

Índices Alfabéticos da Classificação Decimal Universal

Eimar Ohman
Täby

Tradução de Giorgio Bullaty
Revisão de Padre Astério Campos
Membro da IBBD/CDU

RESUMO

Discussão em torno dos índices alfabéticos da classificação Decimal Universal — CDU — uma vez que diferentes trabalhos apresentados, nos últimos anos, tem abordado problemas referentes às tabelas. O estudo dos índices é feito nas edições abreviadas, desenvolvidas e médias.

A classificação Decimal Universal tem sido muito discutida e criticada nestes últimos anos (1-5). Mas essa discussão tem versado, quase exclusivamente, sobre as tabelas sistemáticas. Dos índices alfabéticos quase não se fez menção. Por outro lado, fizeram-se, durante estes anos, interessantes publicações sobre os métodos de elaboração de fichários e registros de palavras-chaves, que são efetivamente indispensáveis aos usuários (6-9), assim como sobre a confecção mecânica dos índices da CDU (10). Neste artigo serão tratados apenas os índices alfabéticos impressos.

Gostaria de começar com duas citações:

"Este índice alfabético, que é a parte mais importante do sistema, consta de títulos tirados de uma multiplicidade de fontes, cuja aceitação os usuários consideraram oportuna no decurso de 54 anos".

"À parte sistemática das tabelas da CDU costuma-se juntar uma parte alfabética. Mas esse registro serve, apenas, como orientação geral e aproximada do lugar em que se deve procurar na parte sistemática."

A primeira citação é de Melvil Dewey, tirada da introdução por ele escrita, no ano de 1926, e repetida nas edições posteriores da "Classificação Decimal de Dewey" (CDD). A segunda deriva da "10.ª Edição Abreviada Holandesa da CDU" (1970). Trata-se de duas concepções essencialmente diversas sobre a importância dos índices alfabéticos.

Tenho a impressão que, entre os especialistas em CDU, existe uma subestima dos índices alfabéticos, como mais ou menos está expresso na segunda das citações acima mencionadas. Nada há, evidentemente, que objetar contra a seguinte advertência que se encontra na Edição Média Alemã da CDU: "A experiência de muitos anos adverte que jamais devemos efetuar classificações exclusivamente pelo índice". Com muita frequência ouvem-se também opiniões como esta: "Nunca utilizo o índice alfabético; quem conhece bem as tabelas hierárquicas pode arranjar-se exclusivamente com elas". Este é um exagero que não ajuda aos principiantes. Desejo mencionar aqui uma citação da mesma introdução de Dewey: Mesmo o usuário, que sabe exatamente onde encontrar nas tabelas o material procurado, pode, pela consulta ao índice, descobrir material afim, que, de outro modo, lhe passaria despercebido". Será esta afirmação válida em relação aos índices atualmente existentes nas várias edições da CDU? Não o é sempre, como a seguir demonstraremos; existem, além disso, edições sem índice.

Os conceitos afins, de que fala Dewey, são por ele chamados "distributed relatives" e em alemão receberam o nome de "verstreute Verwandte" e em português chamaremos de "afins dispersos". Quando, a seguir, estabelecer discussão sobre os índices de assunto, estarei tentando responder à pergunta sobre a maneira como esses "parentes pobres" são considerados nos índices da CDU. É melhor começarmos pela edição abreviada trilingue (1958), na qual se encontram, lado a lado, três índices alfabéticos de assunto (alemão, inglês e francês).

Tradução autorizada pela FID. Artigo publicado em DK-Mitteilung 16 (2): 7-10, 1972.

Edições Abreviadas

A edição abreviada trilingue é muitas vezes considerada padrão das edições abreviadas. Entretanto, os três índices respectivos não se correspondem entre si como era de esperar. Isto pode ser explicado por razões lingüísticas, mas só em parte. Vê-se, por exemplo, à primeira vista, que o índice francês é muito mais curto que os outros dois: 56 páginas contra 83. Outras diferenças serão mencionadas a seguir. O índice inglês foi elaborado, em grande parte, segundo as idéias de Dewey. Quando consultamos, por exemplo, a palavra "Furnaces" (Fornos), encontramos imediatamente remissiva para dois outros termos: "Ovens" e "Kilns", que, na realidade, podem ser considerados como quase-sinônimos. Na palavra "Ovens" encontramos os seguintes números de classificação que não se encontram no termo "Furnaces" (é traduzido aqui o texto alemão da edição abreviada trilingue):

- 643.33 Fornos. Fogões. Aparelhos de cozinha e de aquecimento
 666.3.041 Fornos. Cozimento (Cerâmica)
 683.95 Fogões e fornos de cozinha (Ferragens)

Em "Knills" encontramos de novo 666.3.041; este termo é muitas vezes usado em inglês para indicar fornos da cerâmica.

Poder-se-ia talvez observar que as remissivas mencionadas não são completas. Se tivéssemos encontrado também "Stoves", teríamos chegado a 683.94, o que é reproduzido no texto alemão com "Ofen". Além disto teríamos gostado de encontrar referências a "Heaths" e "Grates". O último termo nos leva a 662.932 Grelhas. Apesar disto, as remissivas inglesas são muito úteis.

Não há cousa parecida nos índices alemães e franceses. Se em "Ofen" encontrássemos, p. ex., referências a "Herde" (fogões), "Hochófen" (altos-fornos), "Kamine" (chaminés) e "Kuchenöfen" (fornos de cozinha), assim como em "Fours" (fornos) referência a "Fourneaux", "Pöeles", etc. aumentaria consideravelmente o valor dos dois últimos índices mencionados. Causam estranheza as lacunas do índice alemão. Em "Ofen" só encontramos três números, 621.783, 66.041 e 669.041. Sente-se aí a falta de 683.9 Fornos. Aparelhos de aquecimento. Mas, continuemos com o termo "Furnace" do índice inglês. Encontramos, então, duas importantes remissivas, não para quase-sinônimos, mas para conceitos afins. A mais importante, sem dúvida, é a de 662.9 Tecnologia da combustão. A seguir, temos em 66.043, portanto em materiais de construção para fornos, uma referência a 666.76 Materiais refratários. Os leitores poderão chamar a atenção para o fato que nas tabelas existem remissivas e estas podem ser suficientes. Voltarei a isto mais tarde; aliás, na edição abreviada as

remissivas são bem poucas. Não existem, então, remissivas no índice alemão da edição abreviada trilingue? Existem, mas são poucas. Por outro lado, podem ser descobertas com facilidade por serem impressas em itálico. Em média, porém, só existem três destas remissivas para cada página de três colunas. Elas se dividem, por partes quase iguais, em dois tipos diferentes, como os exemplos seguintes indicam:

- Abbildungen (Gravuras) *siehe auch* (ver também) Lichtbilder (Fotografias), Photographien, Portraits
 Ackerböden (Solo arável) *siehe* (ver) Bodenkunde
 Ciência do solo)

No primeiro caso ("Ver também") trata-se de remissivas para palavras-chaves usadas no índice com o mesmo significado, tais como se encontram em grande quantidade no índice inglês (indicadas com "Cf. " = Ver também). Para uma, duas ou talvez três remissivas de "Ver também" de cada página do índice alemão existem de 20 para trinta no índice inglês. No segundo caso, ao invés, trata-se de termos que não se encontram nas tabelas sistemáticas, mas provavelmente são procurados por muitos usuários. Faz-se então referências a uma palavra-chave que leva ao ponto exato nas tabelas. Também este tipo de remissiva aparece mais frequentemente no índice inglês do que no alemão, mas a diferença não é tão marcante. No índice francês não existem remissivas.

Edições desenvolvidas

Todos os utilizadores da CDU têm, por motivos lingüísticos, que limitar-se a duas ou, talvez, três edições. Por motivos óbvios se dá preferência à língua materna. Quando, por exemplo, se trata de um sueco, como é o caso do autor, e não existe edição desenvolvida em sueco, é forçoso fazer uso de edições em língua estrangeira, nos casos em que a edição abreviada não é suficiente. Devo limitar-se às edições desenvolvidas alemã, inglesa e francesa.

Os índices da segunda edição francesa e da terceira alemã têm, aproximadamente, 40 e 20 anos de idade respectivamente. Estão, evidentemente, antiquados. Contudo, deve-se admitir que o trabalho de Carl Walther, que se encontra no índice da edição desenvolvida alemã, presta ainda bons serviços. Se, através deste índice, somos conduzidos a um número já eliminado, há sempre a possibilidade de encontrar o número novo, sendo a maneira mais fácil de fazê-lo a consulta ao "Comulated UDC Supplement", publicado pela FID em 1967. Aí a maioria dos números é acompanhada de texto alemão. Consideravelmente mais complicada é a utilização da segunda edição francesa, há muito esgotada e muito desatualizada. Infelizmente, agora está esgotado também o índice alemão.

ÍNDICES ALFABÉTICOS DA CLASSIFICAÇÃO DECIMAL UNIVERSAL

A edição desenvolvida inglesa não tem, infelizmente, índice uniforme e completo. Esta edição é publicada em fascículos, sendo que a maioria deles, mas não todos, tem índice alfabético resumido. Merecem especial interesse os índices dos fascículos publicados seguidamente a partir de 1968. Tomo como exemplos o fascículo que contém a classe 618 *Gynaecology. Obstetrics* e o que contém a classe 636/639 *Animal Breeding. Animal Produce. Hunting. Fislting*. Encontramos, naturalmente, aqui, como em geral nos índices antigos das tabelas desenvolvidas, forte ligação com a terminologia das tabelas sistemáticas. Mas os seguintes fatos devem ser observados.

Encontram-se remissivas "Ver" e "Ver também". Por exemplo:

Insemination, artificial *see* (Ver) Artificial insemination

Abnormalities *see also* (Ver também) Anomalies
Além disso, encontram-se, às vezes, sinônimos dos termos usados no texto da tabela. Exemplo:

na tabela 618.156 Wall between urinary bladder and vagina

no índice "Paries vesico-vaginalis" 618.156.

Deparamos, então, com dispositivos, mesmo que não sejam numerosos, introduzidos com a finalidade de facilitar a utilização das tabelas. Por outro lado, a separação dos índices importa numa grande limitação das remissivas. De 618 nunca se faz remissiva para outras partes de 61 Medicina. Para descobrirmos, p. ex., que inseminação artificial se encontra também em 636, devemos, examinar o índice de 636/639.

Os conhecedores da língua russa podem tirar proveito do índice da edição desenvolvida russa publicada em 1970. Consta de dois tomos, tendo no total mais de setecentas páginas. Pelo que sei, não se tenciona publicar, no momento, novo índice da edição desenvolvida alemã. Os usuários de língua alemã devem, por conseguinte, contentar-se, para o futuro, com o índice da edição média.

Edições médias

No momento, só existem três edições médias da CDU: a alemã (DK-Handausgabe), a francesa (Edition Moyenne) e a russa (Srednee Izdanie). Sobre esta última não me é possível, infelizmente, entrar aqui em detalhes. As duas primeiras foram publicadas em 1967. Já no ano seguinte foi editado o segundo volume que é o índice alfabético da edição alemã. O respectivo índice francês ainda não foi publicado. Devo, por conseguinte, limitar-me ao índice da edição média alemã. Este índice foi preparado por meio de um computador eletrônico. Os métodos utilizados acham-se descritos no prefácio de Karl Fill ao índice. Outros detalhes podem ser fornecidos pelas diversas publicações da "Zentralstelle für maschinelle Dokumentation"

(Órgão Central de Documentação por meio de máquinas) (10). Hoje parece natural que os índices **futuros** sejam elaborados por meio de máquinas. É, portanto, também desejável que esta primeira tentativa seja submetida a um exame crítico. Esta edição apresenta, inegavelmente, muitas falhas, que não podem, de certo, ser atribuídas à máquina. Elas são, antes, devidas à insuficiente elaboração do material apresentado.

Antes de tudo, é de lamentar que faltem aqui completamente as remissivas ("Ver" também) existentes na edição abreviada trilingue. Em "Ofen" (Fornos) encontramos (para tomar o mesmo exemplo) onze entradas, desde "Ofen" (Bauhandwerk) (Ofício de pedreiro) até "Ofen" (Zentralheizungen) (Aquecimentos centrais). Seriam aqui muito úteis ao usuário remissivas para "Glasschmelzöfen" (Fornos de fundição de vidro) e "Hoch-öfen" (Altos-fornos). Não se pode, naturalmente, fazer remissiva para todos os termos em que "Ofen" aparece como segundo ou terceiro elemento, mas nos casos mencionados trata-se de setores técnicos importantes. As remissivas existentes nas tabelas sistemáticas, mesmo sendo abundantes, não dão, infelizmente, uma visão geral de que se necessita muitas vezes. Se, p. ex., examinarmos nas tabelas todos os onze números para "Ofen" (Fornos), que encontramos no índice, verificamos que em nenhum lugar se faz remissiva para 542.41 "Ofen" (Experimentaltchemie) (Fornos — Química experimental), para 621.745.3 "Ofen" (Giessereitechnfk) (Fornos — Técnica de fundição) e para 683.9 "Ofen" (Heizgerate) (Fornos — Aparelhos de aquecimento). Destes três números, por outro lado, se faz remissiva para outros números indicados no índice. As remissivas são, portanto, muitas vezes, unilaterais, o que dificulta a visão de conjunto. Mais importantes seriam sinônimos para os verbetes que se encontram nas tabelas, mas estes faltam completamente. Para que seja verdade que "o índice alfabético torna a CDU um Thesaurus universal, como Fill afirma no prefácio, é preciso que os sinônimos estejam presentes em grande abundância. Outro exemplo. Encontramos três lugares para "Kühlen (Refrigerar), p. ex., na indústria do vidro 666.1.038, cinco lugares para "Kühlung" (Refrigeração) nos altos fornos 669.162.214, quatro lugares para "Abkühlen" (Esfriar) na técnica química 66.065.3 e dois lugares para "Abkühlung" (Esfriamento) na técnica de fundição 621.746.6. Aqui deve o pobre usuário procurar sempre e sempre mais, pois há ainda uma grande quantidade de outros verbetes que fazem remissivas para subdivisões das tabelas que não se encontram absolutamente nas quatro entradas acima mencionadas, p. ex., as seguintes: "Kühlanlagen" (Kälte-technik) (Instalações frigoríficas — Técnica de refrigeração) 621.565, "Kühler" (Pneumatische Maschinen) (Refrigerador — Máquinas pneumáti-

cas) 621.5.048, "Kühlgruben" (Warmebehandlung) (Poços de refrigeração — Tratamento térmico) 621.784.8, "Kühlmitter (Kerntechnik) (Agentes frigoríficos — Técnica nuclear) 621.039.534, "Abkühlungsgrosse" (Klimatechnik) (Grandeza de esfriamento — Técnica de ar condicionado) 628.862, "Ofenabkühlung" (Warmebehandlung) (Esfriamento de fornos — Tratamento térmico) 621.784.8. Aqui um sistema de remissivas teria facilitado muito a procura penosa dos usuários. É bastante surpreendente que os verbetes utilizados sejam por demais incompletos. "Kühlung" (Refrigeração) ou "Abkühlung" (Esfriamento) são métodos importantes no tratamento térmico de materiais metálicos. Este assunto é incluído em 621.78. Nas tabelas encontramos:

621.78.081 Esfriamento lento, p. ex. em fornos
.082 Esfriamento normal, p. ex. em ar parado
.084 Arrefecimento em líquidos

Temos, além disto, o seguinte número principal:

621.784.8 Instalações e dispositivos para esfriamento lento e normal. Esfriamento de fornos. Poços de refrigeração

Em "Abschrecken" (Arrefecimento) há no índice remissiva para 621.78.084. Por outro lado, apesar de cuidadosa busca, não pude encontrar nenhuma remissiva para .081 e .082. O número principal 621.784.8 é encontrado em "Ofenabkühlung" (Esfriamento de fornos) e em "Kühlgruben" (Poços de refrigeração), mas não se encontra em "Abkühlung" (Esfriamento), onde certamente se procura em primeiro lugar.

Entradas compostas de adjetivo e substantivo são muitas vezes invertidas. Assim, se encontra 615.816 não somente em "Künstliche Atmung" (Respiração artificial), como também em "Atmung, künstliche" (Artificial, respiração), como também, porém, estas inversões são incompreensíveis e levam a confusões. Exemplo:

Künstliche Augen (Glasindustrie)
(Olhos artificiais — Indústria do vidro) 666.189.1
Augen, künstliche (Augenheilkunde)
(Artificiais, olhos - Oftalmologia) 617.7-089.28
Pflege der Tiere (Zoologie) (Tratamento dos animais — Zoologia) 591.615
Tiere, Pflege (Tierzucht) (Animais, tratamento — Criação de animais) 636.083.3
Quando procuramos "Künstliche Augen" (Olhos artificiais) e encontramos somente 666.189.1, somos levados a crer que não há outro número. Mas se procuramos "Augen, künstliche" (Olhos artificiais), obtemos só o número de medicina. No segundo exemplo, a cousa é algo diferente. Chega-se tanto através de "Pflege" (Tratamento), como através de "Tiere" (Animais), enquanto que à "Zoologie" (Zoologia) chega-se apenas através do "Pflege".

Muitas vezes faltam inversões nos lugares onde se esperava que estivessem. Encontramos, assim:

Legiertes Stahl (Eisenhüttenkunde)
(Aços especiais - Siderurgia) 669.15-194
Legiertes Gusseisen (Eisenhüttenkunde)
(Ferro gusa — Siderurgia) 669.15-196
Acredito que a maioria dos usuários neste caso procura em primeiro lugar por "Stahl" (Aço) e por "Gusseisen" (Ferro gusa). Mas nesses lugares não encontram os respectivos materiais de liga. Convém observar, a esta altura, que o termo "Stahl" (Aço) se apresenta no índice de maneira muito incompleta. Encontramos apenas:
Stahl (Baustoffe) (Aço — Materiais de construção) 691.714

Todos sabem que o aço não é só material de construção e o número mais importante para este conceito é 669.14. Só encontramos este número no índice em ligas de ferro com carbono (Eisenlegierungen mit Kohlenstoff). O texto nas tabelas diz "Legierungen des Eisens mit Kohlenstoff (ausser Gusseisen). Stahl im allgemeinen" (Ligas de ferro com carbono, exceto ferro gusa. Aços em geral). O que gera ainda muita confusão é que em "Farben" (Tintura) se encontra a tintura de vidro, papel, manteiga etc, enquanto que a tintura importante dos têxteis se esconde algumas colunas adiante em "Färberei" (Tinturaria). Os seguintes itens estão também bastante confusos:

Färben (Glasindustrie) (Coloração
- Indústria do vidro) 666.1.031.1
Färben von Glas (Glasindustrie)
Coloração de vidros — Indústria do vidro) 666.1.058.7

O primeiro número é usado para coloração de fundentes de vidro, enquanto o segundo se aplica à pintura sobre vidro. Não seria possível especificar melhor os conceitos colocados entre parêntese? Esclarecerei melhor esta última pergunta com alguns exemplos. Encontramos:

Künstliche Befruchtung (Frauenkrankheiten) (Inseminação artificial — Doenças de senhoras) 618.177-089.11
Trata-se aqui de uma doença? Se é indispensável usar um conceito superior, melhor seria usar "Gynäkologie" (Ginecologia), ou talvez fosse bastante específico "Medizin" (Medicina). Aqui só é necessário estabelecer a devida diferença entre este conceito e "Künstliche Besamung" (Tiersucht) (Inseminação artificial — Criação de animais). Este é mais um caso em que as remissivas recíprocas são necessárias. Não seria mais indicado adotar aqui o termo internacional "Insemination"? Por vezes, os conceitos colocados entre parênteses são inteiramente dispensáveis:
Blindenschreibmaschinen (Schreibmaschinen) (Máquinas de escrever para cegos — Máquinas de escrever) 681.613.1
Blindenschiftdruckmaschinen (Vervielfältigung)

ÍNDICES ALFABÉTICOS DA CLASSIFICAÇÃO DECIMAL UNIVERSAL

(Máquinas impressoras para cegos — Reprodução) 681.625.92

No segundo exemplo foi escolhido um conceito superior inadequado. O conceito imediato é "Sonderdruck maschinen" (Máquinas impressoras especiais), mas que, nestes casos, se trata de máquinas de escrever ou de máquinas impressoras, depreende-se diretamente dos termos.

Posso apontar ainda alguns exemplos que não necessitam de muita explicação. No índice alfabético, os exemplos que seguem encontram-se na ordem aqui reproduzida:

Slavische Schriften (Escritas eslavas) 003.349
Slavische Strophen (Sprachwissenschaft) (Estrofes eslavas — Lingüística) 801.676

Slawen (Rasscnanhängezahl) (Eslavos — Número auxiliar de raça) (= 81)

Slawische Sprachen (Sprachwissenschaft) Línguas eslavas — Lingüística) 808.1

As duas formas de escrever que aparecem nas tabelas foram reproduzidas inalteradas no índice.

Também as seguintes entradas aparecem uma após a outra no índice:

Holland (Ortsanhangszahl) (Holanda — Número auxiliar de lugar) 425.3

Holländisch (Sprachanhangszahl) (Holandês — Número auxiliar de língua) = 339

Que "Holland" aqui mencionado se encontra na Grã-Bretanha e que ali não se fala holandês deveria ser indicado de alguma forma no índice. Para se encontrar Holanda nos Países-Baixos deve-se procurar em "Niederlande", ou em ortografia holandesa, ou em "Noord-Holland", ou em "Zuid-Holland". Tal coisa não pode agradar muito o usuário.

Espero que os exemplos apresentados acima mostrem de maneira suficiente que o índice alfabético da Edição Média da CDU, apesar das boas qualidades que tem, apresenta muitas falhas. Para que possa ser utilizado com êxito, é necessário já ter ótimo conhecimento da CDU. Em suma, com ele não se consegue nenhum novo adepto da CDU. Conforme já foi dito, as causas das falhas residem a meu ver na superestima do trabalho da máquina e, conseqüentemente, no insuficiente trabalho redacionai.

Que esperam os usuários?

A crítica do índice alfabético existente, como foi feita aqui, baseia-se, naturalmente, no pressuposto de que os usuários esperam realmente algo melhor. Que isto é verdade, experimentei várias vezes nos últimos tempos. Na Suécia preparamos atualmente uma nova abreviada. Diversos grupos de usuários expressaram a opinião de que a mais

importante melhora seria uma considerável ampliação do índice alfabético, que é aliás bem extenso. Foi proposta a aceitação de uma grande quantidade dos termos que aparecem freqüentemente e que na CDU só podem ser expressos por meio de dois pontos (sinal de relação). Até que ponto isto é possível, só poderá ser dito após ulterior exame. Hermann (8) afirma que só é praticável em edições especiais para matérias limitadas. A meu ver deve-se partir do pressuposto de que os usuários esperam encontrar, de futuro, na CDU tantos termos quantos podem procurar agora nos Thesauri (permitidos e proibidos). Não se deve esquecer que no grande Thesaurus americano TEST cerca de um terço dos descritores são combinações pre-coordenadas de conceitos que não podem ser expressos por simples números da CDU (11).

Mesmo que não se possa atender a todos os desejos dos usuários, dever-se-ia levar em conta que os documentalistas pela utilização dos Thesauri e da indexação coordenada baseiam-se nas palavras, mais do que antes. Esta realidade deve levar a índices muito melhores para a CDU, cujo emprego seja mais fácil e dê aos usuários a sensação de confiança.

- 1 - SAMULOWITZ, H. DK-Abgesang oder Neubeginn? *Nachr. Doh*, 21(4): 138, 1970.
- 2 - ARNTZ, Helmut. Die DK-eine Vielfacettklassifikation. *Nachr. Doh*, 21(4): 139-42, 1970.
- 3 - DAHLBERG, Ingetraut. Möglichkeiten einer Neugestaltung der DK. *Nachr. Dok.*, 21(4): 143-51, 1970.
- 4 — FILL, Karl. Dezimalklassifikation — positiv geschen. *DK-Mit.*, 15(5): 17-20, 1970.
- 5 - ÖHMAN, Einar. Die DK-Kritik von H. Arntz im Lichte von Konkordanzuntersuchungen zwischen DK und Thesauren. *Nachr. Doh*, 23(2): 65-8, 1972.
- 6 - BRITISH STANDARD INSTITUTION, London. *Guide to the Universal Decimal Classification (UDC)*. London, 1963. (FID 345)
- 7 — MILLS, J. *Universal Decimal Classification*. New Brunswick, 1964. (Rutgers Series on System for the Intellectual Organization of Information, 6)
- 8 - HERRMAN, P. *Pratische Anwendung der Dezimalklassifikation*. 6. Aufl. Leipzig, 1970.

EIMAR OHMAN

9 — BOSE, H. Chain indexing and Universal Decimal Classification. *Indian Librarian*: 235-9, Mar. 1968.

10 — KOCH, K. -H. Ein Verfahren zur Maschinellen Herstellung von DK-Sachregistern. *Proceedings of Second Seminar on UDC and Mechanized Information Systems*, Frankfurt M., 1-5 June 1970, Copenhagen, 1971. p. 21-42.

11] — WELLISCH, H. A concordance between UDC and Thesaurus of Engineering and Scientific Terms. *Herce Novi Symposium on UDC, Proceedings*, June 1971. 30 p.

ABSTRACT

Discussion about the alphabetical indexes of the Universal Decimal Classification — UDC — Since different publications presented, in the last years, have dealt with problems related to the tables. This study is based on the different editions of UDC.

Linguagem Documentária: Acesso à Informação

Aspectos do problema

Manoel Adolpho Wanderley

Biblioteca Nacional

Escola de Biblioteconomia e Documentação
da FEFIEG

SINOPSE

O processo específico da comunicação documentária é encarado do ponto de vista das relações entre a Linguagem e o universo da representação gráfica, objeto da documentação. Descreve-se uma amostra dos meios de pesquisa bibliográfica criados para tal forma, dos quais as linguagens documentárias constituem a forma mais elaborada. Acentuam-se as implicações lingüísticas da análise do conteúdo, destacando-se as interpretações do plano sintático e do semântico. Culminam tais considerações no exame das relações lógicas e categorias fundamentais das linguagens naturais e documentárias, bem assim na referência sumária a algumas estruturas intelectuais básicas subjacentes a ambas.

O - INTRODUÇÃO

Procede o presente trabalho de surpresa muito à feição da de M. Jourdain. Então não teríamos estado, nós bibliotecários e documentaristas, a fazer lingüística sem o saber — quase diríamos sem a saber? — a criarmos linguagens que desejamos tão significativas quanto a natural? a traduzirmos em códigos os textos em nossa custódia? a parafrasear-lhes o conteúdo em resumos sinaléticos ou informativos, produzindo documentos derivados que dão acesso aos primários, numa palavra, a comunicarmos ao leitor mensagens que, no intuito de

**Dissertação apresentada ao Instituto Brasileiro de Bibliografia e Documentação/Universidade Federal do Rio de Janeiro para obtenção do grau de Mestre em Bibl. oteconomia e Documentação, em 19. 10. 1973.
Orientador: Prof. Sylvio Edmundo Elia, Faculdade de Humanidades Pedro II.**

lhe responderem às questões, os autores confiaram ao nosso zelo e guarda?

Eis algumas das perguntas que seria lícito reduzir tão só a uma: quais as relações entre a documentação e a lingüística? Melhor ainda: se ali onde se entrosa uma com a outra se situam também as chamadas "linguagens documentárias", as LD, tudo se cifraria em sondar-lhes a índole e a condição frente às naturais, bem assim em fixar o papel destas no gerar aquelas, no servir-lhes de modelo, ou no dispensar-lhes o uso.

Ora, os conceitos remetem aos conceitos e, para sustarmos o ressurgir de questões, é mister cingi-las ao problema no que tenha este de crucial. Mas descobrir onde lhe apertam os sapatos, eis uma das mais árduas tarefas do filósofo, como observou Wittgenstein, ¹⁰⁷ que justamente se comprovou em sondar os jogos "lingüísticos" a que igualou os sistemas de comunicação.

Para os que não são filósofos será mais fácil, por vezes, dizer que perguntas não se farão, que outras já se têm por respondidas ainda que a título de hipótese de trabalho. Nesse último caso se incluem as referentes ao significado mesmo da comunicação e ao da linguagem natural, da LN. Demos por tacitamente aceito um conjunto de caracterizações dessa última, que poderíamos sintetizar, a partir de G. Mounin, nos traços que a seguir esboçamos. O que é específico das línguas naturais não é a comunicação — há para esse comércio humano meios diferentes com a mesma função; nem o arbitrário dos seus signos — muitos outros também o são; nem a noção de sistema — valemo-nos de sistemas, parciais ou não, de símbolos não lingüísticos; nem a linearidade da men-

sagem - a música e o cinema transcorrem no tempo; nem tampouco o caráter discreto dos sinais, já que a mensagem também se faz pelo ritmo, pelo tom, e assim por diante.

Próprio das línguas naturais é, em termos de Martinet, a dupla articulação, ou seja, o fato de que toda língua falada "articula a realidade" sob duas formas: em unidades significativas (morfemas) e, no interior de cada unidade significativa, em unidades sonoras mínimas (fonemas)", no dizer de D. Fleury.²²

Devida atenção foi sempre prestada, no fio das presentes cogitações, à distinção entre eixo horizontal, relativo à combinação sintagmática de duas ou mais unidades significativas na cadeia falada, e eixo vertical, paradigmático, atinente às relações associativas de Saussure²³ ou sistemáticas de Hjelmslev,²⁴ no caso vertente porém, em referência às LD, às "representações indexadas" e à organização do "universo dos conhecimentos" que lhes são próprias.

Note-se que, ao nos reportarmos, de modo explícito ou não, a tais conceitos face às LD, excluímos qualquer consideração metódica do que nelas corresponderia diretamente aos grafemas e, indiretamente, aos fonemas, isto é, à segunda articulação. Tal restrição, adotamo-la cômicos embora de quanto possa haver de elucidativo, para o nosso tema, no que viria a ser a "comunicação gráfica" das LD, tópico em que se estudassem a escolha dos signos e o seu modo de combinação.

Mais absoluta ainda — "et pour cause" — foi a limitação, dentre as funções lingüísticas de R. Jakobson,²⁵ à função referencial, de tipo cognitivo e objetivo.

Ao contrário, incluímos sem vacilar a semântica, cuja reintegração na lingüística demos por definitivamente assente, até mesmo como pretende G. Mounin, "no domínio de uma lingüística estrutural de que algo apressadamente a tinham excluído a priori".

Assim quanto à análise do significado mesmo, temos entre os mais válidos os processos suscetíveis de objetivação formal, usados para identificar a estrutura interna da unidade significativa mínima, numa palavra, entre outros, os critérios da teoria lógica do significado que deriva de Hjelmslev.²⁶ Como acentua fortemente, a esse propósito, S. Elia, "entre logística e lingüística deve existir íntima colaboração. Na prática uma língua é uma semiótica em que todas as outras semióticas podem ser traduzidas. Não só todas as outras línguas, mas também todas as outras estruturas semióticas imagináveis".

Bem sabemos que, sem denunciar, ao dizê-lo, algum excesso de controvérsia no cenário lingüístico, seria fácil contrapor tal ou qual teoria a tal outra.

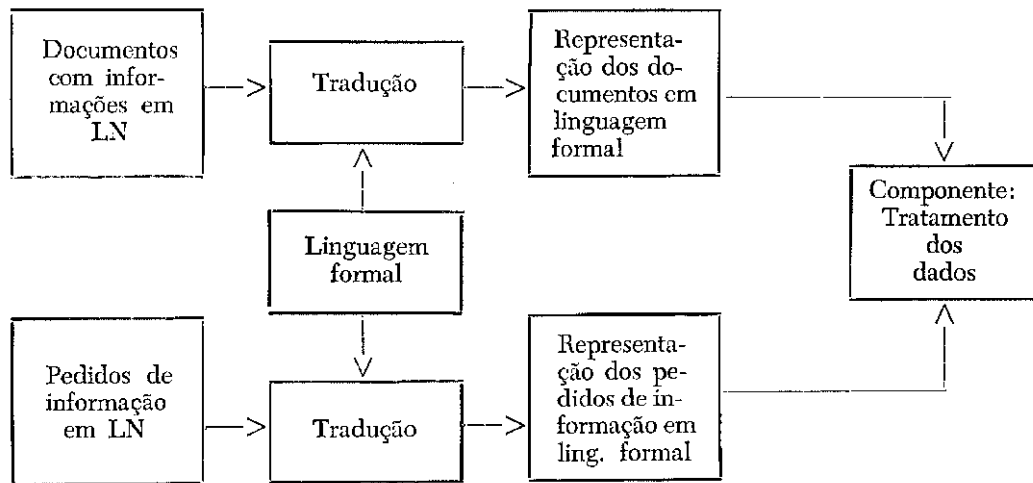
Assim, dois dentre os autores em que buscamos as mais copiosas e ricas informações — Grolier e Coyaud — discrepam, por exemplo, ao apontar o primeiro no segundo, a fidelidade a "une linguistique... assez dépassée" ou em criticar-lhe, em termos de Hjelmslev, preferindo o outro os de Martinet, a consideração exclusiva da "forma da linguagem" com negligência da sua substância. Mas questões semelhantes, de sumo interesse, fugiriam quase por definição ao escopo de nosso trabalho, que é antes delimitar uma área de observações e circunscrever o espaço de um problema.

Por outro lado, na ótica da análise dos sistemas, e já que nos cingimos à interface "linguagens documentárias — linguagem natural", não remontamos às totalidades mais amplas, ao processo de comunicação como tal, nem sequer, dentro dele, ao que, desde J. Moreno,²⁷ se vem chamando "conserva cultural" ou, humanisticamente, "memória cultural".

Não insistiremos, assim, sobre o fascinante tema da biblioteca, da enciclopédia ou "documentatriz" universal, como o expõe A. Moles,²⁸ tese em que ressurge, a nosso ver, em estilo de diagrama, a metáfora clássica de Ortega Y Gasset, do "bibliotecário como filtro", mormente, acrescentemos, no seu afã de que se concretize em proveito social, o "savoir tout, tout de suite, partout" e na sua missão de provedor da cultura.

Na expressão de Moles,²⁹ as diferentes línguas é que funcionariam como "filtros das coisas de importância" e o "velho sonho da enciclopédia universal" viria a realizar-se graças "às memórias mecânicas e às técnicas de análise de conteúdo, aptas a exprimir, em "maquinas", a totalidade das idéias e dos *semantemas* novos produzidos pelo homem a cada instante".

Para análise do conteúdo, sim, se voltará a nossa atenção. Como não nos move, todavia, o exame da língua universal, mas antes, ao nível das LD que talvez lhe venham preparando o caminho, as relações com a linguagem e o seu estudo, limitar-nos-emos, para tornar explícitas as nossas ordenadas conceptuais, a esquematizar aqui o sentido da representação e análise de conteúdo que tomamos por guia. Estas, de acordo com C. Montgomery,³⁰ que tanta luz lançou sobre a questão em foco, constituem um dos componentes do sistema de informação total, com "input" da "aquisição" e "output" para a "utilização da informação", via "tratamento dos dados". Os registros ou documentos em LN adquiridos pelo sistema, bem assim os pedidos nela gerados pelo usuário, são *traduzidos* em alguma linguagem formal, meio de comunicação entre este e aquele e base para a comunicação no sistema. O diagrama seguinte, da autora citada, dispensa qualquer comentário adicional:



Resta, quanto a sistemas, um esclarecimento: o método da análise quantitativa, da avaliação, não o adotamos, em princípio, nas presentes considerações. A medida, por exemplo, dos índices de recuperação ou de pertinência, ou a estimativa de quaisquer outros parâmetros, julgamo-las estranhas a nosso ponto de vista, ao qual subordinamos e reduzimos, também, a faceta computacional e estatística, tanto da documentação quanto da lingüística.

Para apresentar, por fim, em dois ou três traços, a marcha da nossa investigação, diremos que o sentido geral que lhe imprimimos foi, no campo da documentação, o de irmos do significante para o significado. Partimos, assim das interpenetrações dos *domínios da linguagem e do documento*, a propósito do que analisamos o universo da representação escrita e a natureza lingüística dos meios de pesquisa. Voltamos, em seguida, a atenção para as gradações entre a linguagem natural e as *linguagens documentárias*, em relação às quais examinamos a natureza, as classes de símbolos, os componentes e as tentativas de tipologia. Na "busca de métodos e modelos" atentamos não só para a relatividade dos conceitos que lhes têm sido dados face à linguagem natural, mas também para a concatenação maior ou menor dos descritores e a formalização decorrente da automação. Segue-se a *análise do conteúdo*, como atividade documentária essencial, o que significa que a sintaxe é propriamente o alvo das cogitações. Após discriminarmos o conceito de indexação e de classificação sob esse ângulo, esboçamos um modelo abstrato proposto para as gramáticas em questão, de que se enumeram alguns tipos. Alcançamos o nível do significado, indo das estruturas superficiais

para as profundas. Delineam-se aí algumas *teorias lingüísticas com base semântica*, em especial as de Fillmore⁷⁰ e a metateoria de C. Montgomery.⁷⁰ Um dos pontos de convergência das linguagens naturais e das documentárias são, nessa altura, os problemas da polissemia, da homonímia, da sinonímia e da definição. Ocupamo-nos, cingindo-nos a algumas considerações genéricas, das *relações lógicas e categorias básicas* das linguagens naturais, das conjunções em particular e dos componentes semânticos. Delimitada a área das conexões entre lingüística e lógica formal, e a das estruturas intelectuais, retomamos os mesmos tópicos no que concerne às linguagens documentárias, *stricto sensu*. Comparam-se, para esse fim, os pontos de vista relacionais de diversos sistemas significativos, sublinhando-se no que lhes concerne, o aspecto lexicográfico, o papel dos "operadores" e alguns casos de estruturas triádicas.

Duas palavras, por fim, quanto aos termos técnicos e às citações — ou melhor — à forma de exposição. Iniciamo-la sob o manto de um Molière por assim dizer "lingüista". Mas nem por isso haveria fugir, sem boa causa, — e não o fizemos — ao risco de incorrer num dos preciosismos que o seu riso castigou para sempre, o das gírias especializadas avessas ao natural da linguagem.

Não vacilamos, pois, em usar, com moderação, os tecnicismos e, até mesmo, eventualmente, a simbologia matemática. Mais para temer seria, a esse respeito, a conhecida e candente crítica da lógica de Bar-Hillel aos formalismos em voga e sem razão de ser, de que não raro se revestem os estudos devotados à recuperação da informação.

Todavia é essa mesma trivialidade, que lhe parece logicamente argüível em tantos casos, que nos

fez aceitá-la aqui, se os especialistas dela se valem para economia de expressão e precisão dos conceitos. Nem outro motivo nos moveu quanto aos termos que se lhes afigurou oportuno cunhar. Se um autor fixa, como A. Moles,²² — e a substância da proposição que, a seguir, dele daremos é outro dos pressupostos a que acima aludimos — para a medida da cultura, tanto individual quanto coletiva, a fórmula: "C = n° de *culturemas* x de associações realizadas entre eles", por que, nesse passo, "descreever" tudo o que o símbolo "produto" implica, e trocar "culturema" por "elemento de conhecimento ou de percepção memorizada como unidades simples"?

No que toca a citações — e voltando urna última vez a Molière que disse "je prends mon bien partout où je le trouve" — exploramo-las sem restrições, já que a natureza das presentes observações beira a da "resenha" ou do "state-of-the-art report". Respeitamos as normas em uso, sem declinar as páginas do item referido, por facilmente localizáveis em todos os casos. Limitamo-nos, outrossim, a referenciar o autor citante, no caso de citação.

Fora de regra ficará a que acrescentamos nesse ponto, já que nos escapa quem disse algo parecido com o seguinte: "um bom — ou seria um mau? — trabalho científico é o mais curto caminho entre várias citações". Mas preferimos que, ao nosso, não se venha a aplicar nenhuma das versões possíveis, ou para verberá-lo ou para elogiá-lo. Pelo que houver aqui de positivo, vão os nossos agradecimentos ao Instituto Brasileiro de Bibliografia e Documentação que, em boa hora, proporcionou esse convívio profissionalmente fecundo, com tantos especialistas de nomeada no domínio da Ciência da Informação, inclusive com alguns dos mais representativos que enriquecem a lista de autores citados.

A nossa gratidão, pois, ao Professor Silvio Elia, cuja solicitude impediu, nesse sentar "entre deux chaises" que é a nossa dissertação, onde se aproximam documentação e lingüística. (3 cairmos de todo fora da última.

Também o nosso vivo reconhecimento à Presidente do IBBD, Senhora Professora Hagar Espanha Gomes, e à Diretora da Biblioteca Nacional, Senhora Jannice Monte-Mór. Favoreceu-nos nesse caso, não propriamente o permitirem e como ordenarem utilizássemos em proveito de nossas indagações as imensas reservas de saber, respectivamente técnico e humanístico, das Instituições a que também pertencemos por tantos elos. Devemo-lhes, antes, em definitivo, o encorajamento de colegas com quem penamos em nossa primeira incursão no domínio dos "pushdown-stores", sob a orientação do saudoso Professor James W. Perry, ao destrinçarmos os segredos lingüísticos — os outros nós os conhecíamos — do "string" por ele proposto como exem-

plo: "o Brasil tri-campeão".

Por fim, e muito especialmente, timbramos em tornar pública a nossa dívida à Professora Gilda Maria Braga, sem cujo espírito crítico e sentido de exatidão jamais teríamos chegado a proporcionar as nossas possibilidades ao empreendimento em que nos lançamos.

1 DOCUMENTAÇÃO E LINGÜÍSTICA: INTERPRETAÇÕES DOS DOMÍNIOS

Afiguram-se particularmente ricas as relações entre o documento, como suporte da atividade que lhe explora e difunde o potencial de informação, e a linguagem, como objeto privilegiado do domínio da ciência que estuda a "vida dos signos no seio da vida social" (Saussure, 53).

Com efeito, no âmbito do universo bibliográfico e da representação escrita das idéias, por intermédio de códigos lingüísticos — naturais ou artificiais — usados no decurso da comunicação gráfica, como esteio para a análise e indexação do conteúdo e o subsequente ciclo das questões-respostas, nesse triplo plano — abstração feita de outras discriminações — é que se dá o processo documentário.

1.1 — *Universo da representação escrita*

Observe-se, desde logo, que do livro à microforma mais compacta que a biblioteca coloca na mão dos leitores, o que prevalece como veículo das mensagens que ele divulga são textos e registros de natureza predominantemente escrita. É; bem verdade que à sua custódia se confia, outrossim, o que interessa antes à semiologia *lato sensu*, ou seja, outros códigos em outros suportes físicos — não apenas documentos audiovisuais de toda sorte, mas **também**, por **exemplo**, cartões **perfurados** ou fitas e discos magnéticos, em que se armazenam dados.

Entretanto, o que continua a constituir o núcleo mais relevante do seu acervo, o que a preocupa mais genuinamente, fiel, como permanece, à advertência clássica "*verba volant, scripta manent*", é a documentação sob forma gráfica. Não consiste a escrita, com efeito, "numa representação visual e durável da linguagem que a torna transportável", como a qualificou Cohen(62)?

Em modalidades mais ou menos cotidianas ou cultas, para grupos mais ou menos restritos, atinentes a domínios e disciplinas mais ou menos formalizados ou especiais, consoante a função social a que os registros se destinam — em princípio nada se exclui dos repositórios bibliográficos — as linguagens neles preservadas aproximam-se, em maior ou menor grau, das naturais, em cuja definição Cardin (26) as engloba, inclusive suas ramificações apropriadas aos textos científicos.

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

Tanto é verdade que, no dizer de Gleason (66) constitui a língua oral "a realidade última do que se lê ou escreve". Dela, o código escrito seria o substitutivo. Aproxima-os, de fato, marcado paralelismo, capaz - segundo L. Prieto — de alcançar no "Fonematismo" uma tal correspondência que para cada fonema de uma teríamos um grafema na outra, e vice-versa.

Não surgiria, ademais, em certos casos, uma espécie de paralelismo e substituição dupla entre as linguagens naturais, as LN, e as documentárias, as LD, às quais aludiremos adiante?

1.2 — *Natureza lingüística dos meios de busca documentários*

Por outro lado, para o acesso à massa de noções acumuladas em semelhante processo, sempre lançou mão o especialista nele empenhado — do bibliotecário tradicional à sua variante atual, afeita à informática — de instrumentos, em última análise, igualmente lingüísticos.

No "descriptive continuum" a que se reporta F. Jonker (39) como base qualitativa da sua teoria "generalizada" da indexação, o espectro desta se estende dos índices e repertórios multidimensionais de palavras-chave às classificações ditas hierárquicas.

Sem dúvida, teríamos, num extremo, elementos extraídos *dos próprios textos*, de partes suas tidas como reveladoras do conteúdo, quando não de *resumos*, que, de certa forma, os parafraseiam. Tomados dentre os demais termos, porque mais significativos estatística ou semanticamente, confere-se-lhes, ou não, estrutura combinatória global, *verbi gratia*, a pós-coordenação em moldes simplesmente booleanos. Atuam as palavras-chave — como não raro se chamam — na condição de índices referentes ao teor de presumível interesse. Manual ou automaticamente destacados do contexto, ao recorrermos a tais segmentos de frases ou meras palavras isoladas, constituímos-nos, é claro, em tributários das linguagens naturais aludidas. Nada mais espontâneo, a bem pensar, que semelhante expediente. A ele se aparenta a prática, que não é de hoje, de nos valermos de "títulos" para sugerir, em poucas palavras, o que dizem muitas. De hoje, porém, é o explorá-las por permutação ou "rotativamente" nos computadores, como o fazem os vários programas para geração do KWIC (PACT) ou KWOC e os meios similares de pronta notificação, com as palavras-chave preservadas ou não, no contexto imediato.

A naturalidade do intitular — e as deficiências, também, de nos atermos às auto-designações dos textos — aliam-se, de algum modo, à celeridade e à cega obediência das unidades eletrônicas, numa "chave" cômoda para a comunicação do que encerram os escritos.

Mas não se cinge o analista de assunto a essa espécie, por assim dizer, de amostragem dos conteúdos. Vemo-lo, desde que se formaram as primeiras bibliotecas, refinar tal matéria-prima vocabular e elaborar esquemas cada vez mais afastados da linguagem natural para indexar ou referenciar a informação armazenada.

Labora-se aí no outro extremo do espectro, na região das classificações hierárquicas, tão familiares ao mundo bibliográfico.

Que estas constituem uma linguagem artificial é S. R. Ranganathan quem o afirma, o criador da Colon Classification, que abriu amplas perspectivas à teoria e à técnica documentária. "Classification is the process of translation of the name of a specific subject from a natural language to a classificatory language".

Em suma, a culminar no polo representado por tais esquemas e pelos "códigos semânticos", e a partir dos precedentemente assinalados, há pois, toda uma gradação de "linguagens". Fazem-se explícitas, no seu conjunto, articulação cada vez mais ricas, que a crescente formalização no fixar-lhes as descrições, os princípios e os modelos, também torna mais racionais.

Uma ocorrência de cabal, porque deliberada, similitude entre as estruturas desses sistemas "artificiais" e os da comunicação corrente, ao nível das palavras do discurso e dos termos do código, fornece-a o "Semantic Code" de Perry, para a Western Reserve University.

Assim, nele corresponde uma noção como "numeroso" a LYMN, onde L-MN é o fator semântico — unidade — e Y, a "relação analítica" atributiva. Salta à vista a analogia se, num confronto, usarmos para frisá-la, o esquema com que K. Baldinger mostra o jogo complexo de unidades da palavra "apasionado".

4 2 3 2 2 4 4 3 2 3 3 4
... a-p-a-s-i-o-n-a-d-o... ...L Y M N ...
1. 1 1 1 1 1 1 1 1 1 1 la la la la

1 = fonemas (unidade distintiva mínima)

2 = morfemas (unidade significativa mínima, em lista fechada)

3 = lexema (idem, em lista aberta)

4 = contexto (quadro sintagmático)

Propomos **la** para os grafemas.

1.3. — *Implicações lingüísticas das operações documentárias*

Também encerram implicações marcadamente lingüísticas as operações mesmas ligadas ao armazenar e ao recuperar das informações. Nos sistemas correspondentes, apreendidos na sua feição global, a indexação e a análise do "input" anteci-

pam certas "questões" hipotéticas e relações conceituais estabelecidas pelos seus usuários (R. Taylor).

Ao ser descrito o assunto dos textos, têm início as transformações documentárias a que se submete, durante a análise, a forma de representação dos conteúdos, bem assim a dos "pedidos" de informação, na fase subsequente. Enceta-se, numa palavra, a sua tradução para os termos da LD, quando não a mera redução do teor dos documentos, com que se lhes faculta a seleção futura, e isso no grau de pertinência desejado.

Que há forte afinidade dessas com as transformações lingüísticas, pelas quais se converte a construções Canônicas a estrutura das frases formuladas em LN, provam-no a possibilidade e o fato de se aplicarem ao tratamento dos textos, para objetivos documentários, as gramáticas mesmas e os algoritmos da lingüística dita computacional e os instrumentos da chamada tradução automática. Esse tipo de correlações, bem manifestas, por exemplo, aos autores do SYNTOL, é que os levou a perfilhá-las na sua "Syntagmatic Organization Language".

Para o que decorre de semelhante representação documentária, para os extratos, indexações e resumos que a integram, atentaremos um pouco mais longamente em outros pontos.

Neste, cabe notar tão só, quanto à estratégia da busca — fator do êxito nas respostas fornecidas como resultado da recuperação da informação — como se condicionam, até mesmo nessa etapa final, os fatores operacionais e os elementos do discurso. Assim, as variáveis da estrutura sintática das "questões" se cifram, justamente, como as enumera R. Taylor: *no número de termos; no número, tipo e "direção" dos verbos modificadores e conectivos; no nível de complexidade lógica nas interrelações entre os termos, e no grau de similitude da "linguagem de questão" com a do sistema.*

2 - DA LINGUAGEM NATURAL À LINGUAGEM DOCUMENTÁRIA

Nos sistemas de comunicação assim delineados, destacam-se, portanto, entre os componentes necessários — não a ponto, porém, de excluir as próprias linguagens naturais — os que vêm recebendo, genericamente, quer a denominação de *linguagens documentárias* ou a de linguagens de *indexação* — "index language", como prefere J. Melton — quer a de *linguagens descritoras*, — "descriptor language", de B. Vickery (68) ou, ainda, a de *codificações documentárias*, de E. de Grolier (49). Parece-nos dispensável, para o fim que nos move, procurar critérios pelos quais se venha a reconhecer sentidos diferentes entre tais designações. Aplicar-lhes-emos o que decidiu D. Soergel (97) sobre questão terminológica similar: "We take

the following terms as synonymous: information language (IL), information retrieval language, information description language, documentary language".

2.1 — *Natureza e situação das LD*

Ao se lhes fixar o conceito sob qualquer desses nomes, ter-se-ia em mente expressar não só a finalidade essencial a que visam, a saber, a "descrição da informação para a armazenagem, a recuperação, o processamento e a comunicação", mas também uma série de funções básicas que as condicionam ao nível mesmo da prática bibliográfica.

Entre elas destaca-se, de acordo com os que lhes perscrutaram o alcance, como B. Vickery (68) e R. C. Cros:

- a) A capacidade de obter, por ocasião das respostas aos pedidos de informação, um máximo de indicações relevantes ou pertinentes e, para tal, de coincidência entre as descrições de autoria do indexador e as do usuário, numa palavra, a *normalização*.
- b) A economia nos símbolos, a eliminação de redundâncias, a condensação.
- c) A facilidade de "modular" as questões, típica do SYNTOL, ou seja, a de transformá-las, na fase de seleção, generalizando-lhes ou especificando-lhes a compreensão mediante a *organização semântica*.

Tentaremos, pois, em traços amplos, expor-lhes:

- a natureza e a situação em face das LN;
- os símbolos de que lançam mão;
- os seus componentes usuais de maior significado;
- alguns ensaios com que se procurou discriminar-lhes os tipos mais acentuados.

Qual a natureza de semelhantes codificações bibliográficas, a sua razão de ser e justificativa lógica? A resposta a esse tipo de interrogações, ofereça a própria dinâmica do universo de comunicação a que pertencem. Se o tomarmos como referente ou, melhor, nele, o conjunto das linguagens naturais, salta à vista, já o insinuamos, a relatividade das que se qualificam de documentárias.

Na medida mesma em que estas se afirmam em oposição à riqueza daquelas — e isso, via de regra, restringindo-lhes, com mira na representação indexada, o vocabulário ou a gramática, quando não ambos — assumem a condição de linguagem artificial, isto é de sistema de signos "deliberadamente construído, por uma pessoa ou grupo pequeno de pessoas, e num período de tempo comparativamente curto, para determinado objetivo", segundo C. H. Kellog.

Precisão, concisão e adequação ao processamento, sobretudo ao automático, além do poder inferencial, eis o que provem em contrapartida. Compete a quem as projeta, erigir tais qualidades em

parâmetros a serem considerados na sua construção e definição.

É bem de ver que, aí também, no formulá-las e descrevê-las, se revela a tendência metodológica crescente para a formalização, cujas conexões gerais com o domínio das LD consignaremos em outra parte.

Basta-nos mencionar, de passagem, porque incide sobre a caracterização que aqui tentamos, a proposta de um modelo, com base na teoria dos conjuntos, por F. B. Thompson, no "Seminar on computational Linguistics", 1966.

Conforme o seu ponto de vista, algo paradoxal, "a natural language N is an unending chain { L_i }

i e I of more and more discriminating formal languages".

Dada uma linguagem formal L_i, há sempre outra

L_{i+1} mais discriminativa, e isso *ad infinitum*, de tal forma que $L_i \leq L_{i+1} \leq L_{i+2}$. Eis uma cadeia, ou seja um subconjunto linearmente ordenado de um conjunto parcialmente ordenado, definindo-se a *linguagem natural* como uma cadeia máxima de linguagens formais.

A "articulação" do que sabemos sobre um assunto particular não é jamais completa: "se se afirma de um "corpus" de sentenças de N, que ele está em L_i, pode afirmar-se que está também numa linguagem mais discriminativa L_j de X. Para uma palavra W, de X, embora de sentido específico nas sentenças de L_i haverá matizes adicionais na segunda, mais discriminativa.

Do mesmo modo não há propriamente a gramática de N, mas níveis cada vez mais discriminativos de "gramaticidade".

De modo geral, se não denotam sempre a existência de modelos explícitos e formais concernentes ao processo de comunicação em causa, refletem contudo, as definições propostas para as linguagens documentárias, características estruturais e lingüísticas do tipo das que acabamos de investigar, implicadas, de fato, no seu conceito e traduzidas por termos como "símbolos, descrições, questões, indexações, léxicos, gramáticas" e similares.

Quem as dissecou detidamente — M. Coyaud (23) — considera a LD "instrumento de comunicação mediata entre os usuários e os autores do documento", ou melhor, entre aqueles e os "documentalistas (ou bibliotecários)" graças ao qual venham os primeiros a obter "nos menores prazos possíveis, os documentos escritos ou não, ou as referências aos documentos que os interessam".²⁴

Mediata, esclarece, porque põe em jogo dois códigos, e não um só, como a comunicação imediata, em que há transcrição condicionada tão só pela natureza do canal. Um código (fonte) transforma-se organicamente em outro no caso da LD, que

atua como intermediário entre aquele e o destinatário.

O primeiro é, sob esse ângulo, original. A LD — Coyaud²³ o demonstra fartamente em sua tese — o será em maior ou menor grau e, nisso, ela se assemelha às *linguagens intermediárias* (LI) propostas, por vezes, para a tradução automática entre duas línguas naturais.

De menor significado é a distinção devida a J. C. Gardin (23) entre LD diretas e indiretas (LDD e LDI), diferenciáveis por se reportarem, umas a documentos originais (objetos ou textos quaisquer) e as outras, a documentos no segundo grau ou secundários, isto é, a documentos que descrevem, eles próprios os documentos originais.

2.2. — Símbolos das LD

Quanto aos símbolos a que recorrem, as LD dizem respeito, sobretudo, à comunicação escrita. Aproveitamos de preferência o código alfabético e o numérico, ou ambos, mas não lhes é estranho o valerem-se de significantes outros, como o diagrama. Cros fá-las corresponder, no seu *Glossaire* a "um sistema de símbolos utilizados na indexação dos documentos", equivalente esta última, "à análise documentária em que se tende a exprimir o conteúdo dos textos mediante "descritores", com ou sem indicações sintáticas".

A aceção dada aos "signos" da LD é via de regra sensivelmente ampla, já que nela se abarcam "symboles... empruntés ou non à une langue naturelle, "en clair" ou "codes"... etc. (44). Observe-se que não obstante essa "hospitalidade" quanto aos grafemas, a notação é questão crucial para a codificação eficaz.

2.3 — Componentes e princípios das LD

Encaradas do ponto de vista não já da sua finalidade, c sim dos componentes que integram num dado sistema e área, as LD acusam — são concordes no registrar o fato quase todos os autores citados — uma dupla ordem de fatores, suscetíveis, ambas, de assumirem diversos graus de elaboração e articulação ou de ostentarem feição rudimentar, redutível o segundo, nas formas mais frustas, à quase nulidade:

- a) um dicionário ou *léxico documentário*, ou seja, o conjunto de "signos" ou "descritores" (palavras naturais, símbolos alfa-numéricos etc.) concernentes à representação das informações documentárias, e
- b) uma gramática ou *sintaxe*, ou série de regras para combinar os termos.

Cabe notar aí, de imediato, em correspondência acentuada com os dois eixos — vertical e horizontal — em que se estruturam quer o pensamento subjacente ao discurso, quer este próprio: a *organização*

paradigmática atinente ao léxico documentário e significativa de certas relações semânticas estabelecidas a priori, segundo Cros; a *organização sintática gramática*, própria das representações documentárias, "expressão de certas relações lógicas observadas" entre os elementos.

A proporção variável entre o ingrediente léxico e o sintático decorre da própria diversidade do panorama das classificações e codificações, e Coyaud (23) que lhes fez, com minúcia, o levantamento, vê na possibilidade de significarem-se quase as mesmas coisas com "mais ou menos lexemas e operações sobre esses lexemas", uma das explicações para o fato.

Reduzindo a uma espécie de formato único as diferenças nos processos de expressão sintática, N. Bely formula-lhes a estrutura como (R, a, b) em que *a* e *b* designam dois descritores, *R*, a existência de uma relação sintática, e o índice *i*, uma interpretação particular dessa relação, por exemplo, "causalidade" — *a* é agente do evento *b*; "finalidade" — *a* é o destino de *b* — etc. Donde, do ponto de vista da "gramática", diferentes LDs consoante os valores de *i*:

- $i = 0$ linguagens sem sintaxe
- $i = 1$ linguagens com relações sintáticas não específicas
- $i = 2, 3 \dots n$ linguagens cujas ligações sintáticas prevêem 2, 3...n categorias.

2.4 — Ensaio de tipologia das LD

Conviria, ainda, considerar — antes do breve exame, na seção seguinte, de alguns métodos de que se valem os especialistas para descrever os códigos em apreço — duas ou três tentativas de discriminar-lhes as variedades principais.

Há os que põem em foco a totalidade dos sistemas, para distinguir os enciclopédicos dos especializados, como o fazem Grolier (49) e a taxonomia tradicional no mundo bibliográfico.

Afastada qualquer presunção de exaustividade, seria possível adotar, para a classificação do gênero de linguagens a que pertencem as documentárias, o critério proposto por Coyaud (23), a saber, a dosagem, nelas, dos elementos lógicos e lingüísticos.

Definir-se-iam, assim três categorias de sistemas:

- a) os lógico-lingüísticos, mais tradicionais, com ou sem relações sintáticas, as quais, se presentes se regem, a seu turno, por uma sintaxe livre ou rígida;
- b) os fundados na lógica matemática;
- c) os puramente lingüísticos, para operar no quadro das LN.

C. H. Kellog situa-se mais decididamente na área

da comunicação homem-máquina e distingue, atinente às "IR languages", isto é, às linguagens de recuperação da informação, os sistemas com base em dados estruturados, com base lógica e com base na cooperação "on line" entre o usuário e a unidade de processamento.

Também da simulação, pela máquina, das atividades do homem no campo da informação, ou melhor, dentro deste, do processo em que se realiza a tradução da LN para a *linguagem formal* e vice-versa, parte C. Montgomery (70) para quem esta última é um "system of content categorization represented by a subject authority list, or in a more advanced application, a complex system which specifies the syntactic and semantic content — e. g: some enriched version of the propositional calculus". No levantamento que faz de modelos e técnicas na lingüística computacional há para assinalar, além de algoritmos para a análise morfológica, sistemas para a análise sintática, limitados ou globais (full-scale), para a análise semântica, e para a combinação da análise sintática e semântica, onde se enquadram os destinados ao processamento do texto c do ciclo das questões-respostas, bem assim, os que visam à recuperação de "fatos".

Todavia é, talvez, mais comum se aterem os autores no campo, por assim dizer, da lingüística documentária, a um aspecto preferencial sobre o qual incide a sua tipologia. Contaríamos, assim, num setor, com *léxicos* especiais, para cuja metodologia oferece Gardin um modelo, e no conjunto das quais quadrariam, de certa forma os textos clássicos, isto é, os grandes sistemas bibliográficos.

No outro setor — no da *sintaxe* — proliferam os inventários e categorizações de gramáticas, ou dos modelos e teorias que se lhes reportam. Eis aí o domínio, por excelência, das formalizações, às quais teremos oportunidade de retornar em outro passo.

3 - EM BUSCA DE MÉTODOS E MODELOS PARA AS LD

Com o escopo de firmar o modelo de linguagem documentária, fixar-lhe as funções e estruturas ou descrever-lhe as características há uma profusão de tentativas e a busca constante de critérios, na literatura especializada.

Que metodologia se esboça nesta, quais as tendências aí reinantes, que castas de teorias vêm a prevalecer? Eis os tópicos que, muito por alto, nos propomos perulstrar, sempre com fito de frisar as conexões da Lingüística com a Documentação.

Com prévia consideração conviria, contudo, lembrar algumas correlações acima entrevistas, e referir sempre o estudo de quaisquer códigos de indexação ao que equivaleria, quase, a umas tantas regras triviais para interpretá-los.

3.1 — *Distância variável entre as LD e as LN*

- a) Enquanto *linguagens intermediárias* — "meta-linguagens" para Zinkin (23) — *as LD distam desigualmente das LN, isto é, da sua matriz e sentido final.*

Ao tentarmos dar conta das conexões de umas com as outras tornou-se manifesta a acentuada mobilidade, ou melhor, a relatividade dos respectivos conceitos.

Constatou-se a passagem contínua dos sistemas em que sobrelevam elementos não formais, calcados na ordem espontânea, para construções mais e mais artificiais, embora conversíveis no fundo, a estruturas lingüísticas típicas da comunicação natural. "Chassez le naturel, il revient au galop" — ainda uma vez valeria a velha observação.

Da mesma forma, sob o prisma dos significantes, se deparamos com descritores onde o alfabeto ocorre tal como na escrita corrente, verificamos, igualmente, o destinarem-se as letras a notações de códigos especiais, ou o aliarem-se, não raro, aos números e a recursos gráficos vários. Isso quando não meramente substituídas por esses últimos na codificação dos termos.

Quanto à dimensão vertical ou paradigmática, a mera lista não estruturada vem a mudar-se, por gradações incessantes, nas hierarquias classificatórias e nas relações semânticas complexas, subjacentes ao discurso usual.

Assim, também, no tocante ao eixo horizontal, é força reconhecer-se o recurso a sintaxes documentárias apartadas ou aproximadas das que se formulam para as LN. Fala-se no SYNTOL, por exemplo, em "esquema sintático" — representação da estrutura de uma expressão natural sob forma de um esquema — ou na atribuição de categorias "quase-gramaticais" às palavras-chave.

Daí a inclusão freqüente, no plano em questão, de gramáticas e modelos lingüísticos propostos para as LN.

3.2 — *Concatenação maior ou menor dos descritores*

- b) *Nos produtos da análise documentária, o grau de concatenação dos descritores varia da ausência desta à complexidade da LN.*

Quando atentamos para as operações em jogo na análise do conteúdo ou, se quisermos, na representação indexada, verificamos de pronto, ainda uma vez, a maior *on* menor preservação dos textos e contextos de origem.

Na massa de instrumentos gerados para o acesso à informação bibliográfica discriminar-se-ia sempre uma dupla vertente bem caracterizada, no que concerne à indexação, por Coyaud (24), ao en-

globar este, de um lado, os que derivam da simples seleção de certos elementos incluídos nos documentos, e de outro, os provenientes das transformações das noções mais relevantes neles contidas, já que os produtos, em ambos os casos, podem consistir, quer em dados isolados, quer em seqüências de termos concatenados ou em frases, torna-se disponível enfim, ao usuário, uma gama de meios de busca onde se abarcam, quanto aos primeiros, índices, como os de permutação e seleções de frases extraídas dos textos, e, em referência às transposições, grafos de frases naturais integrais; conjuntos de termos definidos num léxico organizado, com ou sem relações sintáticas; resumo. Veríamos, nesses últimos, a versão abreviada dos textos, em LN ou não, ou ainda, o "telegraph abstract".

No seu Glossaire, Cros define de modo algo parecido, a respeito da análise documentária: o *extrato* — mera seleção de palavras ou frases; a *indexação* feita por meio de descritores, com ou sem indicações sintáticas e o *resumo*, texto abreviado em LN.

3.3 — *Formalização das LD e automação*

- c) *O reforço das estruturas lógicas e operacionais é uma função do grau de automação previsto para o sistema.*

A circunstância de apoiar-se ou não, um dado sistema, em aparelhamento automático, pode significar a consolidação ou a atenuação no modelo que o informa, das articulações lógicas, e uma elucidação maior ou menor das relações essenciais no que concerne ao ingrediente lingüístico.

Como observa L. Schultz, atribuir as tarefas de processar a linguagem às máquinas implica em conhecer aquela a ponto de permitir o "design" dos processos e a avaliação do processamento.

E um dos mestres da lingüística computacional, P. Garvin (45) ao sugerir uma taxonomia para os problemas que lhe são essenciais chega a distinções próximas das precedentes ao identificar questões de ordem estatística e de ordem não estatística, diferenciando nessas últimas, dois tipos — os textuais e não textuais — "textual data — coherent text of some or another or anything else". Se é bem verdade o que postula Upenski (apud Soergel, 97), que "everything which can be formalized can be automatic". não é menos verdade que tudo que é mecanizado deve ser de antemão formalizado.

Côncios, por conseguinte, das flutuações que manifestam, em sua estrutura, os meios disponíveis para a informação documentária, cujas implicações, enquanto linguagens, é nosso intuito sondar, e sensíveis ao impacto da automação mas não voltados para ela, passemos em revista, sob o ponto

de vista do método, o que montaria tão só a uma amostra de como têm sido tratadas. E o faremos, como a propósito de sua classificação, reportando-nos aos que as encaram na sua realidade global concreta e, em seguida, aos que se cingiram de preferência a esquemas formais ou a modelos e abstrações que lhes dizem respeito.

3.4 — *As categorias fundamentais de Grolier*

Se, nalgum momento, tornou-se mais clara a consciência de que há uma zona crítica de correlações entre Lingüística e a Documentação, pode-se afirmar que quem o marcou foi, após os debates da Conferência de Dorking, Eric de Grolier com o seu trabalho para a UNESCO, "Étude sur les catégories générales applicables aux classifications et codifications documentaires" (49).

O centro de interesse são os *códigos*, definidos como sistemas de "símbolos para representar a informação e de regras para combiná-los". Deles, a *classificação* representa a variedade cujas "palavras — conjuntos de símbolos que designam um conceito, ou semantemas — se acham adstritas a relações de ordem entre si".

Fiel ao título, e com viva sensibilidade aos aspectos estruturais das linguagens em apreço, — de que é testemunho a terceira parte dedicada ao estudo de processos análogos usados nas línguas naturais ou artificiais — esmiuça Grolier certas classificações documentárias enciclopédicas e as codificações especializadas, do ângulo, sobretudo, das "categorias gerais".

Destarte, o que de método ali se esboça com respeito à descrição consiste de um lado, em analisá-las quanto à extensão, tomando-as na sua qualidade de universais, ou não, antes de o fazer na sua distribuição geográfica; e de outro, em identificar-lhes as relações entre os termos e os processos que tendem a construí-las como um todo, e não como meras enumerações.

De fato, na consideração de sistemas conceituais funda os seus trabalhos a própria lexicologia, mais e mais estruturalista.

Que o documentalista fere ali um ponto crucial da índole lingüística da informação bibliográfica não há necessidade de sublinhar. As relações entre os termos de uma LD — no plano sintagmático como no paradigmático — eis a área mesma em que se imbricam as facetas léxicas e sintáticas, para não falar na lógica e na semântica, da comunicação.

Embora com a prudência de evitar qualquer categorização prematura, são copiosas as suas observações sobre os pressupostos das codificações correntes no universo documentário e valiosa a sua crítica do modo por que se interrelacionam com as linguagens em geral, haja vista os comentários nas notas ao fim da obra.

3.5 — *Modelo garviniano de J. Melton*

Preocupações metodológicas manifestas constituem o próprio cerne das sugestões de J. S. Melton, inspiradas num "definitional model of language" do referido Garvin (68).

A seu ver, impõe-se tal conceito estrutural porque apto a definir a linguagem e a discriminá-la como "natural" ou como atinente a outros simbolismos, inclusive matemáticos ou lógicos. Dando conta da associação da forma ao sentido, vale para o discurso oral ou escrito, e não só para identificar-lhes os elementos discretos, mas também para fazê-lo em termos de relações e não de mera listagem. Acresce, por fim, que satisfaz aos requisitos estritos do processamento no computador.

O modelo garviniano envolve três conjuntos de níveis: o de *estruturação*, com o plano fonêmico (ou grafêmico) e o morfêmico, onde se identificam as unidades, os signos da linguagem dotados de forma e sentido; o da *integração*, em que estes valem não por formarem meras cadeias seqüenciais, mas por atuarem como um todo (fused-units); o da *organização*, o grupo mais elevado.

A título de exemplo, atente-se para um dentre os que oferece J. Melton com o fito de mostrar como entram em jogo tais fatores numa LD. Assim, para o documento "A study showing the effects of the use of the tape recorder upon individual spelling", o código facetado no esquema de Tauber e Lilley (68) seria:

RDEX PLSI TSCM MDLP MVPN MSTR TLDP
MSED TCIT PEAC PEÃO

Nele, a ordem das expressões não é significativa como o é na frase inglesa "students learn French" ou "Franca students learn". A primeira letra determina o tipo da categoria "gramatical", ou seja — nota a autora — "R está para P" assim como "P está para R".

Já no "telegraphic abstract" do mesmo título — de que reproduzimos apenas trechos — os "interfixos", indicados por dois pontos, operam ao nível de integração de Garvin, dos "fused-units", e reforçam-se as analogias com o conceito gramatical de concordância: o "indicador de função" de ordem inferior (.) concorda "gramaticalmente" com o de ordem superior (...).

.. KEJ, experimental group, control group, student,
. KIS, intermediate, grade... KEC, spelling,
. KQJ, textbook, lecture, demonstration.. KAM,
testing,. KQJ, Calif At.. KAP, experimental
group, spelling, achievement,.. KUP, student,
teacher, attitude.

Nota: "English words uncoded"

Em suma, realizado ou não, o modelo em questão — não pudemos apurá-lo — afigura-se-nos, no entanto, rico de sugestões, como o reconhece a pró-

pria crítica algo rigorosa de Grolier (50) This sign [Mrs. Meltons paper] will be promising for the application of structural linguistics to documentation".

3.6 — Modelo de descrição de Coyaud

Empreendimento dos mais sistemáticos levou-o a termo, sob forma de descrição lingüística, M. Coyaud (23) na sua tese de doutorado. Adaptando terminologia e conceitos de A. Martinet (23), submeteu a exame um "corpus" correspondente ao domínio das LD, tal como historicamente ele se afirmou e expandiu. Nele, de fato, algumas das grandes classificações bibliográficas que se sucederam a partir do século XIX, figuram ao lado de sistemas que a pesquisa vem inovando na área da documentação, da informática, ou melhor, da recuperação da informação, bem assim da tradução automática e de outros setores fecundados pelo computador a serviço da linguagem.

Já que a segunda articulação da linguagem, a dos significantes, — fonemas ou grafemas — despertou-lhe pouco interesse, concentrou o autor sua atenção na primeira, na "face dos significados" (L. Prieto), atendo-se a ela nos "semas" cujas unidades era seu objetivo identificar, e isso segundo a ordem crescente em que se estruturam para cada LD.

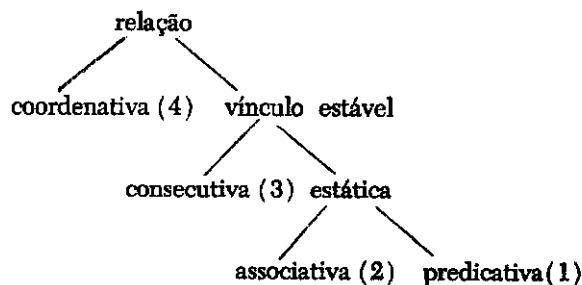
Como seqüência típica, têm-se, pois, -das menores as maiores, os monemas, tanto os lexemas quanto os morfemas, e as unidades sintáticas, isto é, a palavra, o sintagma — distinto da precedente por razões práticas — o enunciado e o discurso. Partindo, assim da forma mínima, isto é, dos monemas, considera, no que lhes toca: os semantemas, ou melhor, como os designa Coyaud (23), os "lexemes", a que, por vezes, também se refere, obediente à praxe documentária, sob o nome de *termos*; os morfemas de que, pela mesma razão, discrimina sob a denominação de *relações*, as *analíticas*, *explícitas* e *a priori*, e as *sintéticas*, estabelecidas *a posteriori*, entre as unidades do léxico, no momento da indexação.

Espécimes de *lexemas* seriam, por exemplo: na CDU, 491 "línguas indo-européias diversas", em que 4 é lingüística; no SYNTOL, a forma canônica de um termo da LN, e, na classificação para o Bureau International de Travail, de Foskett (23), Bbr "reumáticos, cardíacos, diabéticos, tuberculosos", onde B significa "Indústrias".

Quanto aos morfemas, a *relação analítica do indivíduo à atividade* exprime-a, neste último código, a minúscula *b*; na CDU, com respeito à classe 4 apontada acima, o "ponto de vista genético" aparece através de 91, ou o "tipológico", de 494/499, línguas isolantes e aglutinantes".

Com a função dos *morfemas sintéticos*, o símbolo

": " indica, na CDU, "relação geral" não orientada, entre os morfemas, e "+", a "relação de reunião". Se, por um lado, no esquema de Foskett, as nove facetas traduzem "relações sintéticas", como A, "Occupational safety and Health: General"; B, já mencionado, "Industries, special classes of workers"; C, "Sources of Hazards" e assim por diante, por outro lado, no SYNTOL, a forma geral para as estruturas em exame é:



Amostras, enfim, das relações sintáticas, ou seja, das que surgem ao deixarmos o plano dos monemas, são as que apresentamos a seguir, atinentes à combinação das formas mínimas em unidades superiores.

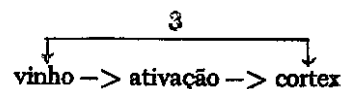
No que diz respeito às palavras, teríamos bem caracterizadas, no código semântico da Western Reserve University como um conjunto de lexemas afetados de um caso, BWFL. 5X. NUBT. 001, "anti-hemorrágico" e, no tocante a *smiagmas*, em SYNTOL, para "ação do hormônio sobre o rim", "hormônio —> rim".

No código de Pagès (23) iw equívale a "propriedade (w) do comportamento (i)".

A formulação, em SYNTOL, do enunciado completo "o vinho ativa o cortex":



comportaria, como expansão, o sintagma complementar E1 E2



De modo geral, do ponto de vista da metodologia e da capacidade de penetração nas estruturas lingüísticas analisadas, nada existe de comparável ao trabalho de Coyaud. Podemos afirmá-lo sem nos opor, contudo, à observação de Grolier (50): "La thèse de Maurice Coyaud... est une utile mise au point qui s'attache cependant beaucoup

plus à la dimension syntagmatique de ces langages qu'à leur dimension paradigmaticque".

3.7 _ Modelos matemáticos em lingüística

Uma das mais fortes tendências na lingüística em geral, é por certo, a formalização, isto é, no tocante a seus objetos, a representação esteiada numa linguagem formal, a que se alia a construção de modelos, notadamente dos matemáticos.

Nesse quadrante das ciências humanas, as entidades e processos que lhes constituem o interesse específico podem ser expressos — e o vêm sendo crescente — em termos de teorias lógicas e quantitativas recrutadas na bateria de meios à disposição, hoje, do rigor da análise. Axiomas, definições, teoremas, por um lado, modelos determinísticos ou probabilísticos do outro: eis o instrumental que se oferece ao investigador.

As vantagens inegáveis e os inconvenientes eventuais da adoção de tais métodos, pesou-os Edmundson com a precisão requerida por quem lhes sintetiza o espectro amplo e esboça, por assim dizer, um modelo do próprio modelo.

A seu favor há a compatibilidade que asseguram com as demais teorias e modelos, o indicarem o tipo de dados e de métodos a serem usados, o permitirem a manipulação das variáveis ou o enfoque global, sem contar, o facultarem o recurso às técnicas de processamento eletrônico.

Os erros eventuais a que podem conduzir, são, de certa forma, o reverso de suas qualidades: omissão ou não-pertinência das variáveis, incorreção nas relações ou nos valores atribuídos aos parâmetros, inflexibilidade e incoerência, quando não a abusiva identificação com o mundo real.

Tem-se beneficiado, por conseqüência, o estudo das LD com a incessante dialética — ousamos dizê-lo — entre os modelos, propostos no plano dedutivo,

e o mundo real, donde resulta, como frisa Edmundson, o perene reformulá-los por via também indutiva.

Os formalismos múltiplos agrupa-os o mencionado autor em modelos *determinísticos*, a saber, os lógicos, analíticos, algébricos, geométricos, e modelos *estocásticos*, ou seja, probalísticos e estatísticos. Desde a lógica das proposições e dos predicados (Bar-Hillel, 32), da teoria dos conjuntos (Mooers, 32), das relações (Hillmann, 32), das transformações (Chomsky 32), ou da álgebra dos semigrupos, até a álgebra booleana, a topologia, a teoria dos grafos (Tesnière 32, Lccerf 32) e a probabilidade lógica (Carnap 32, Bar-Hillel 32), toda uma gama de instrumentos matemáticos tem prestado eventualmente a sua contribuição fecunda.

Menos, sem dúvida, do que em teorias gerais, redundo o recurso a tais instrumentos metodológicos em formalizações parciais atinentes ao que Edmundson denomina entidades lingüísticas. Colima-se, no modelo matemático, a representação abstrata de um objeto ou processo da LN ou da LD. Gramáticas transformacionais, categoriais, análise de dependências, sinonímia, informação semântica, indexação documentária — citamos ao acaso — eis alguns processos e entidades traduzidos numas ou noutras das citadas linguagens lógicas ou quantitativas.

Á guisa de ilustração não apenas dessa ordem de modelos, mas do modo por que o citado autor os apresenta, escolheríamos o que R. Fairthorne (32) propôs, com base num reticulado, isto é, num sistema parcialmente ordenado (S; R) em que cada par de elementos x e y possui uma interseção $x \wedge y$ e uma reunião $x \vee y$. "Quando os elementos x, y, \dots são interpretados como "entradas" (heads) do "Thesaurus" e a relação de ordem parcial como "menos geral que", expressam-se as relações hierárquicas como propriedades dos reticulados":

Reticulado

elemento x

relação de ordem parcial R

interseção $x \wedge y$

reunião $x \vee y$

Thesaurus

"entrada" do Thesaurus x

relação "menos geral que"

a mais geral das "entradas" que é menos geral que x e y

a menos geral das "entradas" que é mais geral que x ou y

DESCONSTRUIR O TEXTO PARA
REFERENCIAR O CONTEÚDO

Delineado — ao menos como tentativa — o espaço em que atuam as LD, ressaltadas as feições que as caracterizam e o esforço envidado para dotar de metodologia adequada a sua formulação e estruturação, voltaremos a atenção dentro das fronteiras que demarcamos, para determinados componentes do sistema de informação — quase diríamos, para convertê-la à escala do mundo moderno e inspirando-nos em Servan-Schreiber (93) — do sistema de "superinformação".

Assim, as relações entre a lingüística e a documentação, põe-nas a nu, de certo modo, a *análise do conteúdo*, a bem conhecida atividade cujo fim é "a pesquisa técnica, com mira na descrição objetiva, sistemática e quantitativa, do conteúdo de um documento".

4.1 — *Indexação e classificação na análise do conteúdo*

Tal operação entrelaça-se, ao que vimos, com a recuperação da informação, onde desempenham função capital o indexar e o classificar o teor dos textos, bem assim o das questões que se lhes reportam.

Problema crucial, mas que transcende o plano estritamente lingüístico é, nessa esfera da comunicação, o da "aboutness" a que se refere Fairthorne (34), ou seja, o da "atênência" dos documentos, o da natureza daquilo sobre que versam e a propósito de que os solicitam os usuários. O julgamento do conteúdo tem por substrato um fundo amplo de experiência social e psicológica, mas apoia-se na evidência dos textos quanto às regras de linguagem, conceitos, interpretações, classe e relações.

Ora, a indexação é segundo o mesmo autor (34), processo extremamente complexo, "inextricável da classificação — isto é — estrutura" do campo do discurso sobre que incide, e da expressão verbal dessa estrutura.

No entanto, a fim de ordenar o exame que motiva o presente trabalho, centraremos, nesta seção, sobre o primeiro tipo de análise, as nossas observações, cingindo-nos, na próxima, ao segundo. Para isso viria a calhar lembrássemos as argutas reflexões de J. Perreault (76) com o fito de formar possível critério para distinguir a *indexação* de quanto está, antes, na alçada da taxonomia dos assuntos.

A tradução dos registros e materiais escritos numa *representação* em linguagem documentária — eis o que tanto define uma quanto a outra das citadas modalidades de tratamento bibliográfico. E o criador de um sucedâneo ou substituto (surrogate) para o texto, que coisa tenta senão indicar, de antemão, as orientações de cada "item", com mira

no liberar a mensagem virtual e no fazê-la alcançar o receptor a que se destina?

Antecipa-se a pertinência ao "documento Z para a questão A", o que traz consigo, aliás toda a problemática do julgamento de valor a que aludimos. Dentro desse gerar subprodutos representativos — dessa ação de escrever "(—) scription", como curiosamente a designa Perreault (76), — teríamos, na *indexação*, a representação dos *elementos dos documentos*, na *classificação*, a dos documentos como *um todo e do "corpus" como um todo*. Mais precisamente, no primeiro caso, surgem duas eventualidades: ou bem o indexador *adscreve* (a-scribes) conceitos aos textos, com base num vocabulário externo, ou bem *descreve-os* (ife-scribes) com as palavras que lhes são próprias, extraídas, pois, do vocabulário interno.

Na classificação, ao contrário, é típica a "hierarquia". Nela, ao menos, com um objetivo último, colima-se a *prescrição* (prescription) de posições dentro de uma organização conceptual sistemática, bem assim a *inscrição* (in-scription) dos documentos em semelhantes posições e a das respectivas "indicações de posições" nos documentos.

Note-se, todavia, que, para indexar, quando se compila léxico documentário controlado, do tipo do thesaurus, onde os conceitos, em forma canônica, se estruturam também por relações de gênero a espécie, introduz-se, implicitamente, um esquema de classe, não integradas — é verdade — num todo único, mas, nem por isso, menos articuladas. Não obstante, a indexação, via de regra, aparece como a atribuição de um conjunto de identificações, etiquetas, palavras-chave, unitermos, descritores, ou o que for — aos elementos significativos dos documentos — quando ela não os assinala tão só pelo destaque — tudo com o escopo de assegurar a identificação do conteúdo e um meio de "recuperá-lo" oportunamente.

Também para o processo em causa têm sido aventadas formalizações com que se lhe busca acrescer o rigor, e o dos sistemas que o abarcam. Landry e Rush julgam, por exemplo, que a "teoria geral da armazenagem e recuperação da informação" se formularia, com mais justeza, em termos de uma "teoria geral da indexação" e, essa, nos dá teoria dos conjuntos.

Redefinem, assim, a coleção ou acervo como um "espaço documental D", um "document space D", união de dois conjuntos: o dos "documentos entrados" (input documents), $D_{i..}$, e o dos "documentos de análise", $D_{..j}$ isto é "o próprio índice, o thesaurus, o analisador sintático e semântico etc": $D = \{ D_{i..} \cup D_{..j} \}$

Trata-se, como se depreende, de conjunto que contém a sua própria descrição: o índice dos documentos da coleção é, também, "um documento e, portanto, um membro do espaço documental".

Os "dados" - quaisquer entidades, digamos, as "palavras" — vão ter ao índice sob forma de LN, com fortes relações sintáticas e semânticas que se reduzem, marcadamente, após a indexação.

A linguagem do índice, a seu turno, consiste do conjunto de elementos de vocabulários que surgem, não raro, como uma lista ordenada de palavras (símbolos), conjunto esse reunido ao de relações de diferença, de equivalência, do tipo parte/todo, genéricas e intencionais.

O modelo atinge, por via de noções e teorias como a da comunicação, da entropia e outras da mesma ordem, elevado grau de refinamento, embora o conceito em que culmina — a de que a "função essencial do índice é a representação acurada e completa do documento" — seja, na opinião mesma dos autores, altamente intuitivo.

Não nos deteremos, entretanto, em tais formulações abstratas. Inclinados, como estamos, a prosseguir do menos para o mais complexo, ou do discreto para o unificado, e a seguir o trajeto "palavra — conceito — assunto" (104), relegaremos para comentário posterior os fatores predominantemente semânticos e as relações que eles ensejam.

Em consequência, ocupar-nos-fam desde logo, após o exame sumário de algumas técnicas fundadas em correlações estatísticas, que aproveitam à análise do conteúdo, os aspectos nitidamente sintáticos. Não é aí que se situa aquela "armature sacrée de Ia syntaxe (sujet/prédictat)" a que se refere Roland Barthes?

Ainda a esse propósito conviria advertir para o que acarreta, ou exclui, a mediação da máquina. Como hoje se dá em quase todos os passos da atividade bibliográfica, ocorre igualmente, na sorte de análise em foco, a par da persistência dos procedimentos "manuais", prevaletentes, sem dúvida, até o presente, o recurso cada vez mais franco ao processamento automático da informação lingüística. Dentro desta, com efeito, pode-se delimitar, com Garvin (46), ao lado da sub-área da tradução automática, outra, concernente ao "*content processing*", à elaboração mecanizada de resumos e à produção de meios de seleção afins. Ela partilharia com a precedente a necessidade de rotinas de reconhecimento lingüístico que gerem a análise da estrutura no tocante às "sentenças" do texto submetido à unidade de processamento.

Se na primeira é de regra o apoio em dicionário bilíngüe, na segunda o léxico será monolíngüe e incluirá, além de código para a gramática, um código semântico capaz de chamar a subrotina apropriada.

4.2 — *Aspectos da indexação e classificação automáticas*

Mencionaremos antes, contudo, a existência de uma família de métodos com que, no domínio em

pauta, se efetuam a indexação e a classificação automáticas. H. Borko estuda quatro espécies destas últimas: estatística, por permutação, por citação e por associação.

Pressuposto da primeira — a estatística — é que há uma relação, constatada por Luhn (105), entre a frequência com que ocorre determinada palavra no teor de um documento — salvo, naturalmente, a pertencente a categorias genéricas — e a medida em que é um "indicador significativo de um assunto, e vice-versa" (41).

Que há, nisso, o risco de excessiva simplificação, eis o que, de imediato, se patenteia. Donde várias técnicas para acrescer o rigor da indexação, mediante o *cômputo de co-ocorrências de termos*, porque mais significativas que o das meras palavras isoladas, a *seleção, no texto*, das "sentenças com a maior concentração de frequência dos termos e dos pares de termos" (105), ou, ainda, o *correlacionamento* da frequência de um termo no texto com a sua frequência no uso geral.

Já tratamos pela rama, em outro ponto, do método de permutação das palavras, adotado nos índices do tipo KWIC e similares.

O da associação, pretendemos retomá-lo, de relance, sob o prisma da classificação.

Quanto à citação, se já foi argumento, ao tempo do "magister dixit", se ela é o próprio pulso e o deleite do pensamento montaigneano, vêmo-la, hoje, assumir a condição de instrumento preciso para a análise do conteúdo.

Em toda uma "ciência das ciências", numa sociologia da pesquisa, num cálculo bibliométrico se funda o seu valor como indício do que encerram os textos. Como bem sintetiza G. M. Braga: "as citações de ordem de série mais alta levariam à determinação [de um assunto] e as mais baixas à das áreas correlatas".

Não entra, porém, no quadro da presente exposição, voltada antes para os aspectos sistemáticos do código lingüístico, aprofundar essa faceta da comunicação humana, embora nos pareça indiscutível — para citar uma das autoridades na matéria, Guiraud (63) — "que se deve considerar a estabilidade dos signos lingüísticos como uma evidência, um fato cuja observação adquiriu doravante o caráter constante e universal".

4.3 — *Modelo abstrato e tipos de gramáticas*

A intervenção da máquina na indexação fez sobressair o papel da análise sintática e da gramática que esta presume quando o que se tem em vista é inquirir o conteúdo do discurso através de cuja concretização e coerência formal se dá a fluidez da comunicação.

Se determinar os conceitos que o texto a descrever contém — quer o dos documentos, quer o dos pedidos — interessa a *análise semântica*, identificar

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

as relações **que** ligam tais conceitos no plano da linguagem é o próprio da *análise sintática*. O conteúdo assim caracterizado, urge "especificá-lo nos termos do formalismo que constitui a linguagem interna do sistema". É o que mostra C. Montgomery (70) evidenciando, a cada passo, o necessário entrosamento entre os dois processos, como já tivemos a oportunidade de assinalar.

Mas não há esquivar-se — nesse setor ainda menos do que nos demais — aos requisitos da formalização, razão pela qual M. Gross, por exemplo, busca construir, com referência a L, isto é, à linguagem natural, uma gramática que "enumere (gere) as sentenças de L e as reconheça, quando associada a uma rotina de reconhecimentos". O resultado, te-lo-emos sob forma de descrição estrutural. O conjunto finito de palavras (morfemas) de L constitui o léxico V de L e, em V, a operação de *concate-nação* define o conjunto C(V) de "strings" ou cadeias.

Como modelo abstrato geral postularemos com Gross:

- o subconjunto L de C(V) é uma linguagem em V: $L < G(V)$
- a cadeia S tal que Se L é uma sentença de L;
- o conjunto finito de regras que caracteriza todas as sentenças de L e apenas L, é uma gramática de L;
- duas gramáticas são equivalentes se caracterizam a mesma linguagem L.

Quais os critérios para avaliá-las? Garvin (46), seguindo os passos de C. Montgomery, assinala entre as qualidades empíricas a "de darem conta de todas as sentenças gramaticais e apenas delas",

e entre as formais a de se conformarem à simplicidade, à recursividade, isto é o facultarem um conjunto particularmente infinito de condições a partir de umas poucas operações lógicas ou aritméticas, e a "algoritmidade", a que se deve seja a computação exata, garantida e única.

Dois grandes grupos englobam a multiplicidade de gramáticas: as *formais*, fortemente orientadas para um modelo, e as *descritivas*, que o são francamente.

Formais são as de estados finitos (Hockett, 46), de estrutura da frase (Harris 46), de dependência (Hays 46), as "formacionais", transformacionais (Chomsky 46), estratificacionais (Lamb 46). Pertenceriam às descritivas as tagmêmicas (Elson e Pickett 46), as glossemáticas (Hjelmslev 46), de item-e-arranjo (Trager 46) e de palavra-paradigma (Oettinger 46).

4.4 — Exemplos de gramáticas

Proliferam, pois, ao que se vê, as gramáticas de que a resenha de Garvin constitui um levantamento até 1965. Estaria fora de qualquer cogitação esmiuçá-las na sua riqueza e variedade. Mas sem ir além de um relance sobre o modo como uma delas vem a operar, *verbi gratia*, para a recuperação da informação, não é descabido reproduzir-se aqui as regras de escrita com que ilustra S. Kuno, valendo-se de um exemplo clássico, uma "phrase structure grammar".

A partir das regras extraídas da *coluna 1*, proceder-se-ia a derivação descrita na *coluna 2*. Dentro da própria cadeia dá-se uma "left-to-right expansion":

Coluna 1

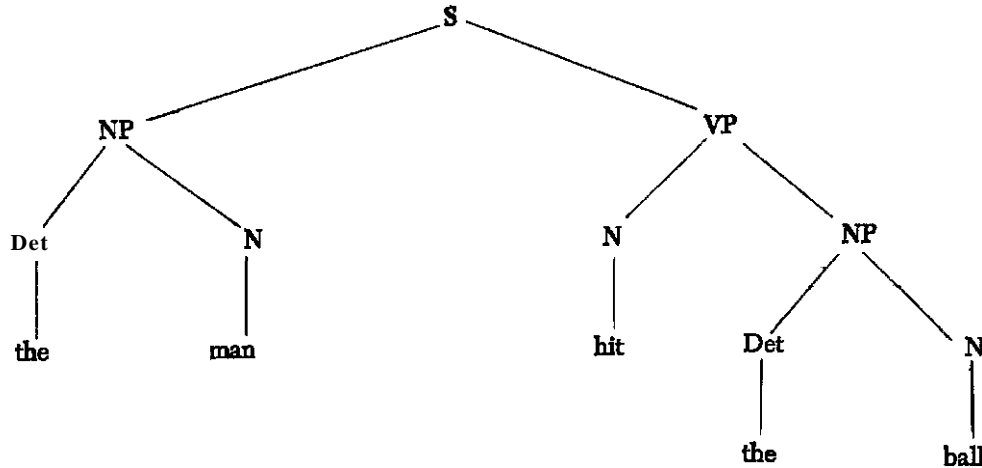
- Regra 1 S -> NP VP
Regra 2 NP -> Det N
Regra 3 VP -> V NP
[Regra 4 Det -> a]
Regra 5 Det -> the
Regra 6 N —> man
Regra 7 N —> bali
Regra 8 V -> hit

Coluna 2

- S
NP VP
Det N VP
the N VP
the man VP
the man V NP
the man hit Det N
the man hit the N
the man hit the bali

Como observa E. Arcaïni "a derivação pode ser representada também pelo diagrama, que fornece

menos informações que a *coluna 1*, pois nele não se precisa a ordem em que se "re-escreve" a cadeia.



A correspondência com o português dá-mo-la a seguir:

- S = Sentença
- NP = grupo nominal
- VP = grupo verbal
- Det = artigo
- N = nome

S. Kuno mostra que se alcança, assim, estrutura de todo equivalente à atribuída à frase em apreço pela análise dos "componentes imediatos", ou seja, segundo a notação adotada, ((the man) (Ut (the ball)))

The II man I hit II the HI ball
the man hit the ball

Quanto a formalizações, acrescente-se apenas a seguinte: uma gramática "sensível ao contexto" abarca um conjunto *P* de regras de "re-escrita do tipo: $\alpha A\beta \rightarrow \alpha \gamma \beta$

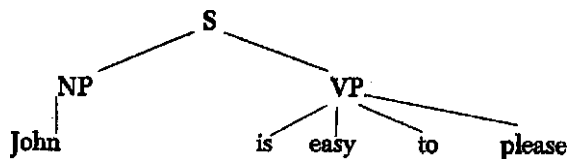
A é membro do vocabulário não terminal *V_n*, ao passo que α, β, γ são cadeias de símbolos de *V_N* e do vocabulário terminal *V_t*, podendo α e β serem nulos, não porém γ .

Uma gramática "context-free", independente do contexto, equivaleria a $A \rightarrow \gamma$, isto é, α e β seriam nulos; outra, mais restrita ainda, compreen-

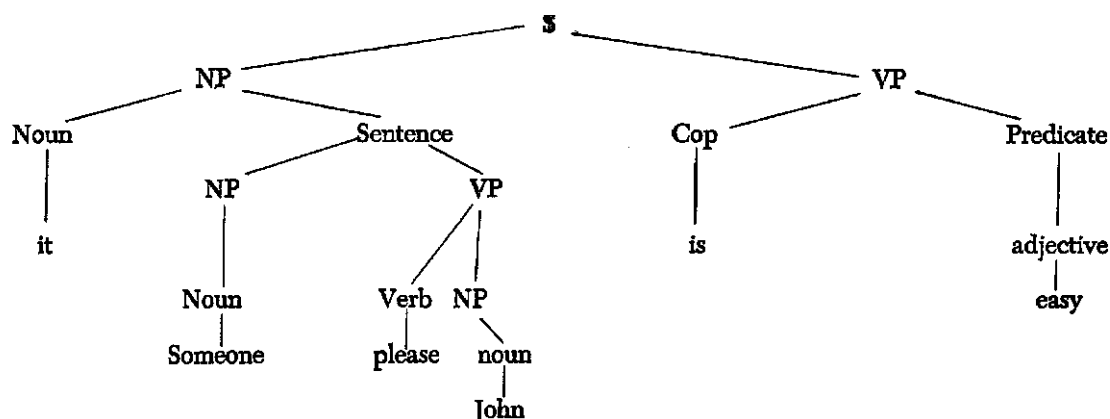
deria regras da forma $A \rightarrow X$, onde *X* seria, uma cadeia de símbolos de *V_t* apenas. No caso acima, *V_t* = the man, hit, ball; *V_n* = S, NP, VP, Det, N. Ainda que descartada — tomamos a dizê-lo — qualquer veleidade de retratar aqui a incessante evolução de tais gramáticas, bastaria para dar vaga idéia de como se ampliam com elas as perspectivas para a documentação, lembrar com S. Kuno ou C. Montgomery (70), o significado, por exemplo, da obra extremamente fecunda de Chomsky (58), e o poder acrescido, por seu intermédio, — que os "transformational parsers", ou melhor, as gramáticas em que se fundam, adquirem em face das de "estrutura de frase".

Com efeito, a gramática transformacional atribui dupla estrutura a cada sentença que venha a gerar: uma *profunda* — que lhe determina a interpretação semântica — e outra, *superficial*, sua realização efetiva e fonética no discurso. Destarte, a estrutura profunda que corresponde à superficial, expressa pela frase "John is easy to please", é a que indica que "John" é objeto de "please", que o sujeito de "please" é "someone" e que o sujeito de "is easy" não é "John", mas "someone... please John".

Não se trataria, pois, de



válida, também, para "John is eager to please", mas de:



Resolve-se, por esse modo, uma série de ambigüidades, uma vez que a estrutura profunda de "John is eager to please" difere dessa última, embora as superficiais de uma e outra sejam, como se mostrou, iguais.

4.5 — Um sistema de análise de conteúdo

Como se integram, na prática, os métodos de análise de conteúdo? Dentre os sistemas de informação fartamente descritos na literatura, sobressai, como um quase protótipo, o SMART, a que se vincula o nome de G. Salton (91). Utilizando um IBM 7097 e um 360 - modelo 65, a sua organização destina-se inclusive a avaliar a eficiência dos vários processos mediante a medida da proporção de material de fato recuperado (recall) e a do material recuperado e pertinente (precision). Para avaliar-se algo do que, em matéria de meios e facilidades, a indexação automática — *sem nenhuma análise manual prévia* — vem a requerer, note-se que o SMART compreende: *dicionário de raízes (stems)*, de *afixos (sufixos)* nos documentos, de *palavras comuns*, a serem suprimidas no verso da indexação, um *thesaurus*, com que se eliminam os sinônimos graças à atribuição de "concept numbers", reduzindo-os, pois, a conceitos; *arranjo hierárquico* dos termos do Thesaurus, a fim de generalizar, especificar ou correlacionar os conceitos; *associação estatística*, para o cômputo das co-ocorrências a que já nos reportamos; *métodos de análise sintática*, destinados a confrontar sentenças e questões com um dicionário de estruturas sintáticas — "Criterion trees" — conferindo-lhes no caso de igualdade semântica o mesmo número;

métodos de reconhecimento estatístico das frases. Salta à vista o caráter eclético do sistema, construído para assegurar até mesmo o controle dos conceitos e significações, aspecto que é nosso intuito considerar adiante.

Um dos seus limites — e isso nos conduz à questão semântica — é o de que o Syntactic Analyzer de Kuno-Oettinger, nele aplicado, a saber, uma gramática "independente do contexto" e processamento de cima para baixo com estrutura de pilha (64) ("pushdown store") tende a produzir múltiplas análises de sentenças antes simples.

5 - DAS ESTRUTURAS SUPERFICIAIS AO NÍVEL SEMÂNTICO

Ao isolar certos grupos de palavras nos limites da sentença, a análise sintática não consegue sempre gerar uma interpretação semântica única: fogem-lhe, não raro, as relações mais finas entre os componentes das frases e, em particular, entre as frases.

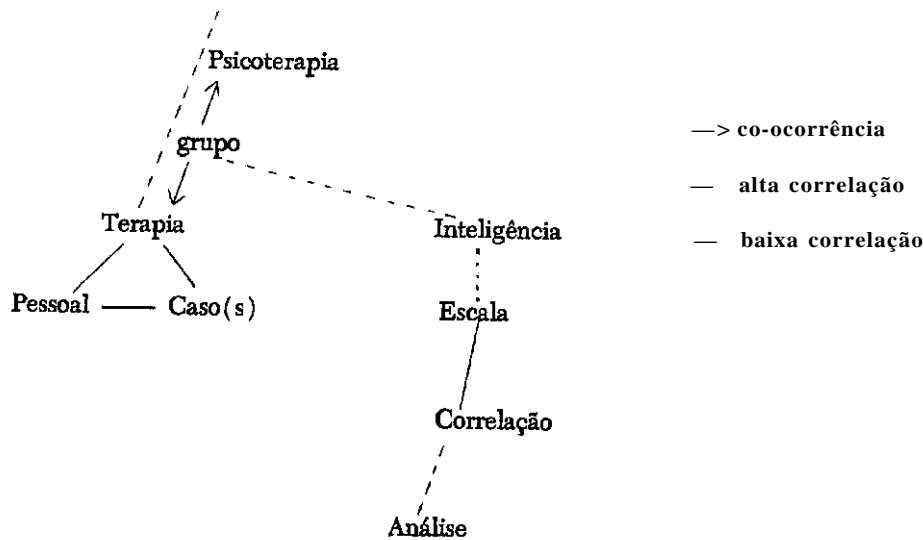
Se as gramáticas transformacionais são capazes de relacionar proposições como: "Smith wrote the report" e "the report was written by Smith", nenhuma delas é capaz de relacionar ambas com "Smith prepared the report" — relação que é crucial para a ciência da informação, ao que observa C. Montgomery (70).

Donde a procura de soluções para semelhantes impasses e a tentativa de chegar efetivamente ao nível semântico na análise do conteúdo. Há, para registrar, até o presente momento, além do tradicional e preponderante recurso a um vocabulário de indexação dos assuntos, que a léxico grafia e

a prática biblioteconômica consagraram de longa data, uma série de experiência e, desde logo, as atinentes à análise por associação, a que já aludimos, ou à classificação automática.

5.1 — *Análise de assunto: associação e classificação automática*

Esteia-se a primeira na possibilidade sugerida por Doyle de representar, derivando-as do texto por computador, associações entre os termos dos índices. Dá-se-lhes um formato de diagrama ou mapa, de que o segmento reproduzido a seguir, com respeito à "Literatura psicológica", fornece uma imagem (14):



O usuário explora o mapa até encontrar o "cluster" (conglomerado) de palavras atinentes a seu tópico. Com respeito ao outro processo, também se lança mão, nele, da técnica de co-ocorrência, mas para a construção de matrizes de conceitos correlativos. Os "clumps" daí resultantes formam "uma espécie de classificação firmada estritamente na coleção de documentos." A seguinte mostra de associação dada por Foskett (41)

"gramática, paradigma, classes gramaticais, adjetivos, proposições, frase, indicador de frase, tempo, fim, raiz, sintaxe, diacrítico." "faria as vezes de uma classe, análoga à convencional, onde, sem embargo, "faltariam muitos focos" (nome, verbo, cláusula) simplesmente porque não ocorrem na coleção documentária. Por valiosas que sejam tais análises estatísticas — e incorporáveis a sistemas mais amplos, onde se lhes confia função como a do "feedback" — não parecem proporcionar a almejada descrição do conteúdo semântico.

Já demos, em síntese, configurações do tipo SMART,

de Salton (91), onde se mesclam facilidades de ordem sintática a meios específicos de acesso à significação e aos conceitos das sentenças. Todavia, uma das objeções que C. Montgomery (70) lhe opõe, além da apontada, é a de que sistema com potencial maior de análise seria imprescindível em áreas documentárias mais homogêneas quanto ao assunto, no tocante às quais — "representações do conteúdo e das questões sob forma de combinações booleanas de termos revelam-se claramente insatisfatórias".

5.2 — *Teorias lingüísticas com bases semânticas*

Setor rico de sugestões para a criação de modelos mais sensíveis aos fatores conceptuais é o da "recuperação de fatos" e o da produção de respostas a questões em LN, de que Simmons apresentou clássica resenha. O próprio autor usa o "procedimento econômico de tomar cada componente tão logo se venha a considerá-lo sintaticamente aceitável e a testá-lo quanto à boa formação (well-formedness) semântica".

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

Por outro kdo vem recrudescendo a atividade le-xico-lógica não abandonada, aliás, pelos investi-gadores soviéticos que, ao que consta, jamais sa-crificaram a semântica à fonologia e à gramática. Passa a teoria lingüística a não concentrar-se apenas em modelos gramaticais, tão ao sabor da orienta-ção reinante, e conformes à preferência marcante da formulação matemática pela sintaxe.

Já nos reportamos, anteriormente, à nova dimensão conferida a essa ordem de investigações ao se introduzirem, no modelo da linguagem, estruturas ditas superficiais, a par de outras, atinentes à "si-tuação conceptual subjacente". R. W. Langacker, que assim se exprime, esquematiza tal organização lingüística na figura:



Constata-se não obstante que, ao fascínio da sintaxe, não escapou até mesmo, na sua faceta mais recente, a investigação Chomskyana, tão fértil indiscutivelmente.

"A noção de estrutura profunda como nível básico do componente sintático, não se acha bastante apartada da estrutura idiosincrásica de superfície para refletir categorias e relações de fatos universais", como a argúcia de C. Montgomery (70) acentua. A busca dos universais acusa menos universalidade que era lícito esperar, poderíamos dizê-lo.

No entanto, a consciência de que não há rejeitar a

semântica como extralingüística, ou separá-la do componente sintático, se torna tão imperativa que D. Varga (70) afirma, sem hesitar: "não erraríamos em representar o processo do entendimento como uma seqüência de transformações semânticas que modificam e interrelacionam blocos semânticos distintos".

Tudo se passa, na verdade, como se houvesse entre os domínios da linguagem comum, da linguagem artificial e do pensamento respectivamente da alçada da lingüística, da lógica e epistemologia e da psicologia — a correspondência que O. Clouzot resume no quadro abaixo:

Linguagem comum	Linguagem artificial	Pensamento
Palavra	Termo	Conceito
Frase	Proposição	Julgamento

A área de tensão se situa — percebe-se logo — entre o plano conceptual, num extremo, e o das relações entre os monemas, no outro. Um dos problemas capitais é o paralelismo entre os significados e a organização sintática.

5.3 — *Predicado: Formalismo de Fillmore*

De interesse para a questão em foco, tem havido a proposta de sistemas numerosos de processamento do texto, onde a noção de *predicado* assume o valor privilegiado de núcleo de toda a análise. Ao contrário do que se admitia na lexicografia tradicional, em que somente as conjunções e os elementos semanticamente pobres, denominados por Pagès de "mots-outils" — palavras-ferramenta — eram tidos como termos de ligação, deve-se considerar, além do tradicional predicado univalente, com um só argumento, predicados com dois ou mais argumentos. "Pedro come pão", ou, "consumo de pão por Pedro", eis uma relação do tipo P (x, t). Numa linguagem de informação que visa à sumariação (langue de résumé), os termos de ligação da mensagem — o seu tecido conjuntivo, na imagem de Pagès — embora participem da sua substância podem ser representados pelo *predicado* apenas, omitindo-se os argumentos. Assim, "consumo" viria a satisfazer, em economia, e uma notação optativa seria R. e não do tipo (x) como pão = x.

O sistema proposto por Noël, em 1966 — C. Montgomery (70) toma-o como referência para a sua "Metateoria" — relaciona as estruturas de superfície dos resumos — a "linguagem-objeto" — com as das "entradas" providas pela classificação — a "metalinguagem". Surgem ali as estruturas profundas como instruções ou proposições "relacionais", na mesma linha das relações de "função" (role) de Fillmore (70), e das relações sintagmáticas e paradigmáticas do SYNTOL, a que já aludimos.

A própria C. Montgomery (70) analisa o formalismo de Fillmore (70), que lhe parece possuir a "maior força explanatória, bem como o mais explícito mecanismo para vincular a linguagem formal à LN". Proposição e modalidade estruturam as sentenças, a primeira constituída pelo predicado e os respectivos argumentos nominais, inclusive outras orações, a segunda pelo tempo, modo e negação.

O conjunto ordenado de categorias de "casos" atinentes à estrutura profunda enfeixa — consoante a inovação de Fillmore — as noções básicas atinentes à "função" e aptas a refletir os julgamentos humanos sobre eventos e situações.

Eis algumas delas: *agente* (A), causa — animada — do evento; *instrumento* (I), causa eficiente, instrumento; *paciente* (E "experiencer"), quem recebe o efeito da ação; *objeto* (O), o conteúdo neutro da experiência; *fonte* (S - "source"), estado, local

de origem; fim (G — "goal").

A título de ilustração, teríamos:

- a) Harvey quebrou o espelho com um martelo
(A) (O) (I)
- b) O martelo quebrou o espelho
(I) (O)
- c) O espelho quebrou-[se]
(O)
- d) Anne deu dinheiro a Claire
(A) (O) (E)

Classificam-se os predicados em termos de "estruturas de casos" (case frames). Assim, para "dar" notar-se-ia [— AOEj, para "quebrar", [— (A) (1) O], expressos os "roles" facultativos pelos parênteses.

Toda uma rede de articulações não propriamente sintáticas, mas de significados, vem configurar-se dessa maneira, tanto mais complexa quanto outros níveis de descrição semântica entram em jogo, além da "gramática de casos" e das noções de função subjacentes.

Com efeito, o próprio Fillmore (70) esmiuçou as "pre-condições" para o uso apropriado das sentenças — as "pressuposições" implícitas em certos verbos. Assim, em "julgar", "criticar", "censurar" e similares, os parâmetros da situação de "incorrecção" (badness) envolvem os de "responsabilidade".

Por outro lado, Leech (70) definiu relações lógico-semânticas, por exemplo, a implicação "dedutiva" (a inclui c, se a asserção *aRb* implica a asserção *cRb*), e a "indutiva", que incide sobre argumentos e predicados. "Gosto de frutas" implica, "dedutivamente", "gosto de maçãs".

5.4 — *Metateoria de C. Montgomery*

Ao integrar essa ordem de elementos na sua "Metateoria", postula C. Montgomery (70) que tanto as relações lógicas quanto as outras podem ser representadas por proposições "relacionais" concernentes à estrutura profunda, e requerem tradução "em termos de metalinguagem que especifique um universo particular do discurso, um subconjunto de enciclopédia".

É aí que interviriam as classificações documentárias, de grande relevância para a análise de conteúdo da linguagem natural e, mais especialmente, do ponto de vista que nos interessa, para a integração da teoria lingüística à da ciência da informação. O que é próprio delas é fornecerem explicitamente relações de classes com que se definam os superconjuntos em face dos subconjuntos e vice-versa.

Discrepando, não obstante, de Noël (70), que vê em semelhantes classificações, metalinguagens propriamente ditas, de natureza "pública", elas mesmas caracterizadas por estruturas superficiais e profundas, C. Montgomery (70) julga-as antes representações tão somente aproximadas de uma metalinguagem em "linguagem-objeto", isto é, linguagem natural. Não incluem todos os termos desta, que inclui, ao contrário, todos os da metalinguagem.

Se algumas "relações semânticas" são expressas na chamada "estrutura de superfície" da metalinguagem, e não na da linguagem-objeto, nem toda a informação relacionai o é, como as pressuposições e os "roles" a que nos referimos.

Portanto, o que se afigura de maior significação para a autora citada, o que têm omitido as investigações semânticas é a urgência de fundamentar, numa classificação documentária, a elaboração sistemática de metalinguagem apta, pelo fato de especificar um subconjunto particular da "enciclopédia", a definir as relações lógico-semânticas e a interpretar as sentenças em LN.

Atingido, como acabamos de fazê-lo, o plano da semântica, intentamos mostrar como esta, se a investirmos de suas conexões com a ordem conceptual e lógica, vem a fundar e a informar, por assim dizer, a teoria — ou se quisermos, a "meta-

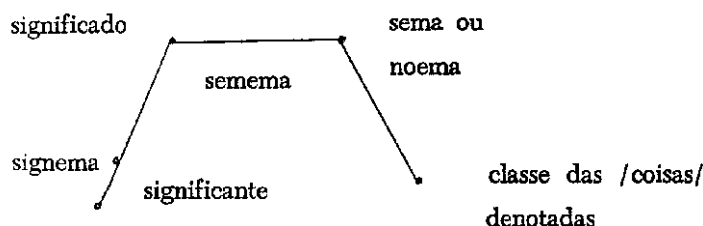
teoria" — e a prática documentária, em especial quanto à análise dos assuntos.

5.5 — *Polissemia, Jiomonímia, sinonímia, definição*

Na área de problemas assim demarcada, algumas questões, já emergentes em outros pontos das presentes considerações, parecem encontrar o terreno de eleição. Embora sem analisá-las sistemática ou exaustivamente, não nos esquivaremos a examinar, num relance, suas incidências na esfera, em apreço, da informação bibliográfica.

Não fosse essa cautela e nos deteríamos a tentar expor, a seu respeito, o afã generalizado de construir-se um modelo semântico do signo lingüístico. K. Heger, justamente, adota como pedra de toque para a sua teoria — nada menos do que meta-metalingüística — os problemas da polissemia, da homonímia e da sinonímia. E não são outros os tópicos que, após tratarmos da definição e antes de nos alçarmos aos das relações ainda nos resta perlustrar rapidamente do ângulo das LD.

Mostrou-lhes as interconexões a conhecida versão trapezoidal do modelo de Heger, mais elaborada que a triangular, e resultantes ambas das transmutações da herança famosa de Ogden e Richards (54) e dos estudos de Baldinger, bem como de outros representantes da onomasiologia de tradição germânica.



O signema é o signo enquanto unidade da língua. Como é sabido, Pottier designa por *sema*, o traço distintivo semântico desde que seja mínimo, e por *semema*, o conjunto de semas específicos:

Sema e Semema. Lembremos o exemplo familiar, onde S^1 = "para sentar", S^2 = "com pés", S^3 = "uma pessoa", S^4 = "com encosto".

S = semema de *cadeira* = S^1, S^2, S^3, S^4

Estaríamos a ouvir o eco, na lingüística, de noções também triviais no domínio da classificação bi-bhoteconômica. Com o próprio Pottier, reconheceríamos tal similitude. Não afirma ele que o semema equívale às "diferenças específicas"? Donde passaríamos às "características" da taxonomia, ao "gênero próximo" e à teoria tradicional da "divisão" com suas raízes platônico-aristotélicas, para não

falar do seu parentesco com a definição.

Esta apresentará, como um semema, cada um dos conceitos, *verbi gratia*, os de "casa" e "cabana", com que Baldinger exemplifica a operação intelectual em apreço:

$$S_1 \text{ (casa)} = (S^1 + S^2 + S^3)$$

1 2 3

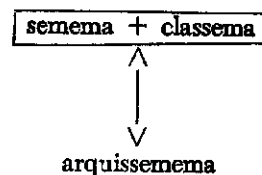
$$S_2 \text{ (cabana)} = (S^1 + S^2 + S^3)$$

1 2 3

Atente-se, como prova final, para o seguinte esquema de Pottier (5): "Todo *lexema* (significante mínimo de designação) terá como significado:

definição
independente
ou *semantema*

definição
relativa



Resta acrescentar que o *classema* vem a ser o conjunto de classes conceptuais gerais e o *arquissemema* o conjunto de semas comuns a vários sememas:
 $S^1 \cap S^2 \dots \cap S^n = \Sigma$.

Ou, na curiosa ilustração que lhe dá Pottier: "2 caniches + 3 "épagneuls" = 5 cães + 3 gatos = 8 animais; 8 animais + 2 homens = 10 seres".
Dá-se a *sinonímia* quando o "essencial de vários sememas, pertencentes a outros tantos significados, é constituído pelo mesmo *noema*", isto é, "sema independente de uma língua dada, considerado como universal semântico", no dizer de Heger.

Eis uma sinonímia parcial (inclusão):

- barco S¹, S², S³
- navio S¹, S², S³, S⁴

Se houver um ou vários semas comuns a diferentes sememas parcialmente análogos, ocorre a *polissemia*, como, no exemplo também do lingüista francês:

- couverture (coberta) S²⁰, S²¹, S²²
- couverture (capa) S²⁰, S²⁰, S²¹

Na *homonímia*, enfim, os sememas são de todo independentes, como os de língua, (órgão) e língua, (termo da lingüística).

Num esquema próximo ao que, com tal fim, adota Baldinger, propõe Soergel (96) um exemplo que torna evidentes as relações típicas dos sinônimos e dos quase-sinônimos.

Denominação	Conceito	Denominação preferêncial (normalizada)
Verstaatlichung (estatização)		Verstaatlichung, Sozialisierung
Nationalisierung (nacionalização)		Sozialisierung 1 (economia)
Sozialisierung (socialização)		Sozialisierung 2 (psicologia social)

equi-valentes {
 sinónimos {
 homónimos {

5.6 — *Polissemia, homonímia, sinonímia, definição nas LD*

Vejamos, pois, no universo dos documentos e dos textos o reflexo dessas equivalências, reduções a elementos mínimos e relações entre termos e/ou significados, ou grupos de tais unidades.

Uma vez que, na *definição nominal*, se fixa a compreensão da palavra em função da de outras palavras, é força indagar se fugindo, é óbvio, ao círculo "vicioso" clássico — não seria possível reduzir-lhes o número. Porque não reconstruir sempre um cabeçalho — recurso típico na documentação — pela coordenação, ou melhor, pós-coordenação dos termos que lhe integram o significado, em vez de usá-lo como unidade no léxico? O mesmo valeria *mutatis mutandis* para os códigos documentários. Segundo R. Pages há de fato, uma linguagem natural de termos de definição. Nem julgam diferentemente os originadores do *General Inquirer* (75) ao buscarem identificar elementos constantes "building blocks" — de definições, com vistas a exprimir um conceito como o produto lógico dos "primitivos" na definição em causa.

De que forma, quanto a esta, se traduziriam tais noções? É que em virtude de sua estrutura supracitada, do tipo $P(x)$, $P(x, y)$, $P(x, y, z)$ ou $P[P(x, y, z)]$, ou, ainda, sob outra forma, do tipo $xR_1 yR_2 z$, a fragmentação ou partição dos cabeçalhos daria múltiplos subcabeçalhos. Estes, — Pagès frisa-o, — embora menos freqüentes que aqueles, que os combinam diversamente, poderiam corresponder efetivamente às questões dos usuários se as fizermos variar sistematicamente, "modulando-as" como diria Gardin (26).

R. Moss (72) lembra a asserção de B. Russell: "I call a vocabulary a "minimum" one if it contains no word which is capable of a verbal definition in terms of other words of the vocabulary". No que se refere à fixação da linguagem "definicionai", como a qualifica o pesquisador francês, é força levar-se em conta a LN "veicular", que prove as "palavras-ferramenta" e as outras, além do repertório de "geradores", procedentes não só do idioma científico geral, mas também do dialecto específico de uma disciplina.

Para determinar esse componente do sistema de informação, o requisito é a análise sócio-psicológica, lógica, operacional, mas, sobretudo, lingüística do comportamento típico do processo do documento.

Note-se que, nas LD, a definição traduz-se implicitamente por essa opção fundamental dos "termos", "descritores", "classes" canônicos, a que as lista dos "Thesauri", esquemas classificatórios, ou meios de busca afins, apõem, se necessário, para definir-lhes a denotação, as "notas de escopo" (scope notes, définitions), como por exemplo: *Revista (documento) e Revista (espetáculo)*, ou, melhor

ainda, nos "Subject Headings" da Library of Congress, para *Simplification in Industry*: "here are entered works on the elimination of excess varieties of products and procedures by a manufacturer..." Assim também, o "non-descriptor" (n. d.): "*Wing-fuselage Interference Effects* (n. d.)" vem definido, no sistema Zator, pela "nota de escopo": "use "Wings plus "Fuselage plus "*Interference*", com * a indicar os "descriptors" de C. Mooers (17), que cunhou o termo hoje tão generalizado para tormento seu. Não temos aí algo como, na LN, a definição "introdutória": **GN, isto é GN_z, GN_x ou GN_y, GN, é GN?**

Tratamos as definições de forma que se cifrasse nelas a referência à paráfrase de Coyaud (24) e Mattoso Gamara. Imposição dos limites que fixamos, eis a razão disso, e não que seja raro o uso da paráfrase, para nada dizer da braquilogia e da abreviatura com fins documentários.

Ao contrário, nos resumos, na condensação, na indexação há transformações parafrásticas a gerarem sucedâneos do texto e a resumi-los. Como aponta Coyaud (24), "no limite, o número de classificação é a definição de um livro, radicalmente abreviado". O estudo dos critérios de equivalência, das interrelações da paráfrase com a sinonímia e a polissemia interessaria, pois, não só a semântica como tal, se não também a documentação.

Não caberia dizer dela o que disse Mounin (5) da definição, a saber, que ela é "la vraie voie de passage entre sémantique et logique"?

Na medida mesma em *que* se afirma o caráter artificial da LD — em que ela se dá como construída e se aparta da natural, — a *sinonímia* é dela excluída. Não, porém, de todo, quando a forma espontânea da comunicação lhe serve de substrato. Demais, se admitirmos que na definição de sinonímia se abarcam a equivalência, o paralelismo ou correspondência ao nível das locuções e frases e das combinações de elementos significativos nem sempre se pode considerá-la descartada *a priori*. É o que se reconhecerá a propósito dos sinônimos extensionais, a cerca dos quais Baldinger lembra o dito de Pascal: "il y a des lieux où il faut appeler Paris, *Paris*, et d'autres où on l'appellerait *capitale de la France*". Algo similar ocorreria no âmbito da LD.

Assim Grolier (49) assinala à saciedade, no tocante a esquemas como a CDU e outros, a confusão possível entre "índices diferentes de sentido muito próximo", como entre *psicologia sexual* (159. 9221) ou *da criança* (159. 922. 7) e a expressão dos mesmos conceitos pela adição, à Psicologia, da subdivisão "-055 segundo o sexo", ou "-053 segundo a idade".

De modo geral, o expediente clássico para a redução da sinonímia ou "quase-sinonímia" é a remoção da forma eliminável à canônica. Para citar um exemplo de van Dijk e van Slype, teríamos o

clássico "ver" ou "use" e UP (usado para) em

Revista *ver* Documento
Documento UP Revista

As relações entre a sinonímia nominal e a definição são imediatamente perceptíveis, o que corrobora a asserção de Coyaud (23) de que o seu estudo nas LD é de grande proveito para o "sémanticien". Mostra-lhe em que contexto, tais ou tais sinônimos são substituíveis ou não, com mira na classificação que lhes seja adequada e para evidenciar a lei de atração sinonímica de Ullmann (23).

No quadro de análise semântica dos conteúdos documentários é que a *polissemia*, o seu reconhecimento e resolução, se revestem de importância, sobretudo no curso do processamento automático dos textos. Qual seria, se não essa, a razão de frequência com que, por exemplo, a expressão "ambiguities" surge na literatura sobre a matéria? Como observa N. Bely é operatória a definição que propõe, ao qualificar de polissêmico "todo termo ou grupo de termos da LN cujo equivalente na LD é variável". Tanto as palavras quanto os grupos de palavras daquela podem corresponder quer a um descritor simples, quer a um grupo de descritores desta. Acentua o autor o caráter condicional de sua tradução, isto é, o fato de depender de fatores incorporados a regras em conexão com os cinco tipos de polissemia que enumera.

Assim "frequência" seria do tipo 1, isto é, palavras-chaves unívocas, mas polivalentes; "complexo", do tipo 2, categorias gramaticais múltiplas, no caso substantivo /adjetivo; "hipocampo", do tipo 3, termo preciso em mais de um domínio; "fibras", do tipo 4, vários sentidos numa mesma área; "centro", do tipo 5, descritores obrigatórios. Diferentes algoritmos se aplicam a cada caso: análise sintática apenas, análise semântica e sintática, análise das palavras vizinhas, do contexto de dependência, e assim por diante.

Coyaud (24) dá o exemplo de uma regra contextual no campo da oftalmologia. Se o termo "chamber" vier precedido de "in vitro", "slored in", "moist" não há retê-lo; nos outros casos "chamber" é parte anatômica do olho, e traduzível pela forma normalizada.

Ao nível da lexicografia, por outro lado, a polissemia virtual é obviada, não raro, por recurso qualificado de "sintático" por van Dijk e van Slype e destinado a evitar as chamadas "falsas combinações". Destas últimas, os citados autores dão como ilustração, o apresentar-se um documento sobre "organização da contabilidade" como resposta a pedido sobre "contabilidade dos trabalhos de organização".

Se adotarmos a técnica dos "indicadores de funções", no caso os "roles" da EJC, com 8 para "as-

sunto principal". 10 para "meios utilizados para realizar o assunto", a indexação do documento em apreço daria "Organização 10" e "Contabilidade 8". Nenhuma confusão, pois, com "Organização 8 e Contabilidade 10", correspondente ao tópico procurado.

Discriminam-se, desde modo, os vários matizes dos termos. O preço é a multiplicação real dos descritores.

6 - ONDE AS ESTRUTURAS INTELLECTUAIS HABITAM ENTRE AS LINGÜÍSTICAS

Por vários trajetos, as interrogações que nos moveram à presente análise, convergiram para um núcleo de *relações* subjacentes aos textos e que lhes servem de arcabouço. Aí viemos ter reiteradamente, ao sondar determinadas estruturas comuns à linguagem e aos sistemas graças aos quais realizam o seu destino os documentos mantidos nessa "conserve culturelle", a que se refere A. Moles (22), como a uma das etapas do processo de comunicação.

Não se insinuam elas na organização paradigmática ou sintagmática, não se lhes deve a trama central na articulação dos vários níveis de elementos lingüísticos, não conduz a análise semântica à busca de um "subconjunto da enciclopédia que lhe sirva de metalinguagem" onde prover-se das categorias mais gerais?

Ocorrer naturalmente interrogar se, ao tomá-las por tema de estudo, se atinge um plano extralingüístico. É um risco a correr-se, de qualquer modo, uma vez que o caminho da linguagem é um caminho de crista, a referir-se sempre a objetos de que se alheia ao significá-los — ... "uma rosa é uma rosa é uma rosa" ... e a fundar-se sempre, por outro lado, em significados e estruturas profundas, algo em que não quer perder-se — ... as palavras para dizê-lo "deixam de ser um problema..."

6.1 — *Relações lógicas e conjunções nas LN*

Qual, pois, a natureza de tais relações? Grolier (5. 1) prefere a definição que lhes deu Lévy, ao identificá-las como conectivos lógicos e ao ver nelas os elementos da significação capazes de vincular todos os termos de indexação constantes do vocabulário, ressalvadas as limitações devidas ao sentido dos termos concatenados.

No âmbito das LN, elas constituem — já o dissemos — um como tecido conjuntivo responsável pelas formas combinadas que se estendem do sintagma lexical até o super-oracional, o parágrafo, o capítulo, toda esta cadeia infinda de símbolos e informações a que monta o grande texto do saber humano.

Que a sua função tem valor nuclear pressentira-o esse poeta da intelecção meridiana que foi Valéry.

"represents"; 3, of constituency — "constitutes"... 6, of logical basis — "is based on"; 7, of source — "comes from /stems from/"... "Destarte, a sentença "our own studies were started as part of a plan", mediante a atribuição do tipo 7, geraria "our own studies stemmed from a plan". Mostra-se aí, na redução à forma canônica, operação de largas perspectivas para a documentação automática. Não caberia, é claro, registrar todas as tentativas feitas com objetivo idêntico. No entanto, se o inventariar os tipos parece válido, é lícito perguntar se a busca dos universais também não o é? "Toda observação sobre qualquer linguagem é, pelo menos parcialmente, um comentário sobre a natureza da linguagem como tal", eis como Postal exprime o princípio fundamental nessa reformulação, hoje em voga, da gramática dos universais. Chomsky não teria também avançado nesse sentido com a reavaliação da de Port-Royal?

A tal gênero de investigações, como é notório, prende-se o nome de Greenberg (24) que, em 1963, propunha 45 universais, como este, reproduzido por Coyaud (24): "Nas frases declarativas com sujeito e objeto nominal, a ordem dominante é quase sempre tal que o sujeito precede o objeto. Os tipos comuns são, pois, VSO, SVO, SOV. "

Quase arriscaríamos, de nossa parte, compulsando o próprio Greenberg, a afirmar a existência de "universais de universais". Que outra coisa implicaria a sua asserção pela qual a tipologia se vincula aos universais: "Sempre que a uma ou mais classes definíveis numa tipologia faltar exemplificação empírica, tal fato é logicamente equivalente a uma generalização sobre todas as línguas"?

Mas há um tipo de universalidade cujas correlações com a ordem lingüística devem ocupar "logicamente" o primeiro plano. E dizemos "logicamente" de modo deliberado, porque é de lógica, ou melhor, de suas interrelações com a semântica que se trata.

Reichenbach figura entre os pioneiros nesse terreno, *por* ele desbravado o bastante para declarar que "a lógica é a *análise da linguagem*", e que o termo "leis lógicas" deveria ser substituído pelo termo "regras da linguagem".

Obviamente o domínio mesmo dessa indagação vem a constituí-lo a semântica, que visa a dar conta do significado dos elementos léxicos, bem assim de como eles se interrelacionam numa "sentença S da língua L", e isso - acima o indicamos - a partir das relações de estrutura profunda, via de regra não presentes nas estruturas sintáticas de superfície.

6.3 — Componentes semânticos de Bierwisch

Quanto à análise semântica — adverte-o M. Bierwisch — os pressupostos são: postulados de significação, ou regras semânticas, introduzidas for-

malmente por Carnap (10); componentes semânticos das significações léxicas.

Digno de nota é o incluírem-se, tanto nuns quanto nos outros, constantes lógicas como "e", "ou", "não" etc. Assim, a forma dos "postulados" para as regras "rapaz — sexo masculino" (boy implies male), "moça — sexo feminino", e assim por diante, viria a ser: "homem - sexo masculino *e* adulto", "mulher — sexo feminino *e* adulto", "rapaz *ou* moça — não adulto", "moça — não rapaz", "homem *ou* mulher, rapaz *ou* moça — (ser) humano".

A definição do significado em termos dos seus componentes semânticos — exemplificar-se-ia como segue: "rapaz: animado e humano *e* do sexo masculino *e não* adulto"; "moça: animado e humano e do sexo feminino *e não* adulto"; "homem: animado *e* do sexo masculino *e* adulto".

O significado de uma palavra é pois "um complexo de componentes semânticos — ou traços, ou indicadores — postos em conexão por constantes lógicas".

Da mesma forma, Bierwisch lança mão de *componentes relacionais* atinentes a verbos transitivos como "hit, meet, love", a verbos com dois objetos "give, take" ou nomes "relacionais", como "father, mother, brother". As entradas lexicais para este último seriam: "brother: **X** child of parent of **Y** *and* male **X**". Em "give" o componente simples "have", isto é, "**X** have **Y**", entra a seu turno numa combinação de componentes semânticos, passando "**X** cause P" a "**Z** cause (**X** have **Y**)", isto é "**Z** é quem produz o estado expresso por P".

Assim a sentença — ainda uma vez o verificamos — esteia a sua representação na estrutura sintática de superfície, onde se combinam os significados lexicais da estrutura profunda; segundo as relações sintáticas em jogo ao que observa Bierwisch, Note-se com ele, pois, que as regras lógicas de transformação e dedução se aplicam às representações semânticas, cujos princípios são universais e são os de lógica formal.

Donde, formulações, como a seguinte, para "o rapaz mata o cão", onde se apõem os mesmos índices àqueles componentes referenciais que se reportam ao mesmo objeto:

- a) [[o rapaz] NP, [mata [o cão] NP₂] VP] S
- b) Humano **X**, *e* sexo masculino *X*-*i* *e não* adulto **X**_i *e* **X**] causa (**X**, muda para (não vivo **X**_i)) *e* animado **X**, *e* cão **X**.

6.4 — Lingüística e lógica formal

Justamente a vinculação da "estrutura profunda" com a lógica formal — equivalente, esta, a "uma linguagem artificial explicitamente concebida e não raro com fim especial" — perseruta-a, com finura a penetração, J. Fodor.

Após mostrar "semelhanças formais" - as derivações sintáticas têm feição de provas lógicas -

destaca a autora, por mais importante, o fato de conterem tanto as descrições lingüísticas quanto os sistemas lógicos, determinadas regras com que definir um conjunto de cadeias "bem-formadas" de símbolos nos seus respectivos vocabulários, e outras com que assegurar-lhes a interpretação.

À pergunta, que é nossa aqui também, sobre o haver ou não, na gramática, uma contraparte para as regras de inferência lógica, responde a autora com Strawson (40), para quem explicar a significância, de uma sentença é dizer quais as condições em que alguém que enuncia faz uma proposição verdadeira, bem assim, mostrar as conseqüências (entailments) da sentença. Como estas podem ser infinitas não há enumerá-las, mas sim dotar o componente semântico de regras de inferência ou fazer com que sejam geradas de alguma formulação da sentença.

Do ponto de vista do lógico há, outrossim, o interesse de confiar ao lingüista a tarefa de prover meios para ligar as suas estruturas coerentes com as estruturas de superfície, aparentemente incoerentes, da linguagem. O que J. Fodor demonstra não ser possível, *verbi gratia*, em certos casos de especificação negativa de quantificadores lógicos a serem introduzidos no formato das regras gramaticais.

Curioso é que se supusermos, nessa troca de métodos, que, da lógica, só procederá a unidade de critérios, encontraremos na própria autora um desmentido, ei-la que afirma que pode o lingüista embarçar-se quanto à seleção dos vários sistemas alternativos que se lhe põem à disposição. Como decidir se dois sistemas lógicos diferem significativamente ou se o fazem por uma questão de notação?

Sem nada dizer do problema crucial de fundarem-se as representações semânticas numa opção entre sistemas igualmente lógicos que reduziriam os operadores uns aos outros, como o *Ou* ao *E* e ao *Não*, ou *E* ao *Ou* e ao *Não*. A tal ponto que se fala, ao que nota Blanche (11), em sistema AN (disjunção e negação) para indicar o de Whitehead-Russell (11), em sistema CN (conjunção e negação) para o de Lukasiewicz (11), e assim por diante.

Já antes constatamos que na mesma LN a paráfrase das conjunções faculta a equivalência de construções como "Se não me engano ou você já veio" ($\bar{p} \supset q$) e "ou eu me engano ou você já veio" ($p \vee q$).

6.5 — Estruturas intelectuais: tríades de Blanche

A elaboração de um esquema unificador, com raízes nas operações essenciais ao pensamento, eis um imperativo se não categórico, pelo menos consolidador das categorias. O mesmo Blanche (12), em

especial, debruçou-se sobre o problema das estruturas intelectuais, no sentido de identificar-lhes o que poderíamos chamar o arquétipo universal. Julga o lógico francês que a conjunção e a negação — ao nível da LN, o *e* e o *não* — se arraigam em estratos subjacentes ao próprio plano intelectual e traduzem funções da consciência em geral: "a consciência poder de adição e de composição, poder de aceitação e recusa". Assim, o princípio de contradição envolve ambos os conceitos: não se pode admitir, ao mesmo tempo, uma proposição e a sua negação: $\sim (p. \sim p.)$. Onde há agrupamento, há conjunção, até mesmo entre os termos que a disjunção relaciona.

Quanto à negação, lembra Blanché (12) o que disse Coirault das díades e tríades. "Antítese (ou oposição), o que faz dois; meio e extremos, o que faz três; tais são os princípios e os ritmos ou se quisermos, os fermentos das imagens construtivas de nosso pensamento verbal..."

Blanché (12) cita, igualmente, Jespersen, cujos exemplos, que constam de "The philosophy of grammar", lhe parecem esclarecedores: "branco, não branco; possível, impossível", ou seja, pares de *contraditórios* que formam alternativa; "branco, preto; doce, amargo", pares de *contrários* que admitem termo médio; "demonstrado, indeciso, refutado; anterior, simultâneo, posterior", *tríades* que comportam um meio entre extremos.

Mas os conceitos se relacionam aos conceitos, como as palavras às palavras, num sistema global, e para tornar explícitas essas estruturas de parentesco, segundo as quais eles se ordenam — já que nos valem de escalas, espectros, árvores, chaves, estrelas, como de esquemas intuitivos de natureza geométrica ou topológica — constrói o autor, a seu turno, uma "generalização esquematizante".

Parte do quadrado das proposições opostas, da lógica tradicional, para considerá-lo como forma frusta de um hexágono, onde cada tríade deriva de inserção de um novo "posto" na díade anterior que substitui.

Não nos cabe aqui esquadrihar "essa estrutura formal do pensamento em geral", extremamente rica. Basta lembrar, com Blanché, que "ela põe em jogo duas ou três operações intelectuais de todo primitivas e elementares", deixando-se analisar ou como "uma dupla de tríade ou como uma tripla díade". Em relação aos "conceitos-cópulas" — igualmente, inerência, inclusão — não ilustra o autor a tríade, mediante um triângulo de contrários $> = < ?$ Uma díade, perguntamos nós, não seria tão só $> < ?$ "A introdução da negação contraditória a partir da tríade, engendra a hexade", ou seja, observa Blanché (12), o sistema de seis símbolos à disposição do matemático:



"Instrumento que permite reconhecer, sob as lacunas e distorções do vocabulário, uma estruturação precisa das idéias" — mais perfeito, a nosso ver, que o "quadrado lógico" de Piaget (12), que o autor francês dissecou, e como retifica, ao projetá-lo sobre o seu esquema — tem o hexágono por domínio o universo dos conceitos, desde as relações fundamentais, a que aludimos, aos "conectores binários", desde as modalidades aos imperativos e valores, para não mencionar os atributos e qualidades.

Demais, respeita-se ao usá-lo, a coexistência da organização dos conceitos por pares contrastados, dicotomias, bipolaridades, com sua organização em escalas graduadas e estruturas lineares, *ambas* na vanguarda da epistemologia e da técnica moderna. Haja vista o que sublinha Blanché (12), o estarem ante os nossos olhos, lado a lado, os computadores digitais e os analógicos.

6.6 — *Relações e conjunções nas LD*

De certa forma essa nervura de relações mais finas e a trama de categorias em que se estruturam os significados interessa duplamente à documentação. Não opera esta sobre e com a linguagem? e ao fazê-lo, não se extrema, por vezes, na criação das LD, a distarem mais ou menos das linguagens naturais?

Os aspectos que tentamos por em foco também são pertinentes às codificações documentárias. Já mencionamos, nesta, a "interdefinibilidade" dos termos e a liberdade de investir no vocabulário, como acentua Pagès - portanto, no paradigma, - "as principais noções em cujo contexto o conceito se torna significativo". E é sobretudo por construção que o fazemos e pela hierarquização.

Às vezes, o serem aptas para tal as LD, pelo que nelas equivaleria à "flexão" ou à "composição", leva até mesmo a incoerências estruturais. Assim, no *código semântico*, como aponta Grolier (49), "isolar" é *SAPT. 003*; "isolado", *SWPT. 003*; "isolante" (particípio), *SU PT. 003*; e "isolante" (substantivo), *SÇPT. LWCT. PUTT. 001*, o que significa "utilizando (Q) a separação (S-PT), agindo sobre (W) a eletricidade (L-CT), produzindo (U) uma proteção (P-TT), e assim por diante.

É claro que se não formos além da condensação documentária e da "codificação telegráfica", as relações deixam, não raro, de ser expressas, por im-

plícitas ou dispensáveis em face do grau de recuperação adotado. Em tal caso, e reduzidos os termos às raízes, o texto "Give algorithms useful for numerical solution of ordinary differential equation" seria, conforme G. Saltan (91), reduzido a "Give Algorithm Use Numer Ordin Differenti Eq...". Se, ao contrário, indagarmos como se exprimem, nas LD propriamente ditas, relações do tipo das traduzidas, nas LN, pelo "e" ou "ou", e por outras conjunções — e isso com maior ou menor precisão Jógica — obteremos para ilustração, símbolos como "+" e "/", da CDU, correspondentes, grosso modo, à operação de reunião. Assim, "622 + 669" significaria "Indústrias minerais e metalúrgicas". No entanto, nela, a regra seria, por assim dizer, a "coordenação assindética", já que se recomenda, no próprio sistema, evitar-se o uso de "+". As entradas separadas tomariam mais rápida a pesquisa.

6.7 — *O ponto de vista relacional em Pagès*

Por outro lado, *as relações* podem ser encaradas antes do ponto de vista paradigmático. Os morfemas (ou monemas, ou semantemas) inclusos na parte estritamente léxica do CODOC apresentam "classificação arborescente" do tipo clássico, embora com ramificações breves. Trata-se, na verdade, de classificação combinatória, segundo Grolier (49).

Na formulação que Pagès propõe, as letras à direita de *a* indicam subdivisões do mesmo tipo que a decimalização, donde $a \supset ab$; a relação $\{\supset\}$ expressa, quer a inclusão estrita, quer outra relação hierárquica com propriedades formais correlatas. Assim, dado a a o valor de "ciência", viria: *ab* método; *ac* problema; *ad* pesquisa... *am* proposições científicas; com subdivisões, a seu turno, para: *ame* conceptualizado; *ami* axioma; *amo* hipótese... De A a Z sucedem-se vinte e duas categorias de noções, de modo que, em Y, figuram relações em geral e em Z, negação.

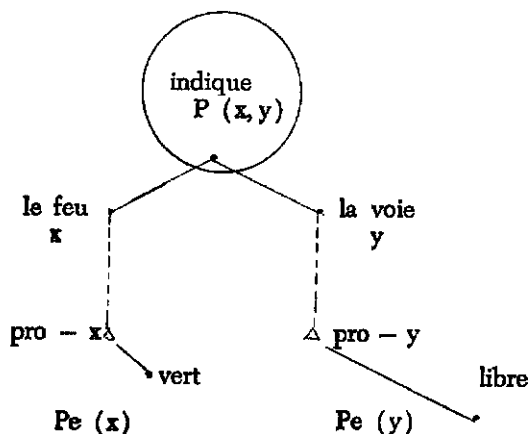
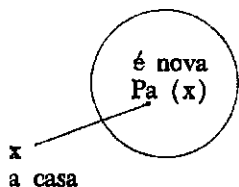
Com semelhante análise do processo de conceptualização, prevê-se o léxico de algo que faculta, mediante a classificação hierárquica, "calibrar", "amplificar" ou "modular" a recuperação, de acordo com a variação das questões, reduzindo-as, por exemplo, de *am* para "*ami*", de "proposição" para "axioma".

As relações sintáticas, em Pagès, abarcam a "união" e a "interseção", da teoria elementar dos conjuntos, bem assim a "tematização" e a "subordinação", assimétricas e com dois aspectos, a saber, "tematizado = tratado por", "tematizante = reportando-se a" e "subordinado/subordinante". E, já que *i* significa "conduta ou processo que se traduz por efeito comportamental no ser vivo" e a notação "tematizado" é, no caso, 5, temos: "ame³ i" para "conceptualização da psicologia".

Equipara-se, pois, a marca predicativa 5 aos "ter-

mos relacionais", que indicam "classes" do discurso afins às preposições ou conjunções da LN. A transformação, pela análise codificada CODOC, de um texto natural em linguagem predicativa — já vimos que o texto compreende argumentos e relações como aR , bR , cR ,... — traduz-se em representações como a dos "estemas" de Tesnière (75). Fornecem elas, no dizer de Pagès, um "como

andaime para a análise lógico-semântica da estrutura das linguagens naturais" e, na LD em apreço, ajudam, particularmente, a segmentar as frases e classificar as palavras em predicados, argumentos, termos lógicos. É o que se infere, quanto às sentenças "a casa é nova" e "le feu vert indique la voie libre" (o sinal verde indica a via livre), dos respectivos diagramas:



Pe predicado do tipo "epíteto"
 Pa predicado do tipo "atributo"
 Pro — i como "pro" em pronome, substituto de...
 (double)

6.8 — *Das palavras-chave às "frases-chave" em Braffort e Leroy.*

Não menos elucidativas se afiguram as observações de Braffort e Leroy concernentes à codificação documentária. Os sistemas de classificação implicam "a possibilidade de estabelecer uma hierarquia das disciplinas segundo a forma de "árvore", no sentido da teoria dos grafos.

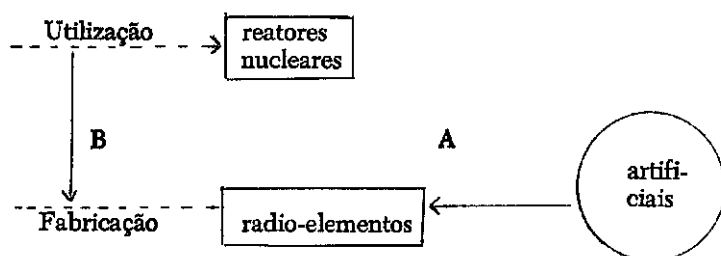
Ora, as interdisciplinas, as atividades científicas "oblíquas", levam à atual transmutação da "árvore de Porfírio" em sistemas de múltiplos entrelaçamentos (boucles) — melhor "diríamos" reticulados. Vejam-se as clássicas observações de Piaget (79) sobre a epistemologia, bem assim a síntese de Queneau para a "Encyclopédie de la Pléiade". Donde o aparecimento de esquemas não só com estruturas algo diversas das classificações tradicionais mas também fundados em palavras-chave, umas e outras a serem denominadas "thesauri" e descritores" em termos documentários mais recentes. "Rubricas" como *Ferro(mina de)*, ferro (nos compostos químicos), dotadas muitas vezes de esclarecimentos, ou seja, das citadas "notas de escopo" com que obviamos as polissemias, oferecem

em tais noções, inclusas entre parênteses, a base do sistema de relações. Apõem-se às palavras-chave esse grupo de elementos puramente relacionais, o que permite se conceba como a LN se amolda à análise documentária. "On a besoin du langage, mais pas de tout le langage. Le problème est donc de trouver une solution optimale", diz Braffort. É de observar-se que tal linguagem se compõe, de fato, de umas tantas sublinguagens relativamente diferentes, em consonância com as subdivisões do pensamento. Na classificação alfa-numérica de Braffort, o conjunto de índices (cotes) associado a cada "rubrica" representa o vocabulário, consistindo a sintaxe nas relações algébricas a que satisfazem as "cotes". A estrutura é bidimensional, com cinco colunas intituladas "Fenômenos inter-escalares, comunicações", "Escala corpuscular", "nuclear", "atômica e molecular" e "macroscópica", e cinco linhas, atinentes respectivamente a "Problemas teóricos", "Produção de fenômenos", "Medida de fenômenos", "Descrição de fenômenos" e "Utilização de fenômenos". A denotaria "Problemas teóricos inter-escalares"; B, "corpusculares"; F, "Produção de fenômenos inter-escalares", e assim por diante, omitindo-se o *i*.

Subentende o sistema "propriedades sintáticas do tipo inflexivo (declinação, conjugação)". Assim, "propriedades químicas dos elementos" traduzir-se-ia por $TABx$, em que AB é o símbolo clássico do elemento e x um sufixo numérico atinente à prepa-

ração do corpo, seus óxidos etc. Note-se que quando o vocabulário e a sintaxe são ricos — caso do domínio técnico-científico atual — urge fixá-los acuradamente. Mais conveniente seria, então, associá-los

ao diagrama, que para o artigo *Tutilisation des réacteurs nucléaires pour les productions des radio-éléments artificiels* poderia ter a disposição:



retângulos:	objetos
círculos:	propriedades
flechas tracejadas:	ações
flechas cheias:	relações
— do tipo A	“relaciona-se
— do tipo B	a”
etc.	“tem por fim”

Inventariar a imensa família de classificações e de relações que as estruturam, enquanto organizações paradigmáticas, ou enquanto subjacentes às articulações do discurso, no eixo sintagmático, escaparia, por muitos ângulos, ao nosso propósito, que é antes a interseção do plano da linguagem com o da documentação do que o levantamento exaustivo de ambos.

Não há, porém, como omitir duas ou três ilustrações e comentários finais a esse respeito, evitada, embora, qualquer insistência quanto a obras de fato fundamentais para a taxonomia bibliográfica, quais sejam as de Sayers, Bliss, Ranganathan, Mills e os trabalhos que resultaram do "Classification Research Group (CRG)". Elas, sob muitos aspectos, poderiam ter sido tanto quanto as citadas anteriormente, o ponto de partida de nossa indagação.

6.9 — Modelo de Gardin para os léxicos documentários

Entre as últimas, afigura-se-nos de primordial relevância para o estudo da lexicografia "documentária", em sua oposição à natural, o modelo proposto por Gardin, a que já nos reportamos. Os conjuntos de termos naturais vêm a ser agrupados em ordem não-significativa, como a alfabética, e teremos glossários, vocabulários especializados e similares; ou podem ser organizados graças a afinidades semântica, e teremos dicionários de sinônimos, de idéias afins, *thesauri*, entre outras formas. O mesmo ocorre com os "léxicos documentários", em

referência aos quais — assinala-se o fato na tipologia em questão — há duas correntes, uma tradicional, em que não se torne explícita a correspondência entre os termos da *LN* e da *LD*, e outra que visa a inventariá-las.

Por outro lado, é clássica a distinção, endossada por Gardin, entre classificações *unidimensionais* e *pluridimensionais*, conforme se estruturarem numa só, ou em mais do que uma dimensão. Por dimensões de uma organização entende-se a natureza das relações analíticas, paradigmáticas, que a constituem, ou, em outras palavras, da relação dos termos à classe a que pertencem.

Um passo adiante nessa caracterização, e cabe enumerar, entre as unidimensionais, as *reais*, sistemas taxonômicos no estilo dos das ciências naturais, onde, de fato, uma só relação informa todos os níveis, e as *aparentes*, em que a multiplicidade se oculta sob a unidade. As multidimensionais abarcam, além das organizações *semânticas*, que refletem definições correntes e, em última análise, a ordem natural, as "*sintáticas*", fundadas não na essência, mas na função das entidades num dado campo de observação — tais são as *classificações de facetas*. Há, por fim, classificações *mixtas*, em que alternam os pontos de vista essenciais e funcionais, e, entre elas, os esquemas dito analítico-sintéticos, tais como o concebe Gardin.

No quadro à esquerda, ao qual convertemos o diagrama de Gardin, tornam-se patentes algumas dessas relações entre os "léxicos":

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

	inorganizados	organizados					
			listas α			Relações analíticas nas LD classificação	
LN		glossários		} THE- SAU- RI	Tipo	Forma	Conteúdo
			Thesauri LN		A	1 relação / real/	Unidimensional
	correspondências LN-LD implícitas		Thesauri LD	B	1 relação aparente	Pluridimensional	
LD	correspondências LN-LD explícitas	positivas: Uniterm negativas: KWIC (contra-dicionário)	classificações	C	várias relações	Pluridimensional	
		Vocabulários					

Quanto às letras do quadro à direita, adaptado de Coyaud (24) para exemplificar os tipos de "relações analíticas" (paradigmáticas) nas LD, há que acrescentar os comentários respectivos. A concerne à inclusão pura "que só ocorre em certas classificações técnico-científicas, como as anatômicas. Em **B**, caso geral, é típico o que se dá com sistema de Foskett (24), por exemplo: o morfema **Dnb** "crâneo" se inclui em "doenças industriais" quando é de fato tão só uma localização delas. Nessa categoria teremos a CDU, os "grupos da WRU, as categorias do MEDLARS, as hierarquias do SYNTOL. Em **C** há tantas relações quantas dimensões, como se dá nas "palavras" da WRU, na "rede de noções" do SYNTOL, no diagrama geral do GRISAL.

6. 10 — *Relações e categorias em sistemas recentes*

Se nos voltarmos, agora, para as categorias e relações fundamentais salta à vista que o estudo clássico de Grolier (49), de que tão amplamente nos valem, não perdeu, com os anos, muito mais do que a atualidade. O próprio autor complementou-o com nova resenha em 1970 (50), onde se descortina, aliás, manifesto retorno à classificação documentária enciclopédica. Assim, o *Descriptive inventory*, de J. K. Tykociner engloba doze "domínios de conhecimentos", agrupados em séries: *Arte*, *Hilenergética* (físico-química, astronomia, ciência da terra), *Exeligmologia* (história, no sentido de Bacon, espécie de ciência da evolução), *Pronoética* (técnicas), domínio da re-

gulação (direito), da disseminação (educação e informação), e *Zetética*, domínio da integração. Aí se acrescentam criatividade artística, filosofia, religião, ao que hoje se vem definindo como a "ciência das ciências", ou como a "cientologia" dos soviéticos.

Tykociner valeu-se da interseção de vários domínios para simbolizar-lhes as subdivisões: A₁ (**B₁B₁**) C₁D₁, "raios cósmicos", onde A₁ (**B₁B₁**) é astrofísica, e C₁ física da matéria inter-estelar. "L'idée est bonne, la réalisation l'est moins", observa Grolier (50).

Respeito à "classificação — biblioteconômica — bibliográfica" soviética — a BRK — de estrutura clássica, e inspirada na CDU e em Bliss, mostra-se-nos também, no referido artigo, que é de 60% o seu grau de "modernismo", isto é, a percentagem destinada às ciências exatas, contra 20% a CDD e na CDU. Os restantes 40% representam as "humanidades". No plano de conjunto, a saber, "A marxismo-leninismo", **B-E, Z-R, SE, Ju** Universo e seu conhecimento", **Ja** Literatura enciclopédica", dicotomiza-se, por vezes, de modo "algo simples", na opinião do autor francês: "URSS de um lado, mundo não soviético de outro; ou, ainda, teoria marxista-leninista/teorias pre-marxistas e não marxistas". Não raro há tricotomias: países socialistas/países em desenvolvimento/países capitalistas.

6. 10.1 — "Operadores" de Farradane e do Precis

Uma vez vista as "categorias" da BRK, não há pas-

sar em silêncio, porque de maior interesse para o problema em foco, as pesquisas do CRG a que já aludimos, um dos resultados dos quais é o PRECIS — *Preserved Context Index System*, a que se vincula o nome de Austin. Nas pegadas de Farradane, optou-se, nele, pela construção dos assuntos a partir dos "quanta" básicos da linguagem, enfoque mais consentâneo com as vistas atuais dos processos mentais" do que a mera subdivisão do universo dos conhecimentos até os conceitos individuais. Há duas categorias destes: *entidades*, isto é, construções mentais (mentefacts) e objetos concretos (artefacts); e *atributos*, isto é, "propriedades, atividades e propriedades de atividades". Quatro relações fundamentais servem de conexão entre os conceitos: *gênero — espécie*, coberta pelas referências cruzadas do tipo "ver também"; *atributiva*, entre entidades e atributos; *possessiva*, entre objetos e partes, e *interativa*, entre objetos. Nesse quadro, para unificar a estrutura gramatical das "sentenças do índice", análogas aos "analets" de Farradane (41), e manter o grau de pre-coordenação que o nome PRECIS sugere, fixaram-se vinte e dois "operadores", a serem aplicados de tal forma que os elementos, num assunto composto, obedecem à estrutura $A > B > C > D$.

Vale isso dizer que se considera o conceito de *D* no contexto de *C*; o de *C*, no de *B*; e o de *B*, no de *A*, isto é, que "em termos de estrutura gramatical, tendemos a organizar a sentença de modo que se tenha esta na "forma passiva", quando se lê a cadeia da esquerda para a direita", e na "forma ativa", se o fizermos no sentido inverso, permutando-se a posição do verbete inicial (lead) com a do modificador (qualifier).

É força reconhecer-se aí a influência da teoria dos sistemas gerais, bem como a de Farradane na noção de "operadores". Os do PRECIS são: (a) forma; (b) objeto, "target"; (/) relação quase-gênérica, diferença, isto é, forma adjetiva do conceito; (p) subsistema, material; (q) propriedade, percepto "percept"; (o) região do estudo, população — amostragem, "sample population"; (1) ponto de vista, perspectiva, (2) conceito ativo; (3) efeito, ação; (4) sistema-chave, "key system"; (5) disciplina; (6) ambiente; (v) conceito coordenado; (w) conceito coordenado correlato; (x) tema coordenado.

Destarte, para o assunto "Intervention by the British government in the dock strikes of 1973: report of a study conducted at Liverpool", a análise dos conceitos seria: (6) Great Britain; (4) Docks industry; (3) strikes 1973; (3) Intervention — by — in; (2) Government; (o) study region; (/) Liverpool. Note-se o "elo" preposicional criado para ligar os termos "strikes" e "government".

Na esteira das reflexões de Farradane, o PRECIS, como aliás, em geral, as pesquisas do CRG, também acusam, à saciedade, a ação — acabamos de

vê-lo — não só desse como fermento intelectual que é a "teoria dos sistemas gerais", cuja formulação original devemos a von Bertalanffy (42), se não também da concepção dos níveis de integração ou de emergência (integrative levels).

Considerando-se sistema *lato sensu* tudo que possa ser tido como entidade, sistema-M (major system), aquele que se pretende analisar, e conjunto-D, o conjunto de subsistemas resultantes da sua decomposição, a estrutura do sistema-M seria a totalidade das relações vigentes entre os subsistemas que pertencem a determinado conjunto D. As *relações* surgem, então, como a totalidade das restrições que incidem sobre as possíveis combinações de dois ou mais sistemas.

São essas algumas das definições propostas por M. Toda, bem assim a de composições e decomposições, materiais ou conceptuais, a de "árvores e famílias de decomposição", todo um instrumental teórico utilizável na "análise e síntese da informação".

Num assunto *verbi gratia*, como "a influência dos lares desfeitos sobre a incidência da delinquência juvenil", "delinquência juvenil" é o sistema "passivo", enquanto "lares desfeitos" é o "ativo", conforme Foskett (42) e o que dissemos acima.

Austin (42) introduziu igualmente a noção de "mudança de sistema", de forma que o conjunto de relações, nessa ótica "sistêmica", vem a ser: 1 — Propriedades dos sistemas; 2 — Segundo sistema/ambiente, relacionado a (3); 3 — Atividades e interações sem mudança material; 4 — Subsistema ativo; 5 — Subsistema passivo; 6 — Introdução dentro do sistema; 7 — Segundo sistema/ambiente relacionado a (8); 8 — Interações com mudança do sistema; 9 — Atributo que define "subclasse (entidade mudada)".

De certa forma a teoria dos sistemas reforça a dos níveis de integração. É típico da primeira afirmar, como o faz M. Toda, que só pelo conhecimento das estruturas dos sistemas macroscópicos se pode, de fato, derivar do microscópico o macroscópico. E o que postula o "não-reducionismo" da teoria dos "integrative levels" se não que o todo é maior que a soma das partes? As entidades se ordenam numa seqüência linear tal que "as unidades constituintes de cada nível, mais as relações que as unem, determinam a natureza do "nível superior", na síntese de Foskett (42).

Mais próximos da questão mesma das relações, resta-nos ainda, do lado britânico, lançar rápido olhar aos trabalhos supracitados de Farradane, cujas idéias procedem antes do prisma psicológico e dos estudos de Vinacke, Isaacs, Piaget, e, principalmente, de Guilford (49). Deste, o modelo de estrutura intelectual envolve três planos: tipos básicos de conceitos, mecanismos mentais e produtos do pensamento. Corresponderiam aos conceitos "figurativos" (figurais), simbólicos e semânticos de

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

Guilford as facetas básicas, do sistema de Farradane, "entidades", "abstratos" (abstracts), "atividades", a que ele teria acrescentado "propriedades". As "classes" se reportariam às combinações de "unidades homogêneas" quanto ao tipo de conceitos básicos. "Termos do sistema", ou "termos heterogêneos", eis como designá-los no caso contrário. "Um passo adiante na análise da informação é a dos assuntos complexos por meio de relações explícitas entre os

termos, onde não se aplicam as relações mais fundamentais da classificação (níveis, classes, grupos heterogêneos)", escreve Farradane. Temos então o clássico quadro de "operadores", que a seguir reproduzimos, decorrente da conjunção de dois tipos de mecanismos mentais, o primeiro atinente à natureza "temporal" da relação, o segundo ao grau de distinção (distinctness) em que ela é percebida.

clareza de percepção	Tempo mental		
	Não-temporal	Temporária	Fixa
co-ocorrente	co-ocorrência /0	Comparação /* agent — atividade	Associação /;
não-distinta	equivalência /= (sinonímia)	Dimensional /+ estado	Pertinência /((genitivo)
distinta	distinta de /) substituída por	Ação sobre /- reação	causação /: dependência

A indexação da sentença "Junção de chapas de aço endurecidas", mediante os respectivos "operadores", daria, no exemplo de Foskett (41):

Aço/: Chapas/ — Junção

/;

Endurecimento

10.2 — Sistema Universal de Wählín e relações de Kervégant

Quiséssemos tentar resenha cabal dos estudos sobre estruturas classificatórias e categorias e faríamos, como se percebe de pronto, longa dissertação. Eis porque nada diremos sobre obra tão "séria, técnica e refletida" quanto a de Z. Dobrowolski, voltada antes para a construção mesma dos sistemas de classificação e para o aspecto da notação, que deixamos de lado. Outras há que é força omitir, com as duas ressalvas seguintes, de que trataremos antes de terminarmos essa linha de considerações num de seus pontos altos, o Simpósio sobre fatores relacionais.

Para Wählín, "domínio dos conhecimentos" e "domínio dos termos" diferem quanto às estruturas, de modo que é força distinguirmos *sistemas* de

documentos e de *termos*. Isso porque, de fato, o objeto da classificação não são propriamente os conceitos, por muito que neles se fundem os documentos, mas estes mesmos, de que se deve levar em conta, no entanto, o conteúdo intelectual. Ressurge aí a tese de Hulme (41), da "abonação literária" (literary warrant): o documento é o esteio da taxonomia bibliográfica.

Demais, o ponto de vista da ação no plano concreto requer igual atenção, donde a necessidade de "sistemas práticos", como os que se destinam a fins industriais e comerciais. Daí um "campo universal", a par de supercampos, campos e subcampos. Dois princípios lógicos merecem menção: a classificação mediante sistema de documento com base nas estruturas hierárquicas, genéricas ou outras, e na pré-coordenação (*ab* e *Im*), e a indexação por meio de certo número de termos equivalentes (a, b, l, m).

Para o arranjo dos termos, no eixo vertical, enumera o autor do US (Universal System of Classification) critérios como o alfabético; o semântico, com base nos conceitos; o sistema de definição e a divisão gramatical de acordo com as classes de palavras. A trilogia — coisas, atividade (processos) e pro-

priedades — corresponde *grosso modo* às três classes — substantivos, verbos e adjetivos. Preposições, conjunções — e a declinação — respondem pelo contexto sintático.

Note-se, enfim, que os "indicadores de função" suprem a insuficiência da lógica do discurso.

Da mesma forma, num retrospecto das reflexões significativas sobre as relações e categorias, destaca-se incontestavelmente a análise de D. Kervégant, do sinal ": " na CDU. É notório que, desde 1957, sugeriu o autor se tornassem explícitas as relações entre os assuntos designados por tal símbolo, acrescentando-se-lhes flechas com orientações diversas e números de ordem.

Assim, no tocante à *Pertinência*, "— > 11—>" indicaria inclusão; —> 12—>, partes, órgãos...; em *Processo*, "—> 21—>" equivaleria a ação;

"—> 211—>" ação favorável, estímulo;

"—> 212—>" desfavorável...; em *Dependência*,

"—> 3<—" teria o sentido de causalidade...;

em *Orientação*, usar-se-ia "—> 41—>" para aspecto, caso particular; "—> 42—>", aplicação... e em

Comparação, "— > 51 — >" para semelhança etc.

Um assunto complexo, ilustra-o Kervégant com o título "Influência do clima sobre a resistência da batata às pragas de vírus", assim traduzido na CDU, colocando-se entre colchetes os "sistemas de conceitos" equivalentes a um termo:

551. 8 - > 2 1 - > [631. 521. 6 - > 4 1 - > 632. 38]

(clima) (resistência a vírus)

- > 4 2 - > 633. 491

(batata)

6.10.3 — Levantamento das relações: D. Soergel

Manancial importante para o estudioso das categorias documentárias são, altm da obra de Grolier, contribuições como as de Pagès, Farradane, Lévy, Perreault, apresentadas ao Simpósio Internacional sobre fatores relacionais na classificação. Entre eles, note-se há uma quase completa soma das relações e indicadores de função, em que se trata de "modo enciclopédico", informação contida, em boa parte, nos trabalhos de Grolier, Perreault, Vickery. É a que devemos à análise exaustiva de D. Soergel. Perquirem-se e comparam-se, nela, os principais esquemas existentes no que toca a "indicadores de função", isto é, "conceitos extremamente gerais aplicáveis a uma larga faixa de domínios (fields) do pensamento".

Antes da lista mesma, apresenta o autor um modelo geral para a estrutura do conjunto de assuntos, partindo, para tal, do "conjunto (A + +) de conceitos primitivos (predicados, atributos) e do con-

junto M de "constantes de relevância", destinadas a especificar a importância do assunto mais simples no quadro do mais complexo. A operação "conexão" prevê a construção de uma rede representativa do assunto complexo: "Termômetro", por exemplo, seria expresso por "instrumentos: { finalidade: [(operação: medida): (< objeto: medida >: temperatura)] }".

Quanto às relações entre os assuntos A e B, enumera Soergel três: a relação $A \wedge B$ dos sistemas parcialmente ordenados e equivalente a "A implica B"; a proximidade (nearness) dos assuntos A e B dos sistemas descritos na primeira relação; as relações do tipo "ver também".

Atenta, ainda, o autor, para o fato, já apontado, de se distinguirem das relações *paradigmáticas*, "essenciais e universalmente aceitas — e portanto implícitas", — as *sintagmáticas*, contingentes, ocasionais e explícitas.

O esquema para a classificação dos "roles" ressalvado o que concerne ao tempo e espaço, e aos pares "todo/parte", "classe/elemento", compreende os tipos básicos: coisa, entidade; material, substância; propriedade, condição; processos, agente; objeto (target), além da Causalidade.

6.10.4 — Natureza relativa dos fatores: F. Lévy

Se quiséssemos caracterizar outra opinião de interesse para o nosso tema, a saber, a que F. Lévy externou no Simpósio mencionado, diríamos que vem a marcá-la — e nisso se revela um porta-voz do SYNTOL — um sentido agudo do valor relativo das categorias, já que é mister, ao que mostrou, firmar-lhes a interpretação no contexto semântico, de natureza lingüística e/ou epistemológica, e que não há reduzi-las a fatores estritamente lógicos. Do contexto global, com efeito, dependerá o valor informativo do que significa a palavra "afeta" nos exemplos:

"(O álcool) *afeta* (a delinqüência)" e "(O álcool) afeta (as células corticais)", valor metafórico no primeiro caso, relativo a processo bio-físico no segundo. O fator lógico, por abstrato, oferece fraco teor de informação. Por outro lado, é em si mesma "epistemologia" a lógica que faculte o se tornarem "significativas as noções na medida que reflitam os "comportamento" do campo em questão".

A compatibilidade entre sistemas — "possível *porque* as categorias deverão ser interpretadas em função dos contextos documentários" — sugere-a o seguinte quadro, em que se ilustram "alguns dos denominadores comuns lógicos":

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

		SYNTOL	WRU
	Características inerentes: (inclusão, predicados, atributos)	ATRIBUTIVO	Catagórico Intrínseco Inclusão Compreensão
Estático	Interpretação modal	EXPRESSIVO	Simulação
	Agente, assunto do processo	DO AGENTE AO PROCESSO	Instrumental
	Objetos do processo	DO PROCESSO AO PRODUTO	Produção
Dinâmico	Causalidade, interdependên- cia entre dois fenômenos ou elementos	CONSECUTIVO	Afetado por

6.10.5 — Tríades de Perreault e outros

Se algo, enfim, pode depreender-se dessa breve incursão pela selva das "relações fundamentais", é que ali onde a unidade lógica seria de regra, encontramos, quando muito, um ar de família entre categorias "nem de todo as mesmas, nem todo outras" em cada caso. Lograr-se-ia o geral consenso quanto a uma díade ou tríade final, nos moldes da de Blanché (12) em que, de certa forma, culminou o exame das estruturas nas LN? É o que tentariam insinuar as reflexões de Perreault (77), "Charmain" do já citado Simpósio sobre fatores relacionais, as quais não acusam, aliás, qualquer referência ao lógico francês.

Ao sondar os fundamentos epistemológicos da questão frisa o autor o fato de que o julgamento, constituem-nos não os conceitos isolados, mas a união lógica e a cópula entre eles. "A articulação é o que faculta o discurso. Operadores, sintagmas, morfemas, "correlatores" são diádicos... criam uma cadeia articuladamente discursiva". Se se apõem indicadores de funções (roles) aos substantivos, temos "A1, B2, C1 e D3"; se elos (links), "[A1, B2

(C1, D3)", mas em nenhum dos casos o grau de articulação de "A α B \rightarrow C β D"; aí, diz Perreault, substituem-se "links" e "roles" por símbolos tais que "oc" é a relação copulativa, " β " a relação de certeza moral, " \rightarrow " a "supra-relação" antecedente/ conseqüente.

No que toca a articulação dos substitutos (surrogates) dos documentos, a que aludimos anteriormente, ou seja, dos documentos secundários — simplificações pertinentes do discurso, já que no dizer de Mills (77) "o propósito da indexação é dizer-nos o que ignorar" — o desiderate é a recuperação efetiva mediante estratégia apta a localizar os itens Y que respondem com precisão adequada à questão X. O que se tem em vista são, pois, aspectos sintáticos dessa "estratégia", isto é, no estilo de Perreault, "a manipulação ritual" [entenda-se: operacional?].

Acena enfim o autor para uma sorte de tripartição reconhecível em muitos dos sistemas relacionais. Teríamos o quadro abaixo dos que compara para ilustrar tal similitude, excluindo o único explicitamente epistemológico, o de Farradane (77), onde há a dupla tripartição — quase diríamos a dupla

MANOEL ADOLPHO WANDERLEY

tríade-que já consignamos. A propósito dele lembra Perreault o que disse R. Sokolowski: "... o con-

ceito de constituição pressupõe a historicidade do sujeito e o câmbio da realidade". Eis o quadro:

DE GROLIER (77)	GARDIN (77)	PERREAULT (77)	HJELMSLEV (77)
Objeto em si mesmo	Coordenativo, Predicativo	Subsumptivo	Interdependência
Objeto na relação ativa	Consecutivo	Determinativo	Determinação
Objeto no ambiente	Associativo	Ordinal	Constelação

7 - CONCLUSÕES

Uma vez reconhecida a área dos problemas em foco, a primeira pergunta que ocorre, espontaneamente, é quanto à condição lingüística da documentação, ou, em outros termos, quanto ao existirem, ou não, as chamadas LD, e no caso de as haver, quanto a seu estatuto, por assim dizer, em face das naturais. Consistiria em mera metáfora a sua realidade ou haveria a lastreá-las alguma consistência ontológica?

Parece-nos que, do levantamento feito, se depreende a flexibilidade extrema do conceito. Registramos, no tocante ao "continuun" da indexação, toda uma seqüência de estruturas de léxicos e de sintaxes a constituírem transição gradual das LN para as LD. Embora o elemento conceptual e os significados de umas e de outras não se diferenciem essencialmente, os significantes podem fazê-lo e ir dos grafemas alfanuméricos e usuais aos códigos e diagramas de diversas sortes. Tampouco, na gama das LD, há uniformidade no que concerne a presença; ou não, de sintaxe, ao fato de se relacionarem ou serem discretas as unidades em que o conteúdo se traduz.

Se nos ativermos, pois, ao polo da máxima coincidência, encontraremos, é óbvio, boa parte da problemática das LN, e a "essência" das LD, se a houver, defini-la-emos, por força, a partir da que fixamos para as primeiras. Questão tormentosa, dada a profusão de teorias concorrentes e a ressonância transcendental logo provocada por tal ordem de cogitações.

Bastaria, no entanto, deixarmos os fatos serem os próprios intérpretes da condição que assumem. Estaríamos deante de usos específicos dados aos componentes e comportamento das linguagens. Como assinala F. Mora, "las funciones lingüísticas son muchas y mui variadas, porque son muchos y mui variados los tipos de actos lingüísticos". Retomando idéias wittgensteinianas ou, pelo menos, a respectiva terminologia — e sem abolir a *priori*, contudo, os universais próprios a esse terreno — acrescenta: "en puridad, no hay lenguaje, sino

juegos lingüísticos".

Ora, justamente os há bem nítidos no âmbito da comunicação e do processo da informação ou, mais restritamente, da "conserva" documentária a que aludimos em vários pontos. Bem precisas são, por exemplo, as funções da "recuperação da informação" e as dos sistemas de "question answering", com fortes afinidades a assemelhá-los e uma única teoria de seleção subjacente a ambos, mas discrimináveis ao menos em opinião tão ábalisada quanto a de C. Mooers (6). Não foi ele o criador da primeira, que tanto êxito logrou? O sistema de *recuperação da informação* — afirma Mooers (6) — é o sistema de informação que permite à pessoa que necessita uma espécie de informação previamente descrita (prescribed) selecionar, da coleção de documentos que contém informação, os que quadram com a descrição. O escopo dos "question answering systems", dos de tipo "pergunta-resposta" é prover respostas às questões que lhes são propostas.

Destarte as funções determinantes de tais processos, e de outros que se lhes aliam no campo da comunicação, especializam as das LN. O mesmo se dá, dentro delas, com outros jogos lingüísticos. Assim, "Fogo!" — nota-o F. Mora — tanto vale para grito de alarme quanto, para "si no se es muy cortés... pedir lumbre y encender el cigarillo". Nem outra é a multiplicidade decorrente das estruturas sintáticas, como verificamos a respeito dos níveis profundos e superficiais dos enunciados.

Haveria, perguntamos, algum fundamento para retirar às linguagens documentárias os foros de "jogo lingüístico" ou, por outras palavras, de linguagem, conquanto artificial, em certos casos, e equivalente à transposição ou codificação sui generis do discurso escrito, como este o é do oral? A nosso ver não, e é mister julgá-la tal, ainda que fosse válida a crítica de Bar-Hillel, ao arguir que "se o anzol com que se fiska o peixe não é por força um peixe em miniatura ou condensado, não há razão intrínseca pela qual um índice deva ser um documento em miniatura ou condensado", nem para que o índice, por muito que consista em

frase constante do documento, lhe veicule o teor da informação.

Porque tudo está a indicar, ao contrário, que é verdadeira a objeção que lhe fazem J. Perry (18) e outros pesquisadores da WRU quando mantêm que "a informação recuperada não pode ser se não a informação originariamente armazenada. A equivalência *semântica* deve realizar-se entre o pedido de informação e a informação armazenada". Eis aí outro ponto que merece ser sublinhado em muito de quanto observamos nas páginas precedentes: o papel central que desempenha, tanto na LN propriamente dita, quanto na LD, a noção de equivalência. Todo um campo conceptual de marcada relevância para a nossa investigação, gira em torno dela: sinonímia, polissemia, ambigüidade, homonímia, e aproxima o uso comum das estruturas lingüísticas, do uso documentário.

Por outro lado, suas relações são incontestáveis com certos aspectos da "analiticidade", da definição, da paráfrase e com outros traços das linguagens, entre os quais, a maior ou menor adaptabilidade à tradução. Se o tradutor, ao que observa Olshewsky, se interessa pela equivalência de significado entre expressões de duas línguas diferentes, o analista lógico, entre duas expressões dentro da mesma língua, não é lícito afirmar que a essa segunda categoria pertence primordialmente quem lida na comunicação a fazer resumos, representações indexadas, codificações de documentos, não obstante participar também da primeira no seu afã de transpor as barreiras lingüísticas? Bastaria aduzir, para corroborar tal asserção, o que frisamos repetidamente, a saber, as ligações estreitas de tradução automática com a recuperação da informação. Assim, contribuiriam para elucidar o problema das polissemias nesta última os algoritmos e programas de cálculo mecânico aptos a resolver a ambigüidade, definida por A. Janiotis como a "situação em que uma expressão na língua-fonte comporta mais do que uma na língua-objeto, ou como a situação em que não basta considerar as palavras ou grupos de palavras isoladamente, mas antes suas combinações aptas a uma tradução com sentido". "Les progrès dans le dépistage des informations à l'aide de mots-clé iront de pair avec la solution des problèmes lexicaux de la machine à traduire", ao que prevê E. Delavenay.

A equivalência estaria conferida, na teoria lingüística, função similar à identidade nas estruturas de índole lógica, e isso inclusive do ângulo do significado, da intensão ou conotação. "Para descobrir se dada proposição é logicamente verdadeira — é Bar-Hillel quem o diz — se duas proposições são logicamente equivalentes, ou se duas expressões são sinonímicas, é preciso conhecer tão só a sua intensão (conotação, sentido)".

Ousaríamos adiantar que a outra face, a do valor-verdade, a do referente, a da coextensividade —

que em princípio excluímos de nossas considerações — encontra talvez, como contrapartida na documentação, dimensões e parâmetros do tipo do índice de pertinência, de recuperação, de "atinação", de precisão, a que, incidentalmente, fizemos alusão. Demais ao nível da análise do conteúdo, ao relacionarmos aos significados as estruturas gramaticais vimos aflorar a noção de correspondências formais — e não temos aí, de novo, algo ligado à equivalência? — entre sentenças incluídas na descrição da frase e outras que não o são. Ao que pondera Chomsky — precisamente quem aprofundou a questão — "tais correspondências são de fácil descrição. Não é difícil definir o que se pode chamar de *transformações*, isto é, as operações formais que convertem qualquer sentença de forma declarativa numa forma correspondente passiva, interrogativa ou outra".

Note-se, por fim, o esforço envidado para a formalização, aqui também, do conceito de equivalência lingüística e — *verbi gratia* do teor de sinonímia (synonymity) de dois textos sinônimos, *Te* e *Ti*. Eles o são, para Shreider, "se corresponderem ao mesmo operador de transformação do Thesaurus $A_t = A_{t_i}$ ". A determinação de tal "synonymity" com respeito ao Thesaurus — é "desenvolvimento natural da idéia de equivalência intensional e igualdade de estrutura intensional de "Carnap", para o qual dois textos terão o mesmo significado (intensão) se o mantiverem em qualquer estado concebível do mundo.

Somente, obtempera Katz, "as previsões feitas a partir de uma língua artificial de tipo carnapiano são... completamente vazias... as construções lingüísticas inscritas sob os títulos "analíticas", "sinônimas" etc. são as que a língua carnapiana "prevê" como analíticas, sinônimas etc. O acerto da previsão de que uma frase *P* tem a propriedade *P* nada mais estabelece que o seguinte: esta frase tem a propriedade *P*".

Não estaríamos aí a identificar o mal de que padecem muitos dos formalismos?

Constatamos, outrossim, na seção atinente à análise de conteúdo, a ocorrência de sistemas em que a recuperação se faz, por um lado, mediante a mera justaposição de termos, canônicos ou não, isto é, mediante LD "assintáticas". Eis um modo de linguagem "tal como aparece na criança, ou para o qual tende a linguagem ao degenerar" — observa-o G. Granger — modo, porém, efetivo, até certo grau de experiência documentária.

Por outro lado, afinando-se o reconhecimento do teor informacional, por mais acurada a "questão" do usuário, entram em jogo as sintaxes, as gramáticas e algoritmos respectivos para a resolução dos textos em seus componentes. Situa-se aí o domínio rico das estruturas lingüísticas e lógicas. Estas, com efeito, integram a informação ao articularem a que os termos do léxico veiculam, de tal sorte

que o todo do significado acaba por condicionar o significado dos elementos encarados de per si. Por outras palavras — e consideradas as relações de equivalência a que acabamos de nos reportar — a introdução de um contexto, dentro do qual estruturas profundas respondem pelas superficiais, atesta a interdependência da sintaxe com respeito à semântica, hoje incorporada na lingüística. Sem a remissão constante ao significado correr-se-ia o mesmo risco, a esse respeito, de nos limitarmos à concepção gramatical da linguagem ou das "sintaxes lógicas" a que tende o nominalismo de teses como a que exprime o célebre conceito de Condülac,

da "ciência como língua bem feita". A distinção, aliás, entre função semântica e função sintática — como o vimos ao escutar-lhes as incidências nas LD e como frisa magistralmente G. Granger — a distinção entre remissão, pelos termos, "aos objetos" e às estruturas lingüísticas, é essencialmente relativa e móvel. No quadro a seguir o autor citado elucida a maneira pela qual "dos sistemas sintáticos tematizados como quase-objetos reparte em seguida o processo". É de observar-se que, no caso vertente, o autor considera a escrita em si mesma e não como transcrição de uma língua falada.

	Formas sintáticas	Conteúdo semântico
Escrita	Regras gráficas	Sons
Língua oral	Regras fonéticas	Palavras
Língua	Gramática	Idéias
Língua formalizada	Lógica	Ligações

Nem é por outra razão, como tivemos o ensejo de verificar, que Salton faz depender a identificação do significado, no que diz respeito ao grupo de palavras, do "reconhecimento adequado das ambigüidades sintáticas e semânticas, da interpretação correta dos homógrafos, do reconhecimento das equivalências semânticas, das relações de palavra a palavra e de uma percepção geral do "background" e ambiência de cada enunciado". Como observação final, tudo nos leva a apontar para o que faz as vezes de fecho de abóboda na complexa construção em que nos detivemos: precisamente as ligações centrais, as articulações mais genéricas das linguagens. "Não pensamos com palavras, pensamos com frases" notava Valéry, e determinar as funções da unidade no julgamento, eis o objetivo de um esforço persistente, a que também tende, de alguma forma, a busca que descrevemos ao nível das categorias documentárias. Com efeito, o problema das relações fundamentais já revela, por si só, até que ponto se aproximam as formas lingüísticas e as formas lógicas, como queria Cassirer e, a *fortiori*, a nosso ver, ambas e as LD, muito menos formalizados que aquelas últimas e equiparáveis não raro — é o que tentamos mostrar — a "jogos lingüísticos" ou usos especiais de umas e de outras, isto é, do discurso e da lógica.

Não acusa origem comum, perguntamos, o longo es-

forço para lograr-se a tábua ideal das categorias fundamentais, quer no plano epistemológico e do conhecimento em geral, quanto aos "modos mais universais do ser e os gêneros supremos do discurso", quer no "espelho do mundo", que é o microcosmo documentário, quanto aos fatores relacionais das "linguagens descritoras"? Tão grande se afigura a semelhança que os componentes de toda síntese intelectual dissecada na teoria, e os das representações indexadas, são objeto, por vezes, da mesma caracterização. Destarte, teremos categorias como a "relação" propriamente dita, em Aristóteles e Kant, e a "relação" na CDU; "lugar e tempo", nos Tópicos e na Colon Classification ou na CDU; "causalidade", dentro da relação na Analítica Transcendental, e "causação" em Farradane, e assim por diante. Igualmente em Peirce, (74) as categorias "cinopitagoreanas" de "firstness", mode ser do que é como é sem referência ao outro, de "secondness", do que é com referência a um segundo, mas não a um terceiro, de "thirdness", do que é pondo em relação um segundo e um terceiro, não sugerem a "triáde" do objeto em si mesmo, do objetivo na relação ativa e do objeto no ambiente, de um Groher, por exemplo?.

De outras feitas, a identidade residiria no número de categorias arroladas e talvez, no fundo, na universalidade de certas operações intelectuais que

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

informam estruturas lógicas como a de Blanche. Mas de modo geral — notamo-lo quanto à documentação ao **tratar das idéias** de F. Lévy — é relativamente ao contexto semântico, e no que concerne antes à epistemologia do que à lógica *stricto sensu* que atuará o princípio de repartição em classes, o qual, em termos de Cassirer, "uma vez descoberto, rege não só a organização do sistema nominal, mas estende-se à totalidade do arranjo sintático da língua e torna-se a expressão verdadeira de sua coerência, de sua "articulação espiritual". O que precede nos parece significar que há, de fato, e se afirma de modo crescente, uma função nuclear que se refere, em todas as linguagens — até mesmo em todas as semióticas como dissemos ao início — às estruturas de caráter universal e em especial, às que interessam a lógica e o modo de ser racional da comunicação humana. Tratar-se-ia de capacidade e disposição inatas, como pretende Chomsky, para os universais lingüísticos nos moldes do princípio "A sobre A"? Parece-nos mais avisado o que avança J. Lyons a comentar o pensamento Chomskiano — que não é "um conhecimento dos princípios formais da linguagem que é inato, mas inata é uma "faculdade" mais geral" que, em circunstâncias adequadas, interage com ela para produzir a competência lingüística. Em suma, de acordo com Piaget, as análises estruturais dos lingüistas evidenciam — não há negar — a originalidade de um sistema de sinais face ao de normas de pensamento ou verdades, e suscitam o problema das relações. Mas "estas existem de fato, e os sinais têm por função exprimir significações cuja natureza é lógica em diferentes graus". Eis por que — isso nos traz de volta à citação de S. Elia feita na introdução — o lingüista Hjelmslev veio a fazer a "hipótese de um nível "sublógico", onde se estabelecem as coordenações lógicas e as lingüísticas".

Seria, enfim, de índole racional essa "linguagem de linguagem" que para Michel Deguy é o próprio da função poética do discurso, de todo excluída de nossas cogitações?

8 - CITAÇÕES BIBLIOGRÁFICAS

- 1 ADDRADOS, Francisco R. — Rasgos semânticos, rasgos gramaticales y rasgos sintáticos. *Revista Española de Lingüística*, 2(2): 249-258, 1972.
- 2 ———— Subclases de palabras, campos semânticos y acepciones. *Revista Española de Lingüística*, 1(2): 335-354, 1971.
- 3 ARCAINI, Enrico — *Principes de linguistique appliquée: structure — fonction — transformation*. Paris, Payot, 1972. 302 p.
- 4 AUSTIN, Derek - PRECIS indexing. *The Information Scientist*, 5(3): 95-114, 1971.
- 5 BALDINGER, Kurt - *Teoria semântica; hacia una semântica moderna*. Madrid, Alcalá, 1970. 278 p.
- 6 BAR-HILLEL, Yehoshua — *Language and information; selected essays on their theory and application*. Reading, Mass., Addison-Wesley; Jerusalem, Jerusalem Academic Press, 1964. 388 p. Comentário por: MOOERS, Calvin Comment on the paper by Bar-Hillel. *American Documentation*, 8(2): 114-6, 1957.
- 7 ———— & CARNAP - Semantic information. In: SARACEVIC, Tefko, ed. *Introduction to Information Science*. New York, Bowker, 1970. 751 p.
- 8 BARTHES, Roland - *Le plaisir du texte*. Paris, Seuil, 1973. 105 p.
- 9 BELY, N.; BORILLO, A.; VIRBEL, J. SIOT-DECAUVILLE, N. - *Procédures d'analyse sémantique appliquées à la documentation scientifique*. Paris, Gauthier Villars, 1970. 240 p.
- 10 BIERWISCH, Manfred - Semantics. In: LYONS, John, ed. *New horizons in linguistics*. Harmondsworth, Penguin books, 1971. 367 p.
- 11 BLANCHE, Robert - *Introduction à la logique contemporaine*. Paris, Armand Colin, 1957. 208 p.
- 12 ———— *Structures intellectuelles; essai sur l'organisation systématique des concepts*. Paris, J. Vrin, 1969. 147 p.
- 13 BORDEN, George A. & NELSON, William F. — Toward a viable classification scheme: some theoretical consideration. *American Documentation*, 20(4): 298-301, 1969.
- 14 BORKO, Harold, ed. *Automated language processing*. New York, Wiley, 1967. 386 p.
- 15 BRAFFORT, Paul & LEROY, André - Des mots-clés aux phrases-clés; les progrès du codage et l'automatisation des fonctions documentaires. *Bulletin des Bibliothèques de France*, 4(9): 383-391, 1959.

MANOEL ADOLPHO WANDERLEY

- 16 BRAGA, Gilda Maria — *Relações bibliométricas entre a Frente de Pesquisa (Research Front) e revisões da literatura: estudo aplicado à Ciência da Informação*. Rio de Janeiro, 1972. 39 p. Tese.
- 17 BRENNER, Claude W. - Experience in setting up and using Zatocoding system. In: SHERA J. H. et alii. — *Information system in documentation*. New York, Interscience Publishers Inc., 1957. p. 178-192.
- 18 CENTER FOR DOCUMENTATION AND COMMUNICATION RESEARCH - Comment on "A logician's reaction". *American Documentation*, 8(2): 117-122, 1957.
- 19 CHOMSKY, Noam - Logical structures in language. *American Documentation*, 3(4):284-91, 1957.
- 20 CHRISTENSEN, N. E. - Sobre la naturaleza del significado. Barcelona, Editorial Labor, 1968. 228 p.
- 21 CLOUZOT, O. — La méthode d'analyse sémantique d'un contenu. *Enseignement Programmé*, 6:21-32, 1962.
- 22 LA COMMUNICATION les images, les sons, les signes, théories et techniques, de N. Wiener et C. Shannon a M. McLuhan. Sous la direction d'Abraham Moles. Paris, Centre d'Etude et de Promotion de la Lecture, 1971. 575 p.
- 23 COYAUD, Maurice — *Intmduction à l'étude des langages documentaires*. Paris, C. Klincksieck, 1966. 148 p.
- 24 ————Linguistique et documentation; *les articulations logiques du discours*. Paris, Larousse, 1972. 171 p.
- 25 — & SIOT-DECAUVILLE, Nelly — *L'analyse automatique des documents*. Paris, Mouton, 1967. 147. p.
- 26 CROS, R. -C; GARDIN, J. -C; LEVY, F - *L'automatisation des recherches documentaires; un modèle général "LE SYNTOL"*. Paris, Gauthier-Villars, 1968. 260 p.
- 27 DELAVENAY, Emile — *La machine à traduire*. Paris, Presses Universitaires de France, 1959.
- 28 DIAZ TE JE RA, Alberto - Puntos de contacte entre Sintaxis y Semântica. *Revista Española de Lingüística*, 1(2): 361-370, 1971.
- 29 DOBROWOLSKI, Z. - *Études sur les constructions des systèmes de classification*. Paris, Gauthier-Villars; Warsawa, PWN, 1969, 303 p.
- 30 DOYLE, Lauren B. — Semantic road maps for literature searchers. In: SARACEVIC, Tefko, ed. — *Introduction to Information Science*, New York, Bowker, 1970. p. 223-39.
- 31 DUCROT, O. — Presupposes et sous entendus. *Langue Française*, 4:44-66, 1965.
- 32 EDMUNDSON, H. P. - Mathematical models in linguistics and language processing. In: BORKO, Harold, ed. -*Automated language processing*. New York, Wiley, p. 33-96.
- 33 ELIA, Silvio — *Orientações da Lingüística moderna*. Rio de Janeiro, Acadêmica, 1955. 244 p.
- 34 FAIRTHORNE, Robert A. - Content analysis, specification and control. *Annual Review of Information Science and Technology*, 4:73-109, 1969.
- 35 ————The patterns of retrieval. *American Documentation*, 7(2):65-70, 1956.
- 36 ———— *Toward information retrieval*. London, Butterworths, 1961. 211 p.
- 37 FARRADANE, J. — Concept organization for information retrieval. *Information Storage and Retrieval*, 3(4): 297-314, 1967.
- 38 FERRATER MORA, José - *Indagaciones sobre el lenguaje*. Madrid, Alianza Editorial, 1970. 223 p.
- 39 FID — *Manuel pratique de reproduction documentaire et de sélection*. Paris, Gauthier-Villars, 1964. 314 p.
- 40 FODOR, Janet Dean — Formal linguistics and formal logic. In: LYONS, John, ed. — *New horizons in linguistics*. Harmondsworth, Penguin books, 1971. p. 198-214.
- 41 FOSKETT, A. C. — *The subject approach to information*. 2. ed. London, Clive Bingley, 1971. 429 p.

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

- 42 FOSKETT, D. J. - Classification and indexing in the social sciences. *Aslib Proceedings*, 22(3):90-101, 1970.
- 43 FRANÇOIS, Frederic — Caractères généraux de langage. In: LE LANGAGE. Encyclopédie de la Pléiade. Paris, Gallimard, 1968. p. 20-45.
- 44 GARDIN, Jean-Claude — Éléments d'un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*, 11(5):171-82, 1966.
- 45 GARVIN, Paul L. - Computing natural language and natural language in computing. In: SEMINAR on computational linguistics. Washington, Govt. Printing office, 1968. p. 74-83.
- 46 ———— An informal survey of modern linguistics. *American Documentation*, 11(4):291-8, 1965.
- 47 GRANGER, Gilles-Gaston - *Pensée formelle et sciences de l'homme*. Paris, Aubier - Montaigne, 1967. 226 p.
- 48 GREENBERG, Joseph H. - Language universais. In: SEMINAR on computational linguistics. Washington, Govt. Printing Office, 1968. p. 84-90.
- 49 GROLIER, Eric de - Étude sur les catégories générales applicables aux classifications et codifications documentaires. Paris, UNESCO, 1962. p. 262.
- 50 ———— Quelques travaux récents en matière de classification encyclopédiques. *Bulletin des Bibliothèques de France*, 15(3):99-126, 1970.
- 51 ———— Synoptic critique. *Information Storage and Retrieval*, 3(4):385-57, 1967.
- 52 GROSS, Maurice - On the equivalence of models of language use in the fields of mechanical translation and information retrieval. In: SARACEVIC, Tefko, ed. *Introduction to information science*. New York, Bowker, 1970. p. 210-8.
- 53 GUIRAUD, Pierre - Langage et théorie de la communication. In: LE LANGAGE Encyclopédie de la Pléiade. Paris, Gallimard, 1968. p. 145-67.
- 54 HEGER, Klaus - L'analyse sémantique du signe linguistique. *Langue Française*, 4:44-66, 1969.
- 55 KATZ, Jerrold J. — *La philosophie du langage*. Paris, Payot, 1971. 269 p.
- 56 KELLOG, Charles H. — Designing artificial languages for information storage and retrieval. In: BORKO, Harold, ed. *Automated language processing*. New York, Wiley, 1967. p. 325-67.
- 57 KERVEGANT, D - Classification et analyse des relations. *Bulletin des Bibliothèques de France*, 4(11):495-511, 1959.
- 58 KUNO, Susumu — Automatic syntactic analysis. In: SEMINAR on computational linguistics. Washington, Govt. Printing Office, 1968, p. 19-41.
- 59 LANDRY, Bertrand C. & RUSH, James E. — Toward a theory of indexing. II. *American Documentation*, 21(5):358-67, 1970.
- 60 LANGACKER, Ronald W. — *A linguagem e sua estrutura; alguns conceitos lingüísticos fundamentais*. Petrópolis, Vozes, 1972. 262 p.
- 61 LEVY, Francis — On the relative nature of relational factors in classification. *Information Storage and Retrieval*, 3(4):315-29, 1967.
- 62 LLORACH, Emilio Alarcos — Les représentations graphiques du langage. In: LE LANGAGE. Encyclopédie de la Pléiade. Paris, Gallimard, 1968, p. 513-67.
- 63 LOGIQUE et connaissance scientifique — Encyclopédie de la Pléiade. Paris, Gallimard, 1967. 1345 p.
- 64 LUCENA, Carlos José Pereira de — *Introdução às estruturas da informação*. Rio de Janeiro, Ao Livro Técnico, 1970. 86 p.
- 65 MANLY, Ron — Detailed discussion of Bar-Hillel's "Theoretical aspects of the mechanization of literature searching". *American Documentation*, 15(2):126-31, 1964.
- 66 MATOSO CÂMARA Jr. J. - *Dicionário de filosofia e gramática referente à língua portuguesa*. 3. ed. Rio de Janeiro, Lozon, 1968. 383 p.
- 67 MEETHAM, Roger — *Informatique et documentation; le traitement automatique de l'information*. Paris, Larousse, 1969. p. 192

MANOEL ADOLPHO WANDERLEY

- 68 MELTON, Jessica — *A use for the techniques of structural linguistics in documentation research*. Cleveland, Western Reserve University, 1964. 20 f.
- 69 MONTGOMERY, Christine A - Automated language processing. *Annual Review of Information Science and Technology*, 4: 145-174, 1969.
- 70 ———— Linguistics and information science. *Journal of the American Society for Information Science*, 23(3): 195-21, 1972.
- 71 MOSS, R. — Minimum vocabularies in information indexing. In: SARACEVIC, Tefko, ed. — *Introduction to information science*. New York, Bowker, 1970. p. 254-63
- 72 MOUNIN, Georges - *Clefs pour la linguistique*. Paris, Seghers, 1968. 189 p.
- 73 MOYNE, J. A. - Information retrieval language. *ASIS Proceedings* 6: 259-63, 1969.
- 74 OLSHEWSKY, Thomas M. - *Problems of the philosophy of language*. New York, Holt, 1969. 774 p.
- 75 PAGES, Robert — Relational aspects of conceptualization in message analysis. *Information Storage and Retrieval*, 3(4):351-75, 1967.
- 76 PERREAULT, Jean M. - Documentary relevance and structural hierarchy. *American Documentation*, 17(3): 136-9, 1966.
- On the articulation of surrogates: an attempt at an epistemological foundation. *Information Storage and Retrieval*, 3(4); 177-92, 1967.
- 78 Perry, James W. & KENT, Allen - *Tools for machine, literature searching; semantic code dictionary, equipment, procedures*. New York, Interscience, 1958. 972 p.
- 79 PIAGET, Jean — *Épistémologie des sciences de l'homme*. Paris, Gallimard, 1970 380 p.
- 80 ———— *Traité de logique; essai de logistiquie opératoire*. Paris, 1949. 423 p.
- 81 PINILLOS, José Luis - La significación desde el punto de vista psicológico. *Revista Española de Lingüística*, 1(1)- 97-117, 1971.
- 82 POSTAL, Paul M. - The method of universal grammar. In: GARVIN, Paul L., ed. — *Method and theory in linguistic*. The Staque, Monton, 1970. p. 113-31.
- 83 POTTIER, Bernard — Hacia una semántica moderna. In: — *Lingüística moderna y filología hispánica*. Madrid, Gredos, 1970. p. 99-133.
- 84 PRIETO, Louis J. - Le sémiologie. In: LE LANGAGE. Encyclopédie de la Pléiade. Paris, Gallimard, 1968. p. 93-125.
- 85 QUENEAU Raymond — *Présentation de l'Encyclopédie de la Pléiade*. Paris, Gallimard, 1956. 62 p.
- 86 RANGANATHAN, S. R. - *Philosophy of library classification*. Copenhagen, Ejnar Munksgaard, 1951. p. 133.
- 87 REICHENBACH, Hans - *Elements of symbolic logic*. New York, MacMillan, 1947. p. 444.
- 88 REY, Alain — Remarques sémantiques. *Langue Française*, 4: 44-66, 1965.
- 89 RODÓN, Eulalia — Categorización, categoria, realización. *Revista Española de Lingüística*, 2(2): 389-406, 1972.
- 90 SALTON, Gerard — Automated language processing. *Annual Review of Information Science and Technology*, 3: 169-192. 1968.
- 91 ———— Automatic processing of foreign language documents. *Journal of the American Society for Information Science*, 21(3): 187-94, 1970.
- 92 SCHULTZ, Louise — Language and the computer. In: Borko, Harold, ed. *Automated language processing*. New York, Wiley, 1967. p. 11-31.
- 93 SERVAN-SCHREIBE R, Jean-Louis - *Le pouvoir d'informer*. Paris, R. Laffontt. 1972. p. 510.
- 94 SHREIDER, Y. A. - On the semantic characteristics of information, In: SARACEVIC, Tefko, ed. *Introduction to information science*. New York, Bowker, 1970.
- 95 SIMMONS. Robert F. - Natural language question — answering systems: 1969. In: SARACEVIC, Tefko, ed: *Introduction to information science*. New York, Bowker, 1970. p. 467-84.

LINGUAGEM DOCUMENTÁRIA: ACESSO À INFORMAÇÃO

- 96 SOERGEL, Dagobert — *Klassifikationssysteme und Thesauri; eine Anleitung zur Herstellung von Klassifikationssysteme und Thesauri im Bereich der Dokumentation*. Frankfurt, Deutschen Gesellschaft für Dokumentation, 1969. 224 p.
- 97 ———— Some remarks on information languages, their analysis and comparison. *Information Storage and Retrieval*, 3(4): 219-91, 1967.
- 98 ———— A Universal source thesaurus as classification generator. *Journal of the American Society for Information Science*, 23(5):299-305, 1972.
- 99 TAULBEE, Orrin E. - Content analysis, specification and control. *Annual Review of Information Science and Technology*, 3: 105-36, 1968.
- 100 TAYLOR, Robert S. - The process of asking questions. *American Documentation*, 13(4):391-400, 1962.
- 101 THOMPSON, Frederick B. - Man-machine communication. In: SEMINAR on computational linguistics. Washington, Govt. Printing Office, 1968. p. 57-68.
- 102 TODA, Masano & SHUFFORD Jr., Emir H. — Logic of systems: introduction to a formal theory of structure. *General Systems Yearbook of the Society for General Systems Research*, 10: 3-27, 1965.
- 103 TRADUCTION automatique et linguistique appliquée. Choix de communications présentées à la Conférence Internationale sur la Traduction Mécanique et l'Analyse Linguistique Appliquée. Paris, Presses Universitaires de France, 1964. p. 286.
- 104 VAN DIJK, Marcel & VAN SLYPE, Georges — *Le service de documentation face à l'explosion de l'information*. Paris, Les Éditions d'Organisation; Bruxelles, Les Presses Universitaires, 1969 p. 265.
- 105 VICISERY, Brian C. - *Technique of information retrieval*. London, Butterworth, 1970. 262 p.
- 106 WAHLIN, Ejnar — Classification systems and their subjects. *American Documentation*, 17(4):199-205, 1966.
- 107 WITTEGENSTEIN, Ludwig. *Notebooks. 1914-1916*, Harper & Brothers, 1961.
- 108 ZULL, Carolyn G. — On the possible relevance of linguistics to information retrieval. *ASIS Proceedings*, 5:285-7, 1968.

NOTAS

1) Para o que se expõe em 4.1, lançamos mão de Landry e Rush que, em sua "teoria geral da indexação" (59), consideram o "document space" como "an R-set (following Russell's notation, that is, it is a set that contains its own description". Interessante nota sobre o aspecto lógico dessa questão já a havia publicado no IBBD B. Inf. 1(3): 85-6, maio/jun. 1955, o Prof. Mario Tourasse Teixeira sob o título "A contradição de Russell e a documentação".

2) Porque só eventualmente nos reportamos ao conceito de "recall", designamo-lo, em 4.5 e em outras passagens, sob a forma corrente de "recuperação", ao invés de o traduzirmos por "revocação", termo proposto pelo Prof. Antonio A. Briquet de Lemos, preferível por mais adequado e fiel ao sentido do original.

3) Ainda em 4.5, optamos por "estrutura de pilha" para a expressão "pushdown store" que C. Montgomery (70) emprega a propósito do SMART e do Syntactic Analyzer. Valemo-nos da versão usada pelo Prof. Carlos J. F. de Lucena para "pushdown stack" em sua "Introdução às estruturas da informação" (64), trabalho em que aprofunda estas últimas quanto à "forma em que as mesmas são implementadas na unidade de armazenamento de computadores digitais".

4) Sobre a "estrutura intelectual" objeto de 6.5 con-signa o próprio R. Blanché (12) o fato de que "ultérieurement, l'idée de l'hexagone est apparue dans un article de L. Hegenberg, A negação, *Revista brasileira de filosofia*, 1957, p. 448-57".

SYNOPSIS

The specific process of documentary communication is envisaged from the point of view of the relations between language and the universe of graphic records which is the object of documentation. As a sample, several tools for bibliographical research, originated to this purpose, are described, attention being paid to the most elaborate of all, the documentary languages. Linguistic implications of content analysis are emphasized, as well as the interpenetrations of the syntactic and semantic levels. These considerations lead to the examination of the logical relations and fundamental categories of natural and documentary languages, and to a brief survey of some basic intellectual structures underlying both.

