema 4, formando o novo *string* 2-3-4, o qual se conecta ao frasema 5 por contrajunção. O *string* resultante 2-3-4-5 serve de pressuposto ao *string* seguinte, que vai de 6 a 12. Vejamos o que acontece nesse *string*.

Efetuada as conjunções 8-9 e 10-11-12, obtemos que o frasema 6, que é subtema do texto, desenvolve-se em 7, que implica 8-9, resultando em 7-8-9; este, por seu turno, implica 10-11-12, de modo que o desenvolvimento de 6 é o *string* que vai de 7 a 12. Ligando-se o subtema ao seu desenvolvimento, obtemos o *string* que vai de 6 a 12. Este, conforme vimos, se conectava ao *string* 2 a 5 por pressuposição, resultando daí o *string* 2 a 12. O frasema 1 é o tema principal do texto, que se desenvolve em 2-12,

desembocando por implicação no frasema 13, que é a conclusão do texto. Podemos representar esquematicamente essas passagens da seguinte maneira:

$$[[1]] \rightarrow [[[[2] \Rightarrow [3]] + [4] // [5]] \rightarrow [6] \rightarrow [[7] \Rightarrow [8] + [9]] \Rightarrow [10] + [11] + [12]]] \Rightarrow [13]$$

$$[[1]] \rightarrow [[2-3-4-5]] \rightarrow [6] \rightarrow [7-8-9] \Rightarrow [10-11-12]] \Rightarrow [13]$$

$$[[1]] \rightarrow [[2-3-4-5]] \rightarrow [6] \rightarrow [7-8-9-10-11-12]] \Rightarrow [13]$$

$$[[1]] \rightarrow [[2-3-4-5]] \rightarrow [6-7-8-9-10-11-12]] \Rightarrow [13]$$

$$[[1]] \rightarrow [2-3-4-5-6-7-8-9-10-11-12]] \Rightarrow [13]$$

Este último *string* resume o próprio texto: é seu *string* principal. Nota-se que sua estrutura básica consiste em um tema ao qual se segue um desenvolvimento, que por sua vez vai implicar numa conclusão que retoma o tema do início. Essa estrutura é comum à maioria dos textos dissertativos. De modo geral, chamaremos de *estrutura canônica* de um texto a esse esquema que obtemos após completada a análise centrípeta. Podemos dizer então que o texto dissertativo se caracteriza por uma estrutura canônica do tipo $[[1]] \rightarrow [] \Rightarrow []$, isto é, introdução-desenvolvimento-conclusão. Deve ser possível, aliás, estabelecer uma tipologia de textos com base na análise de sua estrutura canônica subjacente.

O texto objeto desta análise apresenta ainda uma outra peculiaridade: o *string* 6 a 12 também tem uma estrutura canônica semelhante, o que revela um procedimento bastante comum na dissertação, a saber, o encaixe de um raciocínio conclusivo dentro de outro, de forma recorrente. Assim, cada nova conclusão serve de ponto de partida a um novo raciocínio.

Quanto aos núcleos temáticos, pode-se observar nitidamente no texto em questão a tensão repetição/renovação que garante a informatividade do texto. Verifica-se também como os núcleos temáticos se enlaçam, servindo de mola mestra das conexões. Por exemplo, o texto apresenta como pressuposto (*string* 2 a 5) o desconhecimento da estrutura esquelética dos Tentaculídeos devido à ausência de vestígios diretos deles e

sua conseqüente dificuldade de classificação filogenética. As palavras-chave desse argumento são *desconhecimento da estrutura esquelética* (–subVI), *reconhecimento dificultado* (–IX) e *posicionamento sistemático problemático* (–subIX). Entretanto, os autores anunciam a realização de testes que comprovam que a amostra estudada apresenta vestígios diretos do esqueleto de *Tentaculites* e não apenas vestígios indiretos (impressões sobre a rocha). Aparece então no texto o núcleo VI/subVI, antagônico de –subVI. Conseqüentemente, também o núcleo –IX/–subIX tem seu valor vazio preenchido por IX (*Filo Mollusca*). O texto relata, portanto, a passagem do desconhecimento ao conhecimento. Trata-se da narrativa da descoberta, típica do discurso científico.

Conclusão

Como já alertamos anteriormente, nem todas as conexões possíveis entre os elementos constitutivos de um texto foram aqui relacionadas. Esse objetivo demandaria a pesquisa sobre um *corpus* muito maior e mais variado, o que foge à nossa possibilidade no momento. O que se pretendeu foi, antes, apontar para a possibilidade de estender ao nível transfrástico as regras sintáticas, o que permite, em conseqüência, a abordagem do problema da análise de textos por meio de procedimentos formais. Isso tem importantes implicações, como, por exemplo, a possibilidade de tratamento automático de textos, quer por uma aplicação cibernética do modelo na análise textual quer pela dotação dos computadores de uma “competência textual”, passo fundamental para o desenvolvimento de Inteligência Artificial (IA).

À possibilidade de estabelecer uma tipologia de textos com base na estrutura canônica já fizemos menção. Mais ainda, somos levados a supor que tais estruturas subjacentes, bem como as conexões que as sustentam, façam parte da gramática universal proposta por Chomsky. Afinal, ao passarmos de uma língua para outra, muda a sintaxe, muda o léxico, muda a rede fonológica, a rede semântica e, enfim, a própria visão de mundo subjacente. Entretanto, encontramos procedimentos de conjunção, disjunção, implicação, desenvolvimento, síntese, etc., reveladores da estrutura profunda dos atos comunicativos, que poderíamos identificar com a própria estrutura do pensamento, como querem os lógicos. No entanto, somente o estudo de *corpora* lingüísticos cada vez mais extensos poderá revelar o quão universais são tais regras. De

todo modo, o que se busca não é uma sintaxe universal, mas um modelo universal de descrição sintática, capaz de dar conta das peculiaridades gramaticais de toda e qualquer língua natural. Isso não significa, portanto, que a todas as línguas subjaz uma mesma visão de mundo – muito pelo contrário –, mas que todas as visões de mundo obedecem a um mesmo princípio modelizador. Dito em outras palavras, os recortes culturais variam de uma língua para outra, mas o modo de recortar permanece o mesmo.

Referências bibliográficas

- BEAUGRANDE, R. de; DRESSLER, W.U. (1988) **Introduction to text linguistics**. Londres: Longman.
- BERNÁRDEZ, E. (1982) **Introducción a la lingüística del texto**. Madrid: Espasa/Calpe.
- BORBA, F. S. (1979) **Teoria sintática**. São Paulo: T.A. Queiroz/EDUSP.
- CÂMARA JR., J. M. (1977) **Princípios de lingüística geral**. Rio de Janeiro: Padrão.
- CHARAUDEAU, P. (1972) Sens et signification. **Cahiers de Lexicologie**, n.º 21. Paris: Didier/Larousse.
- CHAROLLES, M. (1978) Introduction aux problèmes de la cohérence des textes. **Langue Française**, n.º 38. Paris: Larousse.
- HALLIDAY, M. A. K.; HASAN, R. (1975) **Cohesion in English**. Londres: Longman.
- MATEUS, M. H. M. et alii (1983) **Gramática da língua portuguesa**. Coimbra: Almedina.
- VAN DIJK, T. A. (1977) **Semiótica narrativa e textual**. São Paulo: Cultrix/EDUSP.
- _____ (1983) **La ciencia del texto**. Barcelona/Buenos Aires: Paidós.