

ANÁLISE
SEMÂNTICA E
ANÁLISE
DOCUMENTÁRIA

Johanna
W. Smit

Um estudo das relações existentes entre de um lado a análise semântica (como proposta por A. Greimas) ou então a análise de textos lato senso, praticada em outras disciplinas — história, sociologia, antropologia cultural, exegese, etc. - e de outro lado o processo empregado pela documentação para transformar o discurso científico em unidades recuperáveis de informação, poderia parecer um capítulo de ficção científica (melhor: intelectual) se não fosse o fato deste tipo de comparação estar começando a se difundir entre alguns documentalistas.

Como hipótese preliminar de trabalho pode-se postular que:

1) as soluções dadas a problemas surgidos com a análise semântica poderão também constituir soluções para problemas documentários, e

2) a prática documentária (entendida aqui como a criação e adequação dos algoritmos de análise e a conseqüente manipulação de programas computacionais) pode perfeitamente ser utilizada como meio de verificação de hipóteses semânticas.

Até agora o problema é colocado parcialmente, em termos da relação ovo/galinha que domina boa parte das discussões sobre a compatibilidade das duas áreas (por exemplo no documentalista Coyaud, 1966 e 1972 ou em Greimas, 1966, pp. 65-68, 90, 163). O que me parece ser mais importante (mas menos discutido) é o paralelo que constituem as *operações*, tanto semânticas quanto documentárias, tendo em vista a *seleção de elementos do conteúdo pertinentes e a rejeição (ou a suspensão provisória) de outros elementos, considerados estilísticos e não pertinentes para a construção do modelo* (Greimas, 1966, p. 159). Ou seja: reformulando os termos da hipótese de trabalho acima exposta (1. teoria semântica e 2. prática documentária) emergem duas áreas de operações mentais muito próximas (a ponto de, nas discussões mais recentes sobre o assunto, tanto documentalistas quanto linguistas não serem mais capazes de definir as fronteiras, chegando a confundir autores reputados de uma área, atribuindo suas pesquisas à outra área) e que, num regime de coabitação, criam condições para que se definam melhor as operações necessárias à análise da significação (no discurso e/ou em documentos).

Não se trata de, a todo custo, impor uma funcionalidade aos trabalhos semânticos (que obviamente nem sempre almejam este estatuto, enquanto que esta funcionalidade está sempre implícita nos trabalhos documentários), mas comparar o comparável: as operações envolvidas em ambos os

processos tendo sempre em vista que algumas das preocupações da teoria semiológica (como por exemplo o conceito de redução aplicado ao *carré sémiotique* introduzido por Greimas, 1970) não são compatíveis com a prática documentária (embora sejam talvez operacionais teoricamente), que se vê obrigada a manter um equilíbrio entre a redução (assegurando a economia do sistema) e a redundância (para que a taxa de recuperação da informação justifique o sistema documentário) (Gardin, 1968 e Allard et al., 1963).

A *análise documentária* corrente baseia-se essencialmente num (falso) pressuposto de homogeneidade epistemológica tanto dos discursos científicos como também das relações entre estes discursos e a própria ciência. Consequentemente, a análise documentária (por intermédio de uma linguagem documentária - LD (2)) trata sua *matéria-prima* (o discurso científico) de uma forma frequentemente ingênua, ignorando toda possibilidade de influência do plano discursivo: somente o *conteúdo* lhe interessa. Deste ponto de vista, parece interessante (ou promissor) confrontar os efeitos *transformadores* da análise documentária, não em relação a seu produto - a recuperação da informação (o que, aliás, se faz frequentemente, sob forma de entrevistas e pesquisas junto aos usuários dos serviços de documentação, lembrando os mais simples ciclos tautológicos de *medidas de satisfação* sem possibilidades de estabelecer comparações) mas comparando-os ao discurso científico do qual as LD *pretendem* ser metalinguagens (por enquanto, no sentido hjelmsleviano: *uma semiótica que tem por conteúdo uma outra semiótica*, 1971, p. 150) e às etapas propostas por Greimas para descrever a significação (1966 e 1970). Assim, pode-se estabelecer um paralelo perfeito entre as fases descritivas analisadas por Greimas (constituição do corpus, normalização e construção do modelo - 1966, pp. 141-171) e as etapas previstas por Coyaud durante uma análise documentária *completa* (amputação, normalização e tradução - 1967, p. 55).

Toda a discussão que seguirá girará em torno desta preocupação semântica das LD, de sua busca de um estatuto metalinguístico; pretensão que, embora seja unanimemente aceita pelas mais variadas tendências documentárias, é interpretada das formas mais diversas possíveis, com matizes que vão da extrema simplificação, passando pela ignorância e pela má- interpretação, à extrema sofisticação.

Portanto, estabelece-se um primeiro dado: as LD funcionam (ou querem funcionar) como um tipo de metalinguagem do discurso científico, o que implica um *sistema construído de definições* (Greimas, 1966, p. 15), fator este que à primeira vista nenhuma novidade apresenta para a documen-

tação. Esta aparece, no entanto, nas implicações um pouco menos imediatas deste *ser metalinguagem*: é o desvendamento do discurso como muito mais que uma linearidade comunicacional, a aceitação de suas ambigüidades, competências redutivas e expansivas, efeitos retóricos ou de redundância, etc., aspectos tratados por uma semiótica em desenvolvimento. Entremendo estas facetas do discurso (salvo honrosas exceções, desconhecidas ou então tendo sua importância extremamente reduzida pela maioria dos documentalistas), outras implicações do *manuseio com o discurso* aparecem, estas de efeitos mais visíveis e devastadores em termos documentários: a suposta relação mecânica existente entre o discurso científico, seu referente e a ciência.

Neste particular, a ingenuidade dos documentalistas foi durante décadas surpreendente, porque não só se admitia como irrefutável a identificação total da palavra com a *coisa* (ou referente), como também se acreditava que as relações expressas pelos sistemas de classificação usados nas bibliotecas estavam diretamente ligadas à organização do conhecimento, quando não com a organização do próprio mundo: a consciência de que se tratava de sistemas de relações entre **palavras** - e não **coisas** - é somente recente (Gardin, 1973, p. 141). Juntando-se a esta visão mecanicista do discurso ~~uma~~ uma fé inabalável na objetividade da ciência, seu desligamento de condições sociais de produção e o desconhecimento da economia que a rege, chega-se ao teorema básico de tantos sistemas documentários ultrapassados semiologicamente: 1) se a ciência (como referente) mantém com o discurso científico uma relação mecânica, direta e imediata; 2) o discurso científico, como *coisa* da documentação, mantém uma relação igualmente mecanicista com sua representação metalinguística, o conjunto das LD.

Exatamente no momento em que a consciência da falsidade deste teorema começou a tomar corpo entre alguns documentalistas (J. Perry, J. Melton, Kuznecov e J.C. Gardin especialmente), a semiologia irrompeu no meio documentário como uma ciência capaz não só de resolver os problemas de sinomínia que tanto o afligiam, mas também como aquela que se preocupa com sua matéria-prima: o discurso.

Embora seja difícil localizar temporalmente esta verdadeira *tomada de consciência semiológica* (entre outras porque ela se fez pouco a pouco e ainda agora só é verbalizada por alguns documentalistas), ela permite toda uma nova colocação de problemas, métodos e objetivos documentários.

Em outras palavras, começou-se a questionar a relação estabelecida

entre a metalinguagem (no caso, o conjunto de LD), e o discurso científico analisado. Se cada metalinguagem, desde que estruturadas suas definições, torna-se um modelo (segundo Greimas), estes modelos de análise são, conforme J. Kristeva, *sistemas formais de estrutura isomorfa ou análoga à estrutura de um outro sistema (do sistema estudado)* (1969, p.29).

Está aí o ponto central das atuais discussões documentárias (mesmo que raramente explicitado): a comparação das estruturas (do discurso e das LD), e não mais a valorização de LD conforme critérios de unidade de entrada, organização hierárquica ou segundo vantagens atribuídas a sistemas de classificação numéricos em detrimento dos alfa-numéricos, tratamento dado a sinônimos, etc.

Dentro desta visão, a centena de LD hoje em dia em uso pode ser dividida em dois grandes grupos, mesmo que de forma provisória, porque as evoluções, mudanças e inovações são múltiplas e os limites entre os três grupos bastante fluidos. Esta primeira divisão de caráter nitidamente semântica estabelece-se entre as LD que se limitam a *selecionar* termos ou frases de documentos originais - desconhecendo totalmente as características semiológicas de qualquer discurso, e aquelas que *trans-formam* semânticamente o conteúdo dos documentos numa linguagem mais condensada e/ou normalizada. Estas funcionam por intermédio daquilo que Gardin chamou as relações sintáticas e analíticas (3).

A - Nas *LD não-semânticas, seletivas*, o texto científico não é tratado como uma variável a ser comparada com um plano taxinômico-classificatório qualquer, mas é o único dado da análise. Por meio de filtros que detectam as *palavras vazias* e métodos estatísticos para *reduzir* sinônimos, os índices coordenados ou permutados assim estabelecidos (sistemas *KWÍÇ KWIT* e *KWOC*) são as aplicações práticas da escola distribucionalista da lingüística americana (Harris, seu principal apóstolo, participou de vários congressos de documentação e influenciou diretamente Luhn, Saltón, Lemmon, Swanson e Climenson, todos documentalistas). Embora as LD concebidas nestes padrões de frequência e coocorrência estatística sejam as únicas que visivelmente tenham contado com um apoio de linguistas (e, não semanticistas), suas falhas são múltiplas (como as de qualquer sistema que pretenda reorganizar o discurso sem levar em conta a *significação*), ressaltando-se: **1)** o pressuposto (ainda falso) de homologia total entre o discurso científico (sendo que a superfície discursiva — a estrutura superficial de Chomsky — é fonte de onde ele extrai todos os seus dados) e a ciência; e **2)** o pressuposto (*idem*) da relação de causa-efeito entre a frequência de uma palavra e a informação que ela traz.

Além do método de contagem estatística em detrimento da significação, os sistemas *KVIC* e *KWIT* limitam-se quase sempre aos *títulos* dos documentos, colocando assim mais uma relação mecanicista no rol dos mecanicismos documentários: a relação título/texto (que foi tratada pela retórica mas por enquanto pouco explorada pela semiologia moderna, fato para o qual Greimas chamou a atenção repetidas vezes).

Os resultados práticos destes sistemas documentários seletivos só podem ser decepcionantes, como realmente o são. Funcionam, o que pode parecer paradoxal, em casos muito específicos, quando tratam campos semânticos altamente algoritmizados: é o caso nos *Chemical Titles*, no *Physindex do Commissariat à l'Energie Atomique* (França) ou no sistema documentário do Space Guidance Center (USA). Pode-se dizer que nestes casos, a normalização do discurso analisado é tamanha, que este se acomoda até com um sistema de análise documentária tão simplista e aleatório quanto um *KWIC*. O mesmo não aconteceria em relação aos discursos das chamadas Ciências Sociais.

A maioria das LD existentes adquiriu suficiente liberdade para não somente selecionar termos dos documentos, mas também traduzi-los para outro código mais normalizado (pouco importa sua proximidade maior ou menor com a linguagem natural). Mesmo assim, entre estas LD, as diferenças são enormes e, mais uma vez, a divisão se fará em termos de estruturas gerais dos sistemas e não pelos critérios documentários anteriormente válidos (LD gerais x especializadas, naturais x artificiais, unívocas x multívocas, unidimensionais x multidimensionais, informativas x indicativas, etc.).

Assim, separam-se as *LD semânticas (B)* das *LD com tradução sintática (C)*, sem que seja necessário enfatizar que os métodos mistos, combinando LD seletivas (*A*) e LD semânticas (*B* ou *C*) são muito correntes.

B - As *LD semânticas* caracterizam-se por uma elasticidade de análise, adquirida pela intersecção dos dois planos fundamentais presentes em todos os processos analíticos que partam dos textos: o plano *taxinômico* (o campo de determinação dos textos, que reúne os campos conceituais supostamente interdefinidos e possivelmente hierarquizados de uma ciência) e o plano *sintático* (a análise do texto, o discurso e seu encadeamento sintático) (Henry e Moscovici, 1968). O grau de predominância do plano taxinômico sobre o sintático determinará a maior ou menor hierarquização da LD: uma predominância muito forte define perfeitamente todas as tradicionais classificações biblioteconômicas (*CDU*, *Dewey*, *Bliss*, *Brown*, *Cutter* e

a *Colon Gassification* de Ranganathan) - reaparecendo o mito da organização do conhecimento humano e da exaustividade dos campos por elas tratados. Suas relações analíticas estão geralmente codificadas (bem ou mal), mas as relações sintáticas representam papéis bastante secundários, originando novamente uma fonte de *amarras* que impedirão uma boa análise documentária (leia-se: uma representação em LD isomorfa ou análoga ao discurso analisado, segundo as palavras de J. Kristeva).

No escalonamento que se estabelece pela predominância do plano taxinômico, há uma nítida evolução temporal entre o desmoronamento desta predominância taxinômica e o aparecimento de LD menos hierárquicas.

A procura da maior liberdade de análise documentária (novamente: visando melhor *representação*) não se fez sem atropelos ou auto-críticas, começando pelos sistemas *a-sintáticos* (os *descritores* de Mooers e os *uni-têrmos* de Taube) que obrigaram a criação das relações sintáticas para impedir as falsas combinações de têrmos do léxico - assim surgem os *roles* no thesaurus do Engineer Joint Council, nos trabalhos de J. Perry e de uma equipe da Western Reserve University para a American Society for Metals ou no sistema da companhia Du Pont (USA).

Em seguida introduz-se a noção de *sintaxe* nas LD, que se apresenta sob duas formas nem sempre contraditórias: os sistemas documentários com *sintaxe livre* (a ordem de citação dos descritores não é significativa) e os sistemas com *sintaxe fixa* (nos quais a ordem dos descritores também significa) e que constituem um grupo bastante difundido de LD: as classificações facetadas (produtos dos trabalhos de Vickery, Kyle e Foskett, principalmente) (ver Gardin, 1965).

Nos estudos teóricos que concernem estas *LD-semânticas*, fazem-se vez ou outra alusões a trabalhos lexicológicos, teorias de campos semânticos e discussões sobre a natureza de sinônimos, antônimos ou polissemias, ficando clara uma visão utilitária: os documentalistas parecem aceitar a influência de outra ciência quando esta parece ser imediatamente *utilizável*. A ignorância de diferentes tendências e escolas semânticas também não contribue para o melhor entendimento da área por parte dos documentalistas.

Nesta corrida pela libertação das LD, embora os progressos sejam imensos, os resultados estão ainda longe de serem satisfatórios, mesmo porque a interação do plano taxinômico e sintático ainda é falha, estabelecida de forma não-sistemática e dependendo de interpretações ou provas empíricas

(Henry e Moscovici, 1968, p. 48), o que impossibilita a *formulação de hipóteses novas, a definição de variáveis que saiam do quadro teórico fixado* idem, pp. 47-48). Se comparadas as fases descritivas destas LD com a teoria de Greimas, nota-se àfora detalhes menores, uma não preocupação com a *estruturação* do modelo documentário, por mais que a própria organização taxinômica dos descritores possa fazer crer o contrário.

C - Exatamente este problema impeliu os documentalistas a procurarem mais longe ainda e chegarem à noção das *LD com tradução sintática*. A procura agora estabelece-se por uma ligação dinâmica entre os dois planos analíticos (o taxinômico e o sintático), o que permite uma melhor representação da dinâmica científica, que não passa de uma constante reorganização de conhecimentos anteriores e portanto também uma reorganização permanente da língua e do discurso científico. Voltando à comparação entre as fases descritivas semânticas e documentárias, as LD pertencentes a este grupo são as únicas *completas* porque chegam à etapa final prevista por Greimas, a estruturação do modelo de análise. As três principais variáveis dos sistemas documentários encontram-se finalmente reunidas harmônica-*mente: a ciência, seu discurso científico e o modelo analítico (a LD).

Para que as LD tivessem esta capacidade dinâmica, foi preciso não somente aumentar o rol de suas relações (sintáticas e analíticas), como também, e sobretudo, organizá-las num conjunto lógico, chegando à *tradução sintática* conforme a terminologia de Coyaud, correspondente à *estruturação do modelo* de análise de Greimas.

Enfim, criou-se meios para que as estruturas das LD se aproximassem mais das estruturas dos discursos científicos analisados, dotando-as de dois planos analíticos interdependentes mas não independentes, possibilitando assim uma real estruturação das LD: este é o mérito principal da LD criada por Braffort e Leroy (o *Grisal*) para a Euratom, do *Semantic Code* desenvolvido pela equipe Perry-Kent-Melton, do *Syntol* de J.C.Gardin, do *General Inquirer* de Philip Stone ou ainda dos trabalhos de Kuznecov, Paduceva e Ermolaevna (Moscou).

Desvenda-se claramente uma opinião um pouco menos utilitária da semiologia por parte dos documentalistas, em vista de uma tomada de consciência mais geral daquilo que sejam as LD e suas funções: passou-se da problemática da palavra, seus sinônimos, sufixos, unidades paradigmáticas e etc. para uma análise do discurso científico, da teoria das linguagens e meta-linguagens e suas gramáticas, enquanto que a bibliografia citada pelos docu-

mentalistas separou-se da lexicologia e da lingüística, para passar a citar a semiologia francesa (Greimas, Benveniste, Barthes e Kristeva) e por outro lado a escola chomskysta, acompanhando atentamente sua evolução para a semântica generativa. (Lakoff).

Finalmente, por mais que este panorama possa dar a entender que a simbiose análise semântica/análise documentária já está perfeitamente estabelecida, ela começa somente a ser conscientizada e muito trabalho ainda deverá ser feito, quer em termos de aprofundamento do conhecimento e difusão das terras já conquistadas, quer como organização de novas expedições.

- (1) O presente artigo resume algumas das conclusões de um trabalho, sob orientação do prof. J. Meyriat, aprovado pela École Pratique des Hautes Etudes em junho de 1973 com relatórios dos profs. R. Barthes e A. Greimas.
- (2) As LD são definidas comumente como ferramentas normalizadas de interpretação do conteúdo dos documentos, num nível de generalidade definido à priori. Elas apresentam-se, regra geral, como léxicos mais ou menos estruturados e hierarquizados, podendo ou não contar com gramáticas próprias (ver Gardin, 1968 e Coypaud, 1966).
- (3) O capítulo das relações que unem ou podem unir os termos de um léxico documental é bastante confuso, principalmente porque a terminologia empregada pelos diferentes autores ainda está longe de uma normalização mínima. Baseando-me nos artigos de Gardin, os seguintes dados sobressaem: as relações entre os termos descritores do conteúdo dos documentos podem ser: 1) analíticas ou 2) sintáticas. Esta dicotomia é essencialmente prática e distingue as relações analíticas (ou semânticas) que aparecem explicitamente entre os termos do léxico antes de qualquer uso deste para a análise documental (são as relações mais previsíveis e por isso já mais semantizadas) das relações sintáticas (ou lógicas) que são menos previsíveis à priori e que unem os termos do léxico no momento da análise documental. Uma relação entre dois ou mais conceitos pode ser expressa analiticamente (por um meio semântico: classificação, fatoração, composição) ou sinteticamente (expressa por roles, liens, álgebra de Boole). Gardin insiste no pragmatismo desta divisão e na necessidade de se prever sempre a compatibilidade e convertibilidade entre as duas estruturas — semântica e sintática, (ver Dijk e Slype, 1969, pp. 82-86; Allard et al., 1963, pp. 27-35; Bely et al., 1970, pp. 8-11; Coypaud, 1966, pp. 22-23 e Gardin, 1973, pp. 144-146). Um paralelo com as relações integrativas e distributivas de Barthes, caso pertinente, resta a ser feito (1966).

BIBLIOGRAFIA

- ALLARD, M.; ELZIÈRE, M.; GARDIN, J.C.; HOURS, F. *Analyse conceptuelle du Coran sur cartes perforées*. Paris, Mouton, 1963. 2v.
- BARTHES, R. *Introduction à l'analyse structurale des récits*. Communications, Paris, (8): 1-27, 1966.
- BELY, N.; BORILLO, A.; SIOT-DECAUVILLE, N.; VIRBEL, J. *Procédures d'analyse sémantique appliquées à la documentation scientifique*. Paris, Gauthier-Villars, 1970. 240 p.
- COYAUD, M. *Introduction à l'étude des langages documentaires*. Paris, Klincksieck, 1966, 148 p.
- COYAUD, M. & SIOT-DECAUVILLE, N. *L'analyse automatique des documents*. Paris, Mouton, 1967. 147 p.
- COYAUD, M. *Linguistique et documentation: les articulations logiques du discours*. Paris, Larousse, 1972, 173 p.
- DIJK, M. van & SLYPE, G. van. *Le service de documentation face à l'explosion de l'information*. Paris, Les Éditions d'Organisation, 1969. 265 p.
- GARDIN, J.C. *Free classifications and faceted classifications*. In: ATHERTON, P., ed. *Classification research*. Copenhagen, Munksgaard, 1965, p. 161-168.
- GARDIN, J.C. *Análise documental e análise estrutural em arqueologia*. In: CORDIER, S., ed. *Lévi-Strauss*. São Paulo. Documentos, 1968. p. 81-86.
- GARDIN, J.C. *Document analysis and linguistic theory*. *Journal of documentation*, London, 29 (2): 137-168, June 1973.
- GREIMAS, A. J. *Sémantique structurale: recherches de méthode*. Paris, Larousse, 1966. 262 p.
- GREIMAS, A. J. *Du sens*. Paris, Seuil, 1970. 313 p.
- HENRY, P. & MOSCOVICI, S. *Problèmes de l'analyse de contenu*. *Langages*, Paris, (11): 36-60, sept. 1968.
- HJELMSLEV, L. *Prolégomènes à une théorie du langage*. Paris, Minuit, 1971. 231 p.
- KRISTEVA, J. *Semeiotikè: recherches pour une sémanalyse*. Paris, Seuil, 1969, 379 p.

The approach nearly visible between the steps followed by semiologists (A. Greimas in this particular case) to find out *meaningness* of discourse, and document analysis, the extension of this approach, its limits, and mainly a new critic view of the value of different types of information languages based on semantic-semiologic dates, that was possible by this comparaison.

Le parallèle qu'on peut apercevoir entre les étapes dégagées par les sémiologues (dans ce cas particulièrement A. Greimas) pour rendre compte de la *signification* des discours et l'analyse documentaire des documents, l'extension de ce parallèle, ses limitations et surtout une nouvelle *critique*, devenue possible par cette comparaison, de la valeur des différents types de langages documentaires à base des données sémantico-sémiologiques.