ANÁLISE SEMIÓTICA DE SINAIS FOTOMÉTRICOS. II - CONSTRUÇÃO DE REDUNDÂNCIAS DE ORDEM SUPERIOR

Jorge de Albuquerque Vieira
Observatório do Valongo - UFRJ

ABSTRACT

We present the computation of the redundancies of higher levels in a natural language outlined in photometrical signals.

The semantical content refers to aspects of the local pollution of the Rio de Janeiro sky.

A basic codification, the alphabet and the first order redundance were developed in previous papers.

Problems concerning the growth of alphabet and a quantitative analysis of tipical "words" in the phenomenous are also studied.

INTRODUÇÃO

Este trabalho tem por objetivo completar a análise semiótica à nível de sintaxe, aplicada em pesquisa de poluição urbana, como descrita em publicações anteriores (Refs. 1 e 2).

Nas referidas publicações, são descritos com detalhes o método, a observação e o instrumental utilizado, de modo que aqui nos limitaremos à análise do problema do estabelecimento de redundânciss de ordem superior à primeira, a qual já foi analisada anteriormente.

O objetivo perseguido é estabelecer parâmetros que caracterizam o comportamento de noites fotométricas, através do ruído nos registros fotométricos, tendo tal ruído seu comportamento relacionado ao da poluição atmosférica local.

A hipótese de trabalho, sugerida por observações anteriores (Ref. 2), é que o ruído fotométrico tem um conteúdo semântico que depende das condições climatológicas particulares e da poluição, sendo que isso implica uma estrutura sintática ao sinal.

Nosso esforço tem sido o de evidenciar a linguagem que contenha essas dimensões sintática e semântica.

Os passos que executamos a seguir envolvem:

- Determinação da faixa de influências intersimbólicas;
- Evolução da redundância e crescimento do alfabeto;
- Restrições sintáticas e mecanismos particulares.

DETERMINAÇÃO DA FAIXA DE INFLUÊNCIAS INTERSIMBÓLICAS

Nos trabalho já desenvolvidos, obtivemos um alfabeto para descrever o conjunto de sinais fotométricos, aliando a cada símbolo o parâmetro informação, o que leva à entropia do sinal e sua redundância.

Esta última é dita de primeira ordem porque foi obtida a partir de símbolos isolados, sem levar em consideração dependências entre os mesmos.

O estabelecimento de redundâncias de ordem superior gera a idéia de faixa de influência intersimbólica ou de alcance sintático (Ref. 3). A linguagem natural, não sendo otimizada, vai apresentar redundâncias que caracterizam sua sintaxe.

A busca por essa sintaxe consiste em tomar os símbolos progressivamente dois a dois (diagramas), três a três (triagramas), etc., considerando os arranjos assim gerados cada um como um único símbolo, calculando-se então a informação contida em cada arranjo, a entropia da população de arranjos e as respectivas redundâncias.

De acordo com o crescimento da ordem de redundância, ou seja, o comprimento da "palavra" composta pelos símbolos originais, notamos os seguintes fatos:

- O número de novos símbolos, gerados por esses arranjos, cresce exponencialmente.
- Quanto maior este número, mais os símbolos tornam-se progressivamente equiprováveis, com uma crescente tendência à independência estatística.

O crescimento do alfabeto vai depender dos vínculos sintáticos existentes e a tendência a equiprobabilidade significa a maximização da função entropia, logo a anulação da redundância, o que pode ser inferido pelas relações:

$$H = -\sum_i p_i \log_2 p_i \quad (1) \qquad R = 1 - H/H_{máx} \quad (2) \qquad H_{máx} = \log_2 n \quad (3)$$

onde H é a entropia, p_i é a probabilidade de ocorrência característica do símbolo s_i, R é a redundância e n é o número de símbolos.

Assim, o alcance sintático dentro do sinal é limitado superiormente e expresso pela geração de "palavras" de tal comprimento que a redundância tende a zero. Esse comprimento máximo de "palavra", além do qual os símbolos ocorrem aleatoriamente e de forma equiprovável, é que delimita o alcance sintático ou a faixa de influência intersimbólica.

Como em nosso trabalho os símbolos significam a duração em tempo na qual as flutuações do sinal ocorrem, essa faixa significa um certo alcance temporal que é estabelecido pelo número de símbolos, não por suas particulares durações: o parâmetro redundância é obtido a partir da entropia e esta depende da contribuição de todos os símbolos.

Assim, não há como especificar explicitamente qual é a faixa exata, em tempo, da influência intersimbólica; só podemos dizer,com exatidão, que há um número máximo de símbolos agregados por um vínculo sintático.

Apresentamos, no gráfico 1, os resultados do cálculo dessa faixa. Devido a grande quantidade de dados, mostramos apenas os resultados obtidos da análise das observações 3 e 4 (ver Ref. 2).

Estas observações foram realizadas na mesma noite, segundo a mesma linha de visada (sobre a cidade) mas sendo a 3 com filtro V e a 4 com filtro B. Esta noite, 26/06/78, apresentava na ocasião formação de névoa.

É um par de observações curioso porque, para esta particular direção, apresenta em uma mesma noite e grosseiramente mesma hora, visões completamente diversas como dadas pelos filtros, quando consideradas as redundâncias de primeira ordem.

O par apresenta redundâncias de primeira ordem extremas,a observação nº 3 com máxima redundância e a nº 4 com mínima.

Traçamos no gráfico a variação da redundância com as ordens, e também a variação da entropia. No caso, a importância maior e da redundância porque este parâmetro permite comparar efetivamente as observações.

Notamos que as redundâncias de quarta ordem indicam a tendência, nas duas curvas, ao limite da faixa sintática. Os erros associados crescem violentamente, como pode ser observado pelas barras, embora esse problema não ocorra no cálculo da informação por símbolo ou da entropia.

Por isso, com erros da ordem de 70% ou mais ligados à redundância de quarta ordem, consideramos os dois últimos pontos duvidosos mas podemos dizer que, provavelmente, a redundância de quinta ordem é bem próxima de zero e a faixa é, no máximo, de 5 símbolos. Ou seja, os vínculos sintáticos relacionam, no máximo, "palavras" de comprimento dado por 5 símbolos.

Um aspecto a ser considerado na sintaxe de cada sinal é que, apesar do sinal 4 partir de uma redundância de primeira ordem mínima, sua evolução é mais suave que a do sinal 3. A elevada redundância inicial deste último cai abruptamente de redundância segunda para a terceira, sugerindo que, talvez por razões físico-químicas, mecanismos responsáveis por esse nível sintático encontrem maior dificuldade em seu estabelecimento.

As duas curvas chegam na quarta ordem quase ao mesmo nível de redundância e com a mesma variação.

Para fins de uma estimativa, se recorrermos à Ref. 2,onde temos a duração em tempo dos símbolos individuais, podemos calcular a duração do símbolo médio de cada alfabeto, por meio das probabilidades, de ocorrência.

Essa duração é, para o sinal 3, de 0,318 segundos e para o sinal 4 é de 0,308 segundos. A estimativa das faixas recai, respectivamente, em 1,59 e 1,54 segundos.

Esse seria, em média, o vínculo sintático no tempo relacionando os mecanismos resultantes das fases componentes do sistema poluente, uma restrição em parte físico-química.

EVOLUÇÃO DA REDUNDÂNCIA E CRESCIMENTO DO ALFABETO

Na construção das várias ordens de redundância, necessitamos agrupar os símbolos em conjuntos sucessivamente maiores, aumentando o comprimento da "palavra". Com a formação dos arranjos entre símbolos do "texto", estes novos símbolos obviamente são em número maior, ocorrendo uma "dilatação" do alfabeto com o crescimento da ordem de redundância.

Se os símbolos ocorrem de forma independente, sem a presença de uma sintaxe, para um "texto" suficientemente longo poderíamos esperar todos os arranjos como previstos pela análise combinatorial.

A presença de uma sintaxe, contudo, restringe a formação livre destes arranjos, associando aos passíveis de formação probabilidades típicas de ocorrência.

No gráfico 2, mostramos o crescimento do alfabeto com a ordem de redundância. É visível que a redundância inicialmente menor do sinal 4 faz com que o crescimento do seu alfabeto seja mais lento que o verificado para o sinal 3 (um alfabeto menor gera menos combinações).

No final da curva, no entanto, notamos que o alfabeto do sinal 4 é levemente maior que o sinal 3: a redundância de ordem 4 encontra, no sinal 4, mais rapidamente o fim dos vínculos sintáticos, tornando os símbolos mais equiprováveis, com uma consequentemente maior diversificação entre eles. Ou seja, o sinal 4 é mais pobre em alfabeto inicial e em vínculos sintáticos.

Essa imagem é compatível com a visão fornecida pelo filtro B: partículas de menor tamanho, não conseguindo formar sistemas estáveis de maior tamanho. (Voltando ainda à Ref. 2, Fig. 3, verificamos como a distribuição da informação, na redundância de primeira ordem do sinal 4 parece seguir a do ruído, nos símbolos de maior duração).

A evolução do alfabeto ou o número de mensagens geradas em um texto é estudada em teoria da informação função do tamanho deste texto. Nossas observações não foram dimensionadas neste sentido, pois limitamos o tamanho dos registros ao mínimo para garantir erros razoáveis, já que ao que tudo indica, um tempo muito longo de observação poderia permitir mudanças drásticas nas condições do sistema observado. Estas mudanças talvez possam, não sabemos ainda com certeza, comprometer a análise da sintaxe, levando a conclusões sobre médias de condições altamente diversas de poluição (por exemplo, o surgimento súbito de uma coluna de fumaça de uma fábrica é completamente destoante da "homogeneidade" local).

Textos maiores permitiriam partir de um comprimento mínimo para observar como as mensagens crescem em número com o crescimento do texto, o que fornece uma constante característica da linguagem logo de sua sintaxe (Ref. 3).

Apesar desta relação vir sendo contestada em trabalhos recentes (Ref. 4), seria importante sua aplicação em nosso estudo.

RESTRIÇÕES SINTÁTICAS E MECANISMOS PARTICULARES

Como visto no item II, a função entropia contém a informação de toda a coletividade dos símbolos, não mostrando como símbolos, logo mecanismos particulares, podem ter importância na sintaxe.

Falamos de mecanismos típicos que podem estar caracterizando as condições ambientes e cuja influência é diluída pelo cálculo envolvendo todos os símbolos.

Uma maneira de evidenciar a importância destes processos é verificar sua grande (ou pequena) informação, frente aos demais.

Um mecanismo destacado por informação máxima significaria um evento raro, atípico, um desvio de condições básicas, etc. Já uma informação mínima caracterizaria o mecanismo comum, típico indicando talvez um vínculo preferencial imposto por restrições físico-químicas e demais restrições ambientais e climatológicas.

É importante frisar que, quando consideramos os diversos comprimentos de "palavra", estamos considerando a associação de vários mecanismos na geração de um composto.

De maneira geral, na construção de redundâncias de ordem n, a notação que utilizamos é:

$s_d = s_i s_j \ldots s_k$, tal que $i+j\ldots+k = d$, onde os índices i,j,k etc., são em número de n.

Ou seja, as durações individuais i,j,k etc., agrupam-se gerando uma duração composta d.

Esta associação pode ser feita de várias formas, podendo indicar uma "preferência" na maneira como os mecanismos se associam o que reflete a sintaxe da linguagem e as restrições ou leis que interessam à semântica.

Podemos ter, por exemplo, um símbolo s_{16} gerado por $s_{10}s_2s_4$, s_8s_8, $s_8s_2s_6$, etc. O mecanismo final, expresso por s_{16}, é gerado por restrições que "obrigam" certos mecanismos primários a agrupamentos que tenham duração final constante.

Para evidenciar qual o mecanismo composto mais típico ou raro no sinal e qual a combinação mais típica ou rara que o gera, traçamos os gráficos 3, 4, 5 e 6 (ver Apêndice).

Mais uma vez, nos restringimos as observações 3 e 4, pelas razões já expostas.

No gráfico 3, consideramos nas abcissas todas as durações compostas d, obtidas nas citadas observações e nas quatro ordens de redundância. Sem discernir os vários arranjos, colocamos nas ordenadas as frequências de ocorrência destas associações, em %.

Obtemos 4 pares de curvas que apresentam tendências a "picos" em certos valores de d.

Estas curvas lembram distribuições maxwellianas e é visível que as curvas pontilhadas, representando o sinal 4, são bem mais "comportadas" que as do sinal 3.

Desvios nas curvas ocorrem em valores altos de d, como em d=8, 11 e 13. Os máximos das curvas deslocam-se de duas em duas unidades, no caso do sinal 4, já no sinal 3 o deslocamento para as 3 últimas ordens de redundância é de 3 em 3 unidades.

Os desvios observados podem representar afastamentos de condições padrões, sendo necessário tentar evidenciar qual a combinação de mecanismos que os geram.

Para isso, vejamos os gráficos 4, 5 e 6:

Como a quantidade de dados é muito grande, estes gráficos são amostras para exemplo. Foram construídos colocando-se no eixo das abcissas as combinações de símbolos básicos que geram mecanismos compostos e no eixo das ordenadas a informação contida nestes mecanismos.

Assim, em 4, temos exemplos de diagramas com durações típicas - procuramos apresentar, para efeito de comparação, durações extremas, de pequeno e grande valor, assim como contribuições dos dois sinais estudados.

Temos, portanto, s (símbolo), I (informação), d (duração). O número contido no círculo indica a observação.

No gráfico 5 temos o mesmo para os triagramas e em 6, para tetragamas.

Nas curvas, as "depressões" indicam baixa informação logo mecanismos mais ou menos comuns; os "picos" representam combinações mais raras; podemos pois inferir, para uma determinada duração final d qual o arranjo mais provável ou mais raro, caracterizando o "peso", a importân— cia, do mecanismo.

O que podemos notar, além deste aspecto, é que quando d aumenta, as curvas tendem a retas paralelas ao eixo das abcissas e com valor máximo de informação, ou seja, estes mecanismos são cada vez mais raros e equiprováveis no sentido de independentes.

Os mecanismos de maior duração são importantes porque não parecem depender da contribuição do ruído instrumental, como sugerido na Ref. 2. Caracterizam bem os processos poluentes tal que um desvio em uma curva de valor d grande evidencia um mecanismo rico em importância semântica.

No gráfico 3, assim como nestes últimos, picos máximos das curvas ou "depressões" máximas podem estar relacionados com a presença de símbolos de duração menor, como os do ruído instrumental, tal que tais estruturas podem estar somente indicando a "contaminação" devida a este ruído.

No entanto, notamos nos gráficos 4, 5 e 6 certos mecanismos muito presentes e que não são gerados pelos símbolos típicos do ruído (do ponto de vista da totalidade dos símbolos - isso distingue um mecanismo expresso pela combinação 1114 de um expresso por 4531).

Estes mecanismos parecem ser intrínsecos das condições exteriores, já que o ruído instrumental é caracterizado com grande preponderância pelo símbolo s_1, o de duração mínima.

CONCLUSÕES

O estabelecimento dos aspectos sintáticos apresentados é fundamental para a construção da semântica ligada aos sinais fotométricos comprometidos pela poluição, ou seja, é fundamental para uma "leitura" do significado dos dados, na descrição das condições poluentes.

A construção da semântica, em termos metodológicos, depende de um acervo de conhecimento já estabelecido, ou seja, deve ser feita com o apoio de conhecimentos já enquadrados em teorias.

Uma visão sintética, que implica a construção de um modelo é mais poderosa que a visão observacional e consequente descrição.

É importante frisar que:

a) os aspectos delineados nos itens II, III e IV podem ser tomados, com o devido cuidado, como protótipos de leis de baixo nível, para a necessária elaboração teórica.

b) Estes aspectos foram obtidos por meio de uma visão sintática como dada pela teoria da informação. Acreditamos que a questão dos mecanismos compostos, apresentada no item IV, transcende o alcance desta visão.

A análise de símbolos simples ou compostos, individualmente, parece requerer a ajuda de teorias de linguística, como aplicadas às linguagens naturais.

Nossa tentativa teórica deve partir inicialmente da mecânica estatística, em termos prioritários, mas com uma ajuda paralela da linguística e ótica da atmosfera.

BIBLIOGRAFIA

VIEIRA, J.A. (1980). "Linguagem Natural e Lei Científica - Um Estudo em Fotometria Astronômica". An. Acad. Brasil. Ciências, 52(2): 235.

VIEIRA, J.A. (1980). "Semiotical Analysis of Photometrical Signals". An. Acad. Brasil. Ciências, 52(3): 467.

GOLDMAN, S. (1968). "Information Theory". Dover Publ. Inc., New York.

MALUF, U.M.M. (1978). "Irrational Metrics and Behavioral Incommensurability: a Framework for Speculation". Fundação Getúlio Vargas.

AGRADECIMENTOS

Esse trabalho foi possível graças ao auxílio prestado pelos Convênios nº 4.3.83.0290.00, FINEP e nº 605/83, FUJB.

APÊNDICE

Para não sobrecarregar os gráficos, com demasiados algarismos, vamos adotar as seguintes convenções: nos gráficos 5 e 6, onde surgem símbolos com vários índices, segundo a notação $S_d = S_i S_j \ldots S_k$ tal que i+j+...+l=d, faremos: Letra maiúscula = ij...k.

Gráfico 5

A = 011	M = 132	Z = 320	L1 = 526
B = 101	N = 130*	A1= 321	M1 = 535
C = 110	O = 141	B1= 411	N1 = 562
D = 111	P = 140	C1= 401	O1 = 553
E = 010	Q = 203	D1= 148	P1 = 634
F = 014	R = 213	E1= 256	Q1 = 724
G = 023	S = 222	F1= 265	R1 = 733
H = 032	T = 230	G1= 364	S1 = 0*13
I = 041	U = 231	H1= 346	
J = 104	V = 302	I1= 445	
K = 114	X = 30*1	J1= 463	
L = 123	Y = 312	K1= 472	

Gráfico 6

(3) d = 4 e 5 e (4) d = 4 e 5

A = 0111	F = 0112	K = 1120	P = 2011	U = 1012
B = 1011	G = 0210	L = 1121	Q = 2101	V = 1021
C = 1101	H = 0211	M = 1201	R = 2110	X = 2010
D = 1111	I = 1102	N = 1210	S = 2111	
E = 1110	J = 1112	O = 1211	T = 0121	

(3) e (4) d = 6

A = 0113	I = 1013	Q = 1131	Z = 2020	H1 = 2201
B = 0130	J = 10*21	R = 1202	A1 = 2021	I1 = 2210
C = 0131	K = 1022	S = 1212	B1 = 2012	J1 = 3011
D = 0211	L = 1031	T = 1220	C1 = 2102	K1 = 3111
E = 0212	M = 1113	U = 1221	D1 = 2112	L1 = 3101
F = 0221	N = 1103	V = 1301	E1 = 2120	M1 = 3110
G = 0310	O = 1122	X = 1310	F1 = 2121	
H = 0311	P = 1130	Y = 1311	G1 = 2211	

GRAFICO I

EVOLUÇÃO DAS REDUNDANCIAS

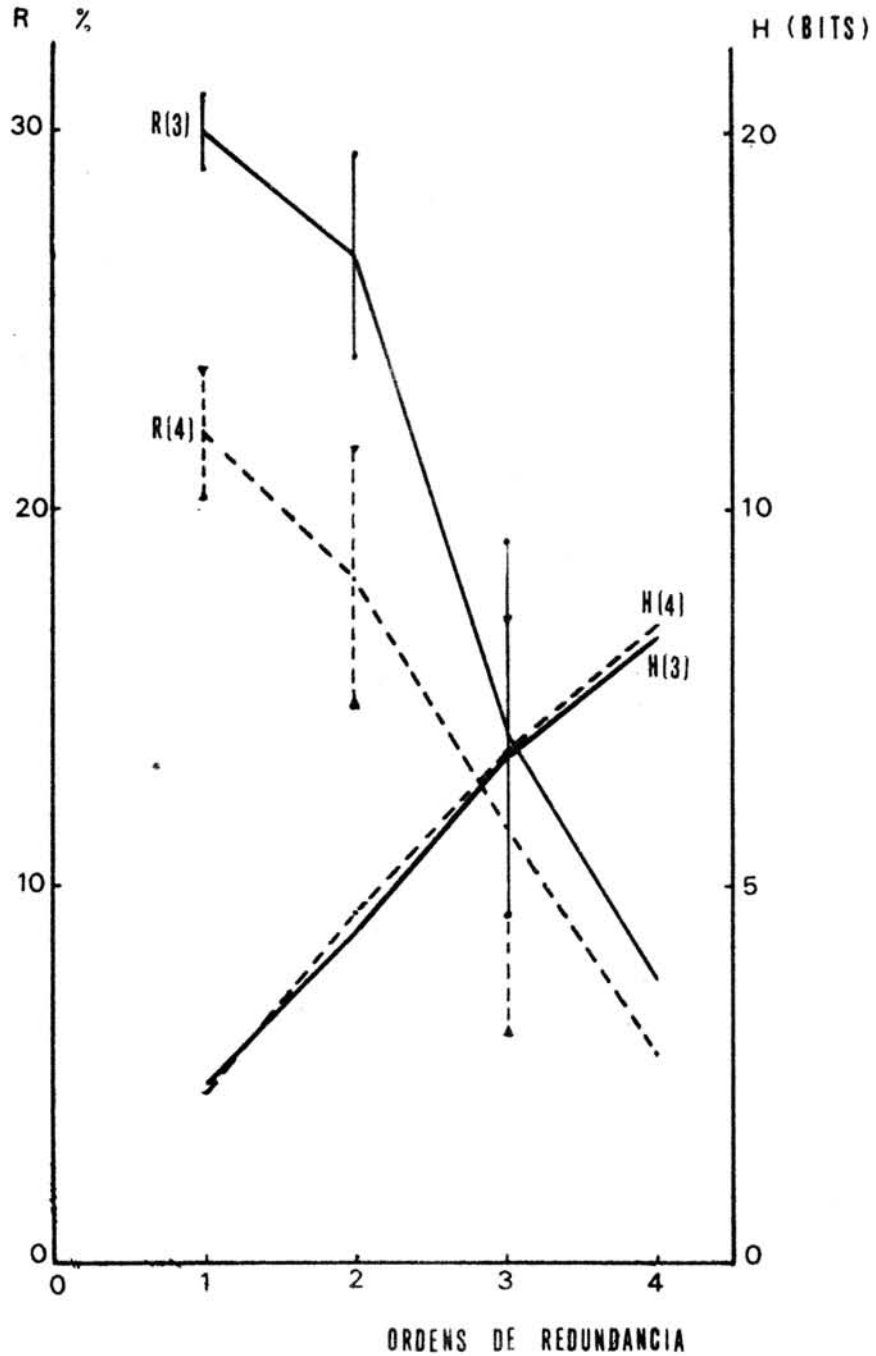


GRAFICO 2

CRESCIMENTO DO ALFABETO

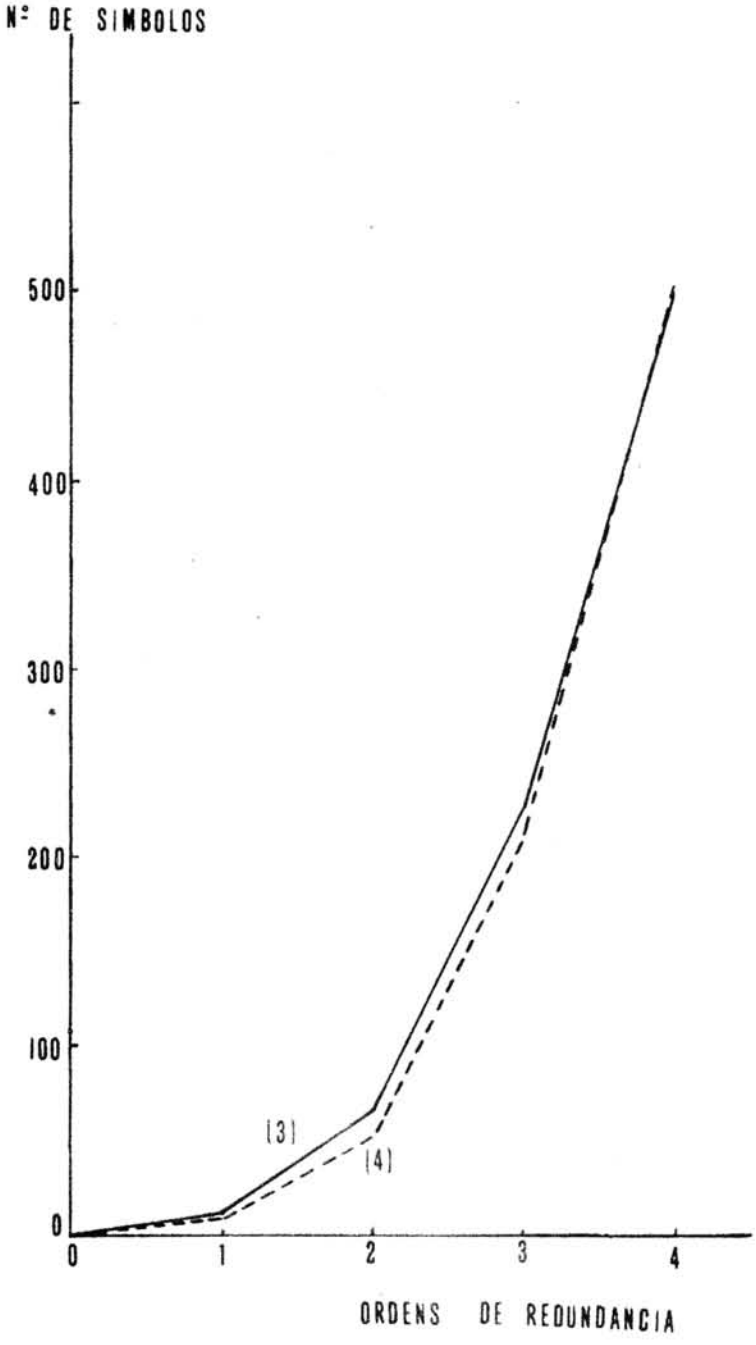


GRAFICO 3

MECANISMOS COMPOSTOS NAS VARIAS REDUNDANCIAS

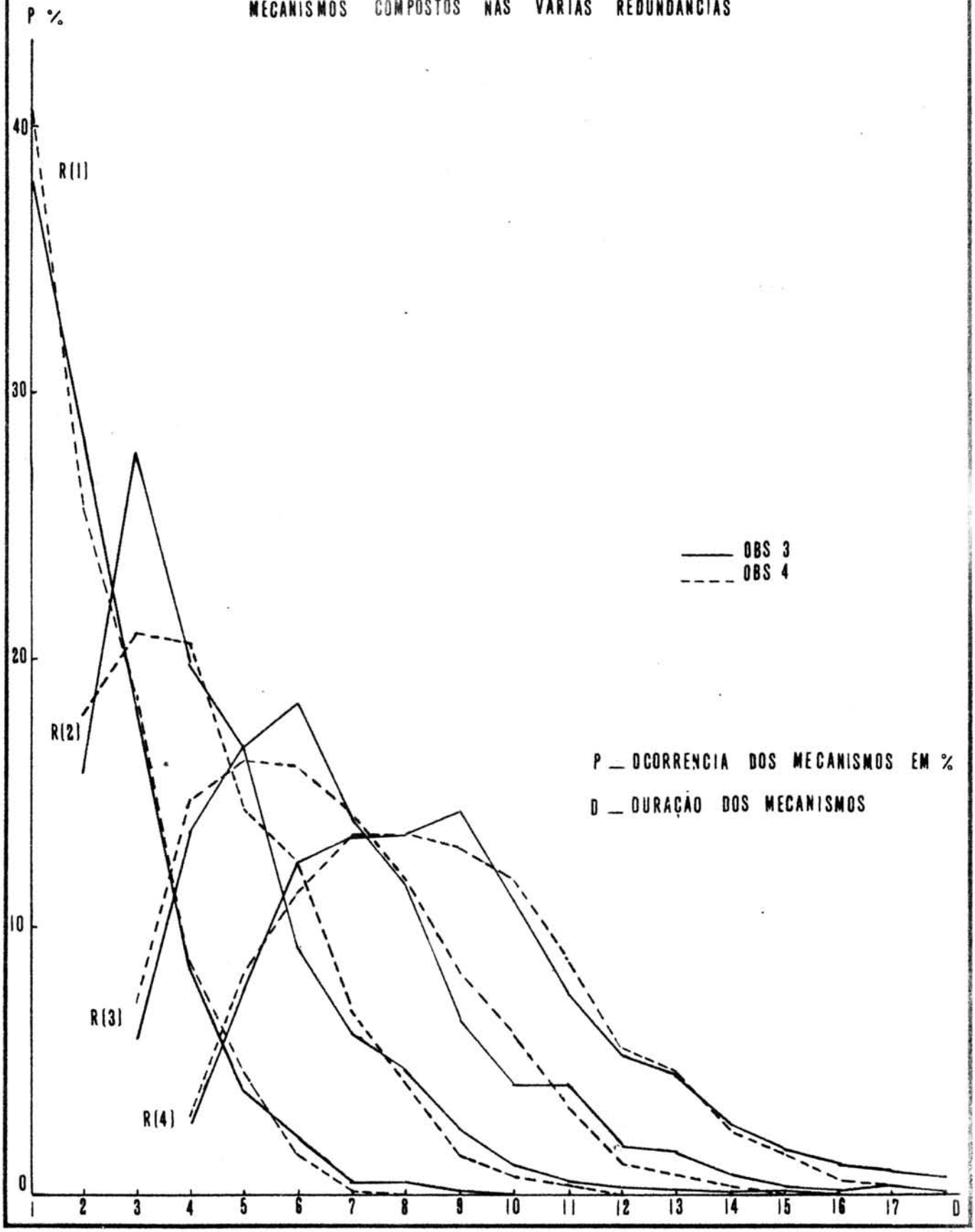
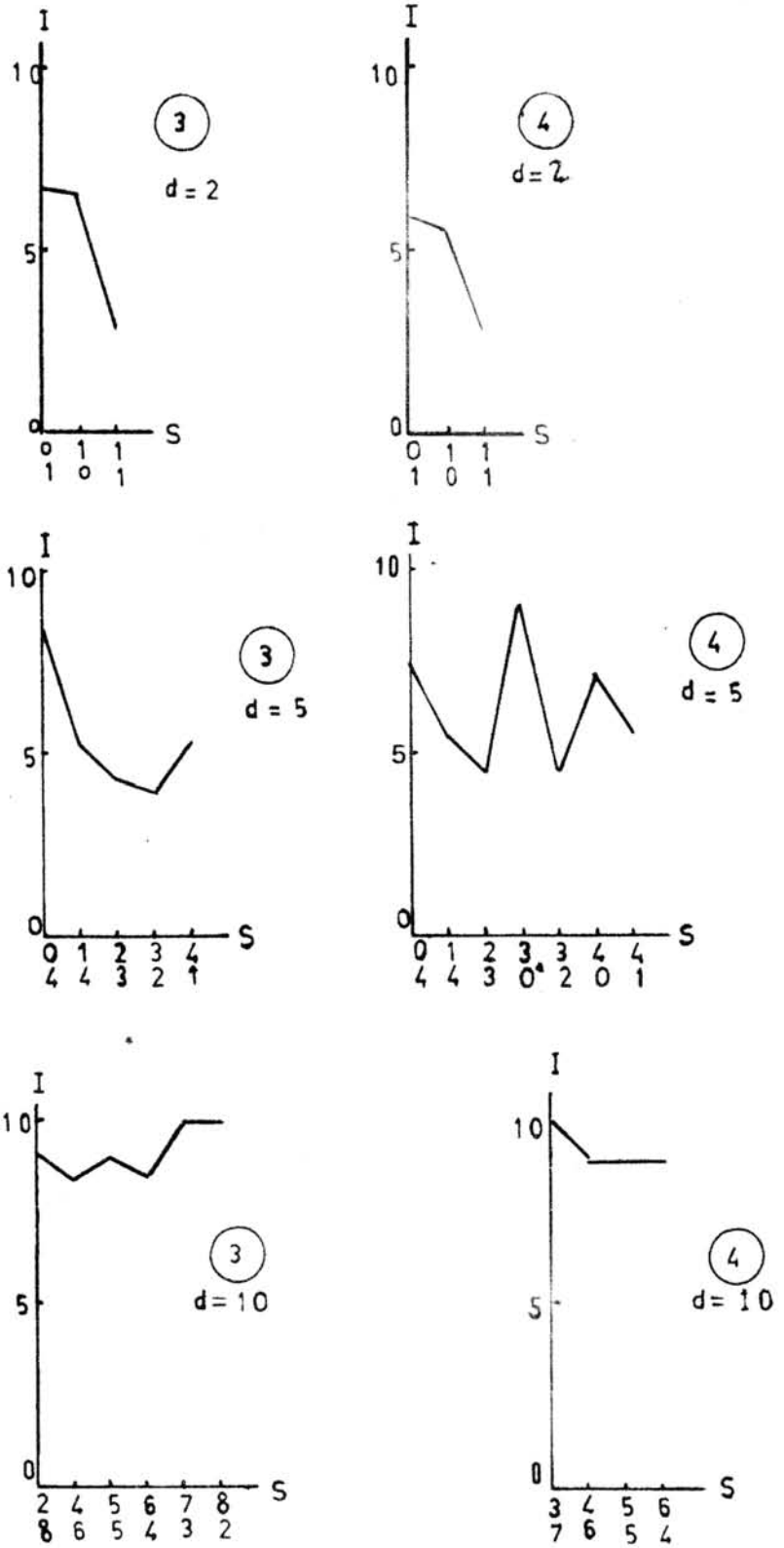


GRAFICO 4

MECANISMOS COMPOSTOS PARTICULARES

DIAGRAMAS



I _ BITS

GRAFICO 5

MECANISMOS COMPOSTOS PARTICULARES

TRIAGRAMAS

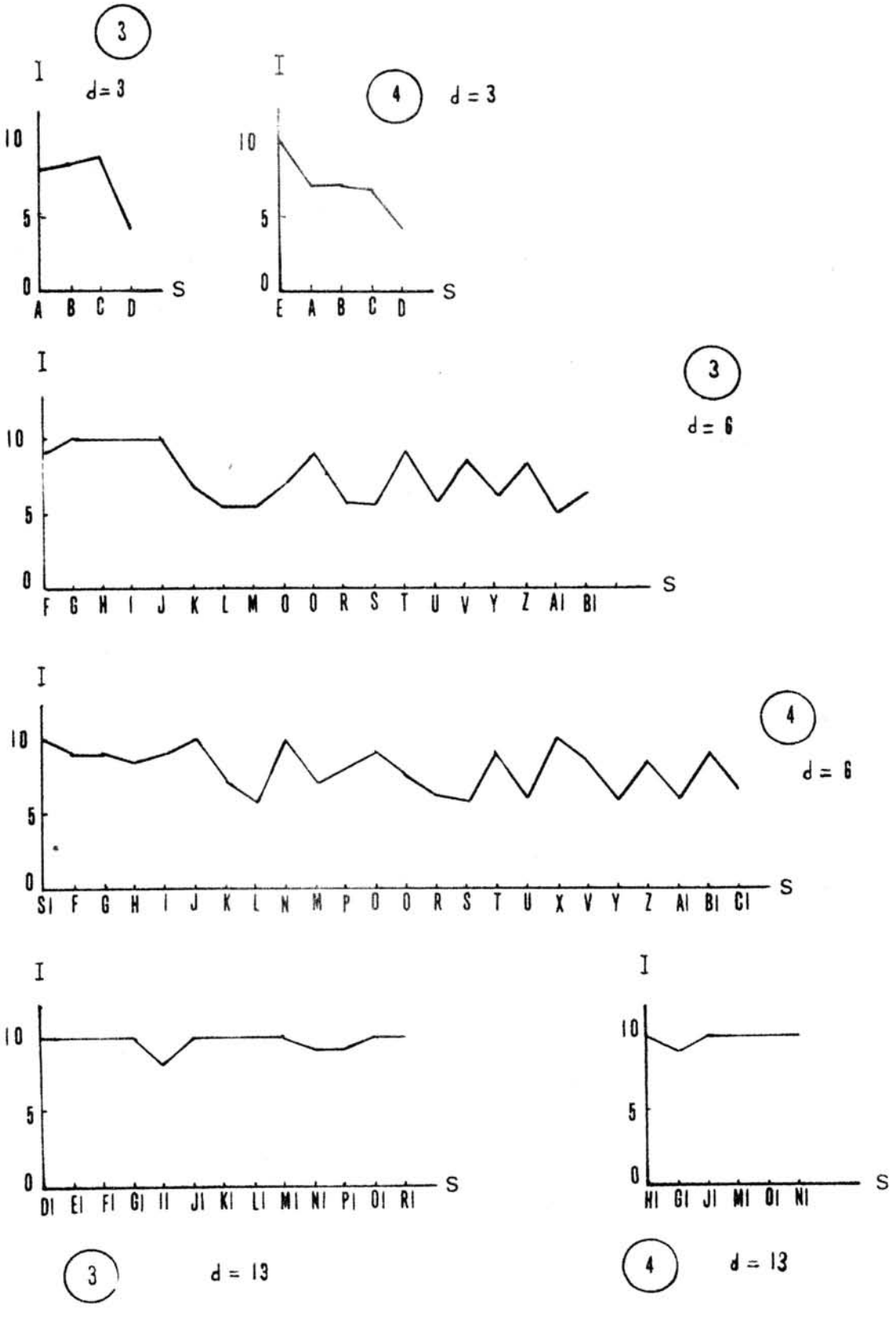


GRAFICO 6

MECANISMOS COMPOSTOS PARTICULARES

TETRAGRAMAS