

UM MÉTODO DE ANÁLISE DE DOCUMENTAÇÃO

Francisco Ruas Santos*

Método de análise de documentos baseado na linguagem natural brasileira, na sua sintaxe e na análise combinatória. Aplica-se a processamento manual ou semi-automatizado de dados. Utilizável em qualquer campo do conhecimento humano, embora a exemplificação apresentada diga respeito, na sua quase totalidade, a transportes brasileiros. Posto em prática pelo Centro de Documentação e Publicação do Ministério dos Transportes, em 1974. A forma completa, com finalidade didática, consta do **Manual de análise de documentação**, editado em 1975. Conjugua a aplicação de linguagem controlada através do thesaurus de transportes brasileiros, editado em 1974 pelo referido Centro, sob o título **Pesquisa integrada na documentação do Ministério dos Transportes, thesaurus**.

1 – INTRODUÇÃO

1.1 – Finalidade, campo de aplicação e conjugação do trabalho

Este trabalho tem por finalidade apresentar, de forma sucinta, um método de análise de documentação baseado na linguagem natural brasileira e sua sintaxe, e na análise combinatória.

Esse método deve ser aplicado no processamento manual ou semi-automatizado de dados. Pode ser utilizado em qualquer campo do conhecimento humano, embora a exemplificação nele apresentada diga respeito, na sua quase totalidade, aos transportes brasileiros.

* Diretor do Centro de Documentação e Publicação do Ministério dos Transportes.

Estabelecido formalmente em 1973, foi posto em prática na análise de documentação de transportes realizada em cursos e trabalhos correntes do Centro de Documentação e Publicação do Ministério dos Transportes (1), desde novembro de 1974. Sua forma completa, com finalidade didática, consta de **Manual de Análise de Documentação**, editado por esse Centro, ora no prelo e possivelmente em condições de ser distribuído quando da realização do 8º Congresso Brasileiro de Biblioteconomia e Documentação.

Finalmente, consignamos que a aplicação do novo método é conjugada com a linguagem controlada de um ou mais *thesauri*, no caso os dos Transportes Brasileiros, já editados em 1974 por aquele Centro, sob a forma preliminar de esboço.

1.2 – Origem do método de análise

Em nosso trabalho de pesquisa no campo da História do Brasil observamos que a vida captada ou expressa nos documentos escritos não se reflete, naturalmente, nas pistas ou cabeçalhos das fichas a eles referentes. Além do mais, a cronologia dos eventos não complementa, via de regra, os cabeçalhos das fichas relacionadas com os eventos históricos.

Na bibliografia histórica, por outro lado, os cabeçalhos de assunto são por demais sumários, não correspondendo, também, de um modo geral, quer à atuação dos personagens, quer às situações históricas. Por tudo isso, praticamente todo o trabalho ulterior de busca, com vistas ao simples achado de assuntos subordinados, recai sobre o pesquisador. O tempo que este dispender na realização de trabalhos infrutíferos é enorme. Se quantificado esse tempo perdido, ficaria evidenciado, a par de outros fatores negativos, o quanto é anti-econômico e irracional o trabalho de pesquisa na conjuntura atual, não só da pesquisa histórica, como, de um modo mais amplo, da literária, entendida como a que lida com fontes escritas.

Ainda em nossas pesquisas históricas verificamos que as circunstâncias de lugar – fatores igualmente essenciais para o conhecimento histórico – também não são expressas, geralmente, nos cabeçalhos usuais das fichas dos nossos catálogos. Eis outra falha que também leva o pesquisador, ou o estudioso, a consumir um tempo precioso em buscas freqüentemente inúteis.

Por tudo isso fomos levados a resumir as ações dos personagens – agentes e protagonistas – e as situações passadas, em forma de cabeçalhos, sempre complementados com as circunstâncias de **lugar e tempo** dos eventos.

(1) Daqui por diante designado por CEDOP, sua sigla.

Acima de tudo verificamos que nesse resumo deviam aparecer os verbos, reduzidos a uma única forma — a forma de ação dos agentes.

De fato, podemos distinguir, nos textos, as formas de ação dos agentes, as quais são peças essenciais no sistema de recuperação da informação do qual estamos tratando. Assim, por exemplo, na oração: “*O Ministro dos Transportes inaugurou a estrada X.*”, a palavra **Inauguração** é a forma de ação que define a ação do agente, Ministro dos Transportes.

Os **protagonistas** são pessoas, as **formas de ação** e os **complementos** são os termos que devem ser incluídos nas tabelas ou lista de assuntos dos **thesauri**.

As pessoas só excepcionalmente ali se encontram, sendo, antes, nomes do catálogo onomástico ou do catálogo de assuntos.

1.3 — Exemplificação

Vejamos, para ilustrar, os seguintes textos: “*O Príncipe D. Pedro proclamou a independência do Brasil.*” e “*O Príncipe D. Pedro, ante a rebeldia da junta de governo local, deslocou-se para Minas Gerais, em 23 de março de 1822; e ali encontrou o apoio do povo.*”

A mesma pessoa, D. Pedro, aparece, no primeiro caso, como **agente**; no segundo, como **protagonista**, na segunda oração e **agente** na primeira.

Exemplo com pessoa jurídica:

“*A RFFSA inicia campanha para reduzir o número de acidentes.*” (RFFSA, **agente**) e “*A RFFSA terá como subsidiária uma empresa pública para a construção da ferrovia Minas — São Paulo.*” (RFFSA, **protagonista**).

Nesses dois últimos exemplos podemos distinguir:

- **outros protagonistas**: — acidentes e ferrovia Minas — São Paulo.
- **formas de ação**: — redução (de reduzir) e construção (de construir).
- **Deslocamento, Junta de Governo, Minas Gerais, 23 de março de 1822, (Empresa) subsidiária, Empresa pública**, nos exemplos, são complementos.

Esses agentes, protagonistas e assuntos complementares seriam transformados em cabeçalhos de assunto, conforme normas para a respectiva formulação.

Os textos dos exemplos contêm, assim, mensagens claras. Mas neles ocorrem, ainda, mensagens “codificadas” que devem ser procuradas e “descodificadas”

a fim de, por sua vez, constituírem pistas ou cabeçalhos para a recuperação da informação. Assim, no segundo exemplo:

– Minas Gerais. História política. 1822

e no terceiro exemplo:

– Transporte ferroviário brasileiro. Ferrovia Minas – São Paulo.

2 – A SINTAXE E O SEU RELACIONAMENTO COM A

ANÁLISE DE DOCUMENTAÇÃO

2.1 – O valor das palavras nos textos.

Salvo o caso em que as palavras têm um sentido hermético e erudito, elas corporificam uma linguagem de comunicação direta de fatos e pensamentos. O natural seria que esses fatos e pensamentos se transformassem diretamente em cabeçalhos de assunto.

2.2 – Os primitivos cabeçalhos

Estranhamente não foi assim que ocorreu no campo da Ciência da Informação, no início do seu desenvolvimento. Viu-se, então, a formulação de cabeçalhos decorrer, via de regra, de uma **combinação de substantivos**, do mais geral para o mais particular, tal como teríamos, utilizando-se o último exemplo:

– BRASIL – Transportes – Ferrovia Minas – São Paulo.

No caso dos dois primeiros exemplos:

– PEDRO I – Biologia – 1822.

Observemos, porém, que, neste caso, o usuário teria de ir aos textos para recuperar a pequena história neles condensada.

Realmente, a simples **combinação de substantivos** não basta para sugerir a atuação do Príncipe D. Pedro naquele dramático ano de 1822, como no caso de se ter:

– PEDRO I. Proclamação. 1822 set 7

e

– PEDRO I. Apoio popular. 1822 março,

com o nome próprio, o substantivo e a forma de ação separados por . (ponto) e com o complemento data.

2.3 – Os cabeçalhos como síntese dos textos

Muito embora os textos devam ser sempre elementos de consulta obrigatória, insubstituível, que não se resolve com cabeçalhos de assunto, tais pistas devem ser, tanto quanto possível, a **síntese dos textos**.

Por outro lado, nos **agentes, protagonistas, formas de ação e complementos**, podemos reconhecer logo os elementos fundamentais da análise sintática: **sujeito; predicado; verbo e objetos; adjuntos**.

2.4 – A análise sintática como chave para a transformação dos textos em pistas para a recuperação da informação

Ora, se a análise sintática é a chave para a compreensão dos textos, por que não aplicá-la na transformação destes em pistas mais objetivas para a pesquisa da informação?

Essa possibilidade é um verdadeiro “ovo de Colombo”, mas nem por isso perde a transcendência que deve ter, para nós, documentaristas ou analistas de documentação.

2.5 – Procedimentos básicos da metodologia da análise de documentação

É possível inferir, do que foi dito até aqui, os procedimentos básicos que compõem nosso método de análise de documentação.

- 1º) reduzir os textos à sua expressão mais significativa, traduzida em elementos sintáticos: **sujeito – agente; predicado – forma de ação; objeto indireto, adjuntos – complementos; sujeito – protagonista e os complementos deste;**
- 2º) buscar nos **thesauri** os termos correspondentes a esses elementos, a fim de compor os cabeçalhos ou pistas para a recuperação dos dados e informações contidas nos textos.

2.6 – Resumo da metodologia de análise de documentação.

Em resumo, a nossa metodologia reduz-se a uma **análise de textos e expressá-la na terminologia controlada dos thesauri**, segundo normas que apresentaremos daqui por diante.

3 – APRESENTAÇÃO DE ASSUNTOS NOS TEXTOS

Nomes, assuntos, idéias, e, conseqüentemente, os termos que os representam, apresentam-se nos documentos de modo **singelo** ou **composto**.

Exemplos do primeiro caso:

José da Silva
DNER
Ferro
Solos.

Exemplos do segundo caso:

Processo de fundição de trilhos
Pavimentação de estradas de 1ª classe.

3.1 – Os termos singelos e compostos relativamente aos thesauri

Os termos (não representando pessoas), singelos e compostos, podem estar ou não nos **thesauri**.

Se estão, têm seu uso **autorizado** e expressarão os assuntos e idéias que representam nas pistas para a recuperação da informação.

Se um termo singelo ou composto não estiver no **thesaurus** (lista alfabética de termos), poderá neste ser incluído, desde que não haja um seu sinônimo já incorporado em qualquer das tabelas dos **thesauri**.

Mas não havendo um sinônimo de termo composto no **thesaurus**, poderá não ser necessária essa inclusão, desde que, dividido em partes, estas já estejam incorporadas nos **thesauri**. Por exemplo:

Produção de soja (termo composto, não incluído)
Produção e soja (termos já incluídos)

Logo, Produção de soja não será, necessariamente, incluída num dos **thesauri**.

A inclusão de termos nos **thesauri** obedece a regras rigorosas, das quais trataremos num dos anexos deste manual.

3.2 – Relacionamento de termos

Este relacionamento – a frase da sintaxe – por sua vez, vai do simples à caracterização de um tema, passando pelo complexo. É indicado por . (ponto).

Exemplos dessa gradação, lançando-se mão de frases:

3.2.1 — Relacionamentos simples

Sem idéia de tempo

Texto: Mineração em Minas Gerais

Relacionamentos:

Minas Gerais. Mineração

Mineração. Minas Gerais

Com complemento de tempo:

Texto: Produção dos estaleiros em 1964-74

Relacionamentos:

Estaleiros. Produção. 1964-74

Ao contrário do caso anterior, temos aqui um relacionamento **irreversível**, porque Produção, como termo capitulado em tabela de **formas de ação**, não pode vir em primeiro lugar, pelo motivo que mais adiante apresentaremos.

Com estrutura cronológica

Texto: Produção de navios mínero-petroleiros em 1974

Relacionamentos:

Navios mínero-petroleiros. Produção. 1974

Navios mínero-petroleiros. 1974. Produção.

Obrigatoriamente, no caso de pessoas físicas e jurídicas, ou equivalentes, deve ser feita essa inversão de data, a fim de que se possa estruturar, e, portanto, recuperar cronologicamente, sua vida, ou evolução através do tempo. Igualmente, quando se tratar de assuntos de particular importância para o órgão de documentação, como no caso deste exemplo, para o CEDOP.

3.2.2 — Relacionamento complexo sem constituir tema

Ocorre, em geral, quando se trata de modo de operar ou fazer, isto é, **processo**, implícito ou explícito.

Texto: Pavimentação de rodovias em solos de sedimentação.

Para formular os relacionamentos, neste, como em casos análogos, é preciso, conforme já evidenciamos, decompor a frase em agentes e/ou protagonistas.

Temos, então:

Agente (s) — não há

Protagonistas:

Rodovias

Solos de sedimentação

Processos tecnológicos (mensagem descodificada ou protagonista implícito)

Forma de ação:

Pavimentação.

Com exceção de Solos de sedimentação, todos esses termos estão nos **thesauri**.

Precisamos, pois, relativamente à frase não existente nos **thesauri**, decompor a em

Solos

Sedimentação

ou

dar entrada na frase, com vistas à sua inclusão nos **thesauri**. Se for este o caso, assinalaremos a frase ou termo a incluir com o sinal (+).

Relacionamentos obtidos:

Simple

Rodovias. Pavimentação (quê de quem)

Rodovias. Solos de sedimentação

(+)

Solos de sedimentação. Rodovias

Solos de sedimentação. Pavimentação

Processos tecnológicos. Rodovias. Pavimentação

Processos tecnológicos. Solos de sedimentação

Pavimentação.

(quê

em

relação

a

quem)

Nos dois últimos relacionamentos ainda temos o tipo simples, pois, na verdade, estamos relacionando Pavimentação de rodovias e Solos de sedimentação, com Processos Tecnológicos.

A chamada indexação coordenada pode resolver problemas de recuperação análogos, através do relacionamento, imediato e de uma só vez, de

Rodovias

Pavimentação

Solos de sedimentação

Processos tecnológicos.

Realmente, comparado o processo da indexação coordenada com o do exemplo, vê-se logo a vantagem do primeiro. Todavia, a indexação coordenada tem suas limitações e estas não permitem usá-la de modo indiscriminado. Eis porque, quando elas ocorrem, é preferível, no processamento manual, empregar relacionamentos, como os apresentados, do tipo complexo, sem constituir tema.

3.2.3 — Temas

É sabido que os temas caracterizam-se pela sua elevada complexidade, superior ao grau mais alto de relacionamento, o qual é, em geral, o que indica **processo, norma de ação**, ou idéias semelhantes, no campo tecnológico.

A complexidade dos temas está, muitas vezes, antes no seu conteúdo do que na forma de expressá-los. Assim, a expressão História dos Transportes Brasileiros, formulada por meio de um só termo do **thesaurus**, indica um conteúdo riquíssimo, o qual envolve, necessariamente, todos os aspectos dos Transportes Brasileiros, desde passado muito remoto até os nossos dias.

Diante de um tema, dois casos podem apresentar-se ao analista:

- a) o tema é do tipo simples, ou expresso num único termo de **thesaurus**, tal o caso do exemplo que acabamos de focalizar;
- b) o tema é do tipo complexo.

Neste caso, o analista deve decompor o tema em relacionamentos simples e/ou complexos, os quais serão as pistas para ulterior recuperação da informação contida no tema.

Exemplo de tema do tipo complexo e de sua decomposição (todos os assuntos referidos a Transportes Brasileiros):

“Navio/A arte e a técnica de uma indústria/

A construção de navios, de grande e médio porte, não é mais segredo para o Brasil. A cada semana novos navios são lançados ao mar; os estaleiros têm encomendas que os manterão ocupados ainda por muito tempo, o mesmo ocorrendo em relação à mão-de-obra, com a ampliação do mercado de trabalho.

A técnica brasileira se aprimora com incrível rapidez, a tal ponto que os estaleiros nacionais não se limitam a atender ao mercado interno: países desenvolvidos como os Estados Unidos e Alemanha Ocidental compram navios fabricados no Brasil, conscientes de que o produto é bom, bem feito, de qualidade igual ao que eles mesmo produzem.

Como se faz um navio, quem são os homens que, no anonimato de um estaleiro, se ocupam da construção naval? O Estaleiro Mauá, da Companhia Comércio e Navegação, foi escolhido, entre outros de idêntica importância, para mostrar a arte e a técnica brasileiras de produzir navios.” (Jornal dos Transportes, agosto de 1974, p 3).

Tema:

Construção naval.

Decomposição:

Indústria de construção naval. Desenvolvimento.
1974

Navios. Exportação. 1974

Navios. Produção. 1974

Estaleiro Mauá. Produção. 1974

Estaleiros. Desenvolvimento. 1974

Construção naval. Aperfeiçoamento. 1974

Observar que todos os relacionamentos aí apresentados são simples. Todavia, houve necessidade em decompor antes o seguinte relacionamento complexo, textual:

“Arte e técnica brasileiras de produzir navios”, em *“Arte brasileira de construir navios”* e *“Técnica brasileira de construir navios”*, os quais puderam ser reduzidos a:

Construção naval. Aperfeiçoamento. 1974.

De vez que Construção naval já é Tecnologia (v. *Thesaurus fundamental*).

4 – RELACIONAMENTOS DO PONTO DE VISTA NUMÉRICO

No exemplo que acabamos de explorar encontramos os seguintes termos ou assuntos:

Construção naval

Desenvolvimento

1974

Navios

Exportação

Produção

Estaleiro Mauá

Estaleiros

Aperfeiçoamento

Tecnologia (implícito em Construção naval).

Temos, portanto, dez termos, ou assuntos, os quais, do ponto de vista da análise combinatória, dariam 3.628.800 permutações simples. Só com este número poderíamos, *a priori*, ter a certeza de recuperar todas as informações contidas no texto, no saldo de todos os seus relacionamentos.

Ora, tal procedimento é proibitivo na prática do processamento manual, ainda que o número de assuntos a permutar fosse bem menor.

A indexação coordenada é uma solução em busca da certeza matemática; mas é parcial, além de só poder ser aplicada nos casos em que não ocorre quantidade muito grande de documentos em faixa muito estreita de assuntos, ou quando os documentos exigem o emprego de poucos termos do *thesaurus* do órgão de documentação.

Como proceder, então, para se ter a certeza moral de que todos os assuntos relevantes apresentam-se nos seus relacionamentos significativos?

É da resolução desse problema que passaremos a tratar, a partir das noções dadas acerca dos relacionamentos dos tipos complexos e simples, e de temas a eles sempre redutíveis.

Antes, porém, de tratar da resolução prática do problema da obtenção da *certeza moral* na recuperação da informação, devemos apresentar alguns elementos fundamentais da problemática da análise de documentação.

5 – ELEMENTOS FUNDAMENTAIS DA PROBLEMÁTICA DA ANÁLISE DE DOCUMENTAÇÃO

Na base de uma cabal e correta análise de textos com vistas à recuperação da informação, estão, antes de mais nada, os seguintes fatos axiomáticos:

- 1º) todo e qualquer documento é um conjunto de mensagens “codificadas” e/ou claras, redutíveis sempre a relacionamentos complexos e/ou simples;
- 2º) reciprocamente, não se pode exigir da recuperação da informação relativa a um documento, mais do que existe neste.

Isso é verdade estabelecida para qualquer texto, desde o de poesia mais hermética até o de um jornal popular carioca, facilmente compreendido por centenas de milhares de pessoas com pouco mais do que instrução primária.

Vista a questão de outro ângulo, podemos afirmar que todo e qualquer texto contém, simultânea, ou separadamente, a atuação de um agente, a presença de **protagonistas** e de **complementos**.

5.1 – Tipos de análise de textos

Logo, podemos reduzir a análise de textos a dois tipos:

- 1) de **profundidade**, implicando em “descodificação de mensagens

- da mais difícil compreensão;
- 2) de **superfície**, ou formal, caracterizada pelo estabelecimento de relacionamentos simples e/ou complexos dos termos de mensagens claras, relativas aos agentes, protagonistas e complementos.

5.2 – Requisitos do analista de documentação

Ora, uma vez que os dois tipos de análise de textos são inseparáveis na prática quotidiana, ainda quando se trata de textos corriqueiros, conclui-se que o analista de documentação precisa preencher os seguintes requisitos interdependentes:

- a) capacidade intelectual, entendida esta como **possibilidade de interpretar** o texto;
- b) **cultura específica**, ou familiaridade com os conhecimentos em que podem estar inseridas as mensagens do documento.

O primeiro requisito, embora inato, supõe, para que se realize, a existência de um certo grau de cultura específica; o segundo requisito pode ser adquirido. Ambos são indispensáveis para caracterizar um verdadeiro analista de documentação.

Além da necessidade de inteligência e cultura, não vemos outras regras para a formação de um analista de documentação que atenda ao primeiro requisito.

Quanto à satisfação do segundo requisito, basta o conhecimento firme das regras da **análise formal**, as quais pretendemos condensar neste trabalho.

É a satisfação dos dois requisitos que dará a certeza moral de que toda a informação contida no texto foi transformada pelo analista em pistas para a sua recuperação; pois a certeza matemática, além de não poder ser atingida na prática, compõe-se, geralmente, de elementos irrealis ou pouco significativos, conforme veremos oportunamente.

5.3 – Regras fundamentais da análise de documentação

Visto como é absolutamente necessário fazer-se, ao mesmo tempo, a **análise de conteúdo** e a **formal**, como deve ser, na prática, o trabalho do analista?

Eis as regras básicas, gerais:

- 19) procurar compreender, e **compreender**, a ou as mensagens do texto, quer as expressas de modo claro, quer de modo oculto;
- 20) verificar a existência, ou não, de um ou mais agentes, e formular

- sinteticamente, em ordem direta, sua atuação;
- 3º) identificar as formas de ação correspondentes aos verbos das orações formuladas;
 - 4º) formular, com a terminologia dos *thesauri*, a oração ou orações relativas aos agentes;
 - 5º) identificar o/ou os protagonistas, os complementos e seus relacionamentos.

1º exemplo

Texto:

"Prodasen procederá legislação brasileira".

Compreensão do texto: envolve, no caso, o conhecimento da existência do Centro de Processamento de Dados do Senado Federal, em Brasília, e que ele, mediante convênio com órgãos da administração federal, porá à disposição destes a informação legislativa que armazenou; e, também, em que nível essa informação poderá apresentar-se.

Agente: Prodasen. Oração em ordem direta: desnecessário, por já se apresentar assim.

Forma de ação: processamento.

Protagonista: legislação brasileira.

Complementos: não há.

2º exemplo

Texto:

"Com o Aviso de 21 de julho último expedido pelo ministério a cargo de V. Exa. recebi para informar o requerimento que o comendador Manoel Antônio Guimarães dirigiu ao Governo de S.M. o Imperador pedindo para uma companhia que projeta organizar o privilégio exclusivo da navegação a vapor nas baías e rios de Paranaguá, Morretes e Antonina, conforme as bases que acompanham o mencionado requerimento.

Em resposta cabe-me assegurar a V. Exa. que com a possível brevidade darei cumprimento ao que V. Exa. me ordena." (Correspondência de 13 ago 1857, do Presidente da Província do Paraná com o Ministério do Império).

Compreensão do texto: Manoel Antônio Guimarães, depois Visconde de Nacar, é o personagem mais importante de Antonina, naquela época; o Presidente da Província que firma o documento é José Antônio Vaz de Carvalhaes.

Agentes: esses dois personagens; orações respectivas:

José Antônio Vaz de Carvalhaes faz comunicação ao Ministro do Império.

José Antônio Vaz de Carvalhaes cita requerimento de Manoel Antônio Guimarães.

Manoel Antônio Guimarães requer privilégio para companhia de navegação.

Manoel Antônio Guimarães pretende organizar companhia de navegação.

Formas de ação:

Comunicação

Citação

Solicitação

Organização.

Orações na terminologia dos **thesauri**:

- José Antônio Vaz de Carvalhaes. Comunicação. Ministro do Império.
- José Antônio Vaz de Carvalhaes. Citação. Requerimento. Manoel Antônio Guimarães.
- Manoel Antônio Guimarães. Solicitação. Privilégio.
- Manoel Antônio Guimarães. Organização. Empresa privada (de navegação).

Protagonistas:

Transporte aquático

Requerimento

Paranaguá

Morretes

Antonina

História dos Transportes (Brasileiros) (1857)

Paraná (província, implícito)

Ministério do Império.

Complementos:

13 ago 1857

Visconde de Nacar

Presidente da Província do Paraná (1857).

Neste exemplo, se escluirmos os dois últimos complementos, temos 2 nomes próprios, 4 formas de ação, 8 protagonistas e 1 complemento, ou quinze termos a que o texto se reduz. Teoricamente, para recuperá-los, uns relativamente aos outros, teríamos que compor 100.588.236 (2) permutações!

(2) Produto dos 15 primeiros números inteiros.

Se, obviamente, não faremos todas essas permutações, como teremos a certeza moral de recuperar os dados e informações contidas no texto?

Primeiramente, dando expressão aos relacionamentos simples, complexos e temas.

Como temas de maior significação, temos:

História dos Transportes Brasileiros no Paraná em 1857

Atuação do Visconde de Nacar em 1857.

Relacionamentos complexos: os já apresentados quanto aos agentes.

Relacionamentos simples: os dos protagonistas entre si, complementados pela data.

Observe-se que o segundo tema já está expresso com os relacionamentos complexos em que Manuel Antônio Guimarães, Visconde de Nacar, aparece como agente. Por conseguinte, ou ficamos com o tema ou com os dois relacionamentos complexos; pois, como pistas para a recuperação da informação, qualquer dos dois procedimentos é satisfatório. Ficando com os dois relacionamentos, já teríamos estes e mais os dois em que entra José Antônio Vaz de Carvalhaes; quatro relacionamentos, portanto. Estes satisfazem plenamente quanto à recuperação de dados e informações referentes aos dois personagens, por volta de agosto de 1857.

Essa afirmativa é válida para todos os casos análogos, desde que reconhecidos e isolados os agentes nos textos a analisar.

Restam-nos, para resolução, os casos dos protagonistas e complementos de tempo, ou data, em número de nove.

Já sabemos que teríamos que formar 362.880 (3) permutações, as quais cobririam todas as hipóteses da recuperação que buscasse os protagonistas e a data, de todos os modos pelos quais podem eles apresentar-se.

Mas, se fôssemos assim proceder, obteríamos alguns relacionamentos irrealis e pouco significativos, e até mesmo ininteligíveis, como, por exemplo: Requerimento. Morretes. 1857 ago 13. Paranaguá. Antonina.

História dos Transportes. Ministério do Império. Transporte aquátil.

(3) Produto dos 9 primeiros números inteiros.

5.4 — As frases e os arranjos simples de termos dois a dois

“*Se repelirmos o natural, ele volta ao galope*”, eis o que diz verso de um poeta da Antiguidade, muito divulgado até hoje.

No exemplo que acabamos de explorar, o **natural** é a **sintaxe, com as frases ou enunciados providos de sentido próprio, completo**. Expulsamos a linguagem natural e ela volta rápido para fazer com que a empreguemos.

De fato, reexaminando-se o texto, vemos que ele nos oferece, naturalmente, as seguintes frases com relacionamentos simples de protagonistas e data:

- Transporte aquático em Paranaguá em 1857
- Transporte aquático em Morretes em 1857
- Transporte aquático em Antonina em 1857
- Ministério do Império e transporte aquático em 1857
- Requerimento ao Ministério do Império, por volta de 13 de agosto de 1857,

num total de 6. Se invertermos os termos desses relacionamentos, como, por exemplo:

— Antonina e o transporte aquático em 1857,

obteríamos mais seis relacionamentos diferentes. Os doze relacionamentos assim feitos, com a permutação da data, seriam elevados ao dobro, ou a vinte e quatro, correspondentes às frases formadas. Ainda é muito, mas já temos a certeza moral de conseguir a recuperação de dados e informações significativos.

Na realidade, fizemos os **arranjos simples, dois a dois**, dos protagonistas, e complementamos todos esses arranjos com a data. Com esta, obtemos 56 arranjos; sem ela, 72. Do primeiro grupo, fariamos apenas 12 realmente significativos, os quais nos dão a **certeza moral** de se poder recuperar as informações baseadas em **relacionamentos simples, derivados das frases em que o texto pode ser decomposto**. Lembremo-nos, aqui e sempre, **de que o limite da recuperação nos é dado pelo conteúdo do texto**.

Em síntese, associou-se a linguagem natural com a terminologia dos **thesauri**, em arranjos parciais de termos dois a dois.

Já estamos, assim, muito aquém não só do número total de arranjos, mas do de permutações possíveis (362 880).

Mas o número total de relacionamentos complexos e simples é muito elevado, mesmo sem considerar a data, para ser feito pelo processamento manual; isto é, 16 (4 complexos, dos agentes; 12, dos protagonistas).

Não poderíamos reduzir esse número, sem prejuízo da certeza moral quanto à recuperação da informação?

6 – CONVENÇÕES PARA REDUZIR O NÚMERO DE RELACIONAMENTOS, FORNECIDAS PELA PRÁTICA

6.1 – Referentes à data

6.1.1 – A data, a não ser que seja necessário destacá-la como efeméride, nunca principia um relacionamento.

Se nos reportarmos ao exemplo anterior, pudemos, aplicando essa convenção, diminuir em 8 o número de arranjos (72 – 8).

6.1.2 – Nos relacionamentos simples só se faz inversão de data quando o primeiro termo é pessoa física, jurídica e assemelhada, e, assim mesmo, neste caso, quando o órgão de documentação precisa estruturar sua evolução cronológica.

Voltando ainda ao exemplo anteriormente focalizado, não seria, então, cabível, ter o relacionamento

– Transporte aquático, em 1857, em Paranaguá, e análogos, o que nos permitiu abater, do total, 12 relacionamentos simples com inversão de data.

6.1.3 – Nos relacionamentos complexos, a data é o último termo e só deixa essa posição, para vir logo depois do agente, quando este for pessoa física, e houver interesse em estruturar a evolução cronológica de pessoa jurídica, ou equivalente.

Nessas condições, nos relacionamentos complexos do exemplo referido, teríamos:

– José Antônio Vaz de Carvalhaes, em 13 de agosto de 1857, comunicação. . .

e

– Manuel Antônio Guimarães, em agosto de 1857 (aproximadamente), solicitação. . . e organização. . .

6.2 – Referentes às formas de ação

As formas de ação, assim capituladas nas tabelas dos *thesauri*, nunca principiam um relacionamento. Por conseguinte, não poderíamos ter:

– Produção. Soja

e, sim,

– Soja. Produção.

Para se poder expressar um relacionamento do tipo Produção. Soja, imposto pela freqüência ou pelas necessidades particulares do órgão de documentação, seria preciso que ele incluísse no seu *thesaurus* ou lista de termos

– Produção de soja.

Observamos, a propósito, que há um certo número de formas de ação que não estão nas respectivas tabelas, no caso dos *thesauri* elaborados pelo CEDOP, já editados e distribuídos a órgãos de documentação e bibliotecas. Constituem assuntos comuns, como é o caso de Lubrificação; e, por isso, podem encabeçar relacionamentos simples, como no caso:

– Lubrificação. Carros de passeio.

6.3 – Referentes à hierarquia dos assuntos

Se levamos em conta a hierarquia de alguns assuntos relativamente a outros, podemos ainda reduzir o número de relacionamentos simples, sem prejuízo da recuperação da informação.

Tomemos o seguinte trecho:

“O entrevistado mencionou a construção de navios graneleiros pelo Estaleiro X como parte do Plano de Construção Naval de 1974.”

Decompondo o trecho em agentes, formas de ação e protagonistas, temos:

José da Silva (entrevistado)

Citação

Navios graneleiros

Construção

Estaleiro X

Plano de Construção Naval, 1974.

Primeiramente, relacionamento complexo, de agente:

José da Silva. Citação. Navios graneleiros (construção).

Relacionamentos simples, sem levar em conta as frases, e aplicando a convenção de 6.1.2:

Navios graneleiros. Construção

Navios graneleiros. Plano de Construção Naval de 1974

Estaleiro X. Plano de Construção Naval de 1974

Plano de Construção Naval de 1974. Navios graneleiros

Plano de Construção Naval. Estaleiro X. Construção.

Portanto, seis arranjos e combinações que nos dariam a certeza moral de recuperar as informações do texto.

Atentando, porém, para este, percebemos que as idéias que contém assim se dispõem, da mais ampla para mais restrita:

- Plano de Construção Naval de 1974
- Estaleiro X (construção
- Navios graneleiros.

Observemos, também, que tais elementos assim podem ser dispostos hierarquicamente:

- Estaleiro X – subordina-se a Plano de Construção Naval
- Navios graneleiros – subordina-se a Estaleiro X.

Se assim é, podemos eliminar os relacionamentos de Navios graneleiros com Plano de Construção Naval de 1974, em número de dois, e Estaleiro X e Plano de Construção Naval, um, ficando, apenas, com três relacionamentos, em vez de seis.

De fato, se vamos mencionar o Estaleiro X como agente que constrói Navios graneleiros, e estes estão relacionados com Construção, ficam totalmente irrelevantes os relacionamentos

- Navios graneleiros. Plano de Construção Naval de 1974
 - Plano de Construção Naval de 1974. Navios graneleiros,
- assim como
- Estaleiro X. Plano de Construção Naval de 1974.

O primeiro relacionamento pode ser eliminado, porque seria mero dobramento de Navios graneleiros. Construção, aliás correspondentes a uma das frases do texto; o segundo, porque o Plano de Construção Naval de 1974 refere-se, diretamente, ao Estaleiro X, também correspondente a outra frase do texto; e, o terceiro, porque é dobramento do relacionamento complexo no qual o Estaleiro X é o agente, e, como tal, frase de sentido mais completo, abrangendo justamente o que o Plano atribui ao Estaleiro:

- Estaleiro X. Construção. Navios graneleiros.

7 – PROCEDIMENTOS PRÁTICOS

Pensamos que já se pode verificar o quanto é objetivo, na análise dos textos dos documentos, buscar ou formular desde logo:

- a(s) oração(ões)
- a(s) frase(s) realmente mais significativas.

De fato, nas orações temos os agentes, suas formas de ação, protagonistas e complementos; nas frases, os relacionamentos aos quais devem corresponder os arranjos simples de termos dois a dois.

Caso as frases conduzam a arranjos mais elevados, devem ser decompostas de modo a que se recaia nesses arranjos dois a dois.

8 — A APRESENTAÇÃO MATERIAL DO MÉTODO

A análise de textos, segundo o método apresentado, é expressa num documento por nós denominado **boletim de análise**.

Esse boletim deve ser o suporte:

- 1º) da **referência** do documento, feita segundo a NB-66, com simplificações;
- 2º) da **lista** de cabeçalhos de assuntos formulados conforme o método;
- 3º) de **apreciação crítica** do texto ou do documento;
- 4º) de **resumo** do texto, em casos muito especiais (resumo com a modernização de textos arcaicos, por exemplo).

Nesse boletim são lançados os elementos usuais do serviço de documentação: classificação, número de registro, local ou locais onde o documento pode ser encontrado, etc.

Em poucas palavras, o nosso **boletim de análise** deve ser o suporte dos vários procedimentos biblioteconômicos ou de documentação.

Muito mais significativo, porém, para nós, é ainda o fato de o nosso **boletim de análise** poder substituir os resumos convencionais dos textos documentais, fazendo com que eles se convertam, de imediato, em cabeçalhos ou pistas dos assuntos mais relevantes ou interessantes contidos nos textos.

Assim, além de servir ao processamento do documento e dos seus dados e informações, o boletim de análise permite que qualquer órgão de documentação faça, também, no todo ou em parte, o mesmo processamento, **mediante simples trabalho mecanográfico**. Ante a escassez, continuamente acentuada, de mão-de-obra especializada em documentação (lembremo-nos sempre da dramática “explosão das informações”) — eis outra vantagem inerente ao nosso boletim de análise.

Assinalemos que, na edição de índices, o boletim faculta substancial economia em papel e trabalho gráfico, visto como os índices podem reduzir-se à reprodução integral do texto dos boletins, ordenados estes pela entrada de autor, ficando os assuntos referidos apenas aos números de ordem dos boletins. Observamos que, relativamente aos índices convencionais, tal procedimento permite reduzir de um a dois terços os índices reproduzidos, **sem prejuízo das informações que eles devem conter**. E, neste particular, associamo-nos àqueles que, por outros caminhos, preconizam o mesmo procedimento salutar.

A título de ilustração, transcrevemos, de forma sincopada, um boletim de análise elaborado no CEDOP:

385
v234a VALE, Manoel Alves do. **Ação administrativa em defesa do desenvolvimento dos Transportes**. RJ, CEDOP, 1971.
433/1974 CEDOP/Brasília

Lista de assuntos

- 1) SERVIÇO DE DOCUMENTAÇÃO. 1971 (a)
- 2) Kalbourian, Jacques, diag
- 3) Mynssen, Maria, il
- 4) Rocha Maia, Augusto César de Sá da, pref... p 5-7
.....
- 8) DRUCKER, PETER ----- (b)
.....
- 11) ANDREAZZA, MÁRIO DAVID. Interpretação. Plano Nacional de Viação, 1973. 1971? p 11-2
.....
- 13) EMILIO GARRASTAZU MEDICI. Transportes ... p 11s
- 14) BELTRÃO, HÉLIO. Administração pública. 1971? . . . p 12, 14
- 15) ADMINISTRAÇÃO PÚBLICA. Beltrão, Hélio. 1971?-----
- 16) DECRETO—LEI 200/1967 fev 25. Princípios . . . p 13
- 17) TRANSPORTES BRASILEIROS. Administração pública
- 18) ADMINISTRAÇÃO PÚBLICA. Transportes
- 19) Reforma Administrativa — Use Decreto-lei 200/1967
.....
- 23) MINISTÉRIO DOS TRANSPORTES. Produção, 1970 out 9 . . . p 17s
.....
- 32) GALVES, CARLOS. Transcrição . . . p 22-4, 34, 75, 77
.....
- 35) PREÇOS. Reajustamento. . . p 25s
.....
- 37) ORÇAMENTO. Atualização . . . p 31s
- 38) ----- Preços -----
- 39) PREÇOS. Jurisprudência. . . p 32s
.....
- 42) ABNT. Normalização. Preços. Reajustamento. 1957 fev. . . p 36s
- 43) ----- UP Associação Brasileira de Normas Técnicas

- (c)
- 44) Associação Brasileira de Normas Técnicas -- Use ABNT

 46) MINISTÉRIO DOS TRANSPORTES. Solicitação. Norma: Preços. 1957. . . p36

 55) Acréscimo ao **thesaurus** de Transportes Brasileiros: Obras Públicas T/AA4Z UP
 Serviços públicos
 TR
 Serviços de utilidade pública (c)

 58) PREÇOS. Legislação. 1957-71 . . . p 36s, 107s

 71) WALD, ARNOLD. Citação . . . p 64, 66

 75) **O DIRIGENTE CONSTRUTOR**, v6, n11, ago 1970. Transcrição . . . p 74
 76) MINISTÉRIO DOS TRANSPORTES. Reconhecimento do mérito. . . p 77

 79) BARBOSA, João Carlos Gurgel. Distinção entre atualização e reajustamento e
 sistemática dos seus cálculos. In: . . . p 83-105
 79.1) PREÇOS. Reajustamento. 1970
 79.2) PREÇOS (índices). Tabelas . . . anexos
 79.3) TABELAS. Preços (índices) -----

 81) PORTARIA 150/1965 mar 17 . . . p 109-18
 81.1) CONCORRÊNCIAS. Legislação. 1965-71. . . p 109-18
 82) PREÇOS. Bibliografia. . . p 121-2

Apreciação crítica

Aspectos interessantes da problemática administrativa focalizada nos preços das empreitadas e seu reajustamento por força da inflação.

Notas à margem do exemplo de boletim

- (a) Convenção interna do CEDOP.
 (b) O travessão indica repetição dos termos situados acima dele.
 (c) Indicação para o serviço de informática, no caso, do CEDOP.
 (d) A terminologia adotada é a do esboço de conjunto de **thesauri** elaborado pelo CEDOP para seu uso e como contribuição ao estudo de uma linguagem controlada a usar pelos órgãos de documentação do Ministério dos Transportes.
 (f) No processamento, seriam feitas, em formulário contínuo, tantas fichas iguais

à referência quantos são os cabeçalhos de assunto e complementos de autoria, número esse majorado com o número de fichas necessárias à indicação dos títulos; a cada uma dessas fichas, menos à de autor, seriam atribuídos os cabeçalhos da **lista de assuntos**, trabalho que hoje em dia, conforme é sabido, pode ser feito de modo semi-automatizado.

A página (capítulo ou parte) será colocada no fim da referência.

9 – CONCLUSÃO

O método de análise apresentado tem como origem, ou processo deflato-
ratório do seu desenvolvimento, nossa experiência no terreno da pesquisa histó-
rica. Obviamente, essa, como qualquer outro tipo de pesquisa literária, deve ser apoi-
ada pela documentação, em particular na sua fase de pesquisa pura.

Assim sendo, buscamos na sintaxe brasileira a base para a formulação dos
cabeçalhos; através de **orações e frases**, estas com relacionamentos que levem a
arranjos simples de termos dois a dois da análise combinatória. E isso, de modo a
que as permutações relativas aos **agentes**, ativos ou passivos, e esses arranjos, dêem-
nos a **certeza moral** quanto à recuperação dos assuntos significativos ou relevantes
contidos nos documentos.

Para os termos simples, ou absolutos, não há problema de relacionamen-
to ou arranjo dois a dois, a não ser o de data e lugar.

Buscamos complementar sempre esses cabeçalhos com as circunstâncias
de tempo e lugar, enfatizamos, e, freqüentemente, em se tratando de pessoas, a data
dos eventos a elas seguem imediatamente.

—Apresentamos o suporte desse trabalho analítico, o **boletim de análise**,
o qual deve ser o suporte, também, dos demais elementos usuais em documentação
(registro, classificação, etc.).

Destacamos que o nosso **boletim de análise** pode substituir os resumos conven-
cionais de textos de documentos, e servir, de pronto, quer à elaboração de índices,
quer, principalmente, ao processamento manual de dados em órgãos de documenta-
ção carentes de mão-de-obra especializada.

Quanto à terminologia adotada, deve ser a do **thesaurus** ou lista de ter-
mos do órgão de documentação. Assim, o método de análise se conjuga com uma
linguagem controlada.

Esse sistema — método de análise e linguagem de indexação —, acentua-
mos, deve ser incutido no documentarista ou analista de documentação, através de

pequeno curso de preparação, metade teórico-prático, metade de aplicação. Os instruendos não precisam ser, necessariamente, formados em Biblioteconomia, Documentação ou Arquivística. Mas devem evidenciar, em seleção prévia, as seguintes qualificações: nível de escolaridade médio, no mínimo, e, como condição eliminatória, predisposição para as tarefas de análise que deverão realizar depois dessa capacitação. No caso do CEDOP, a experiência revelou ser preciso a carga de 120 horas de atividades em classe, com 120 outras de trabalho fora de classe, distribuídas de 1 a 2 meses, para que seja feita a capacitação.

Surpreendentemente, alunos sem nenhuma experiência prévia em biblioteconomia, documentação e arquivo, ficaram capacitados como bons analistas de documentação de transportes.

Abstract

Method for the analyse of documents based on the Brazilian natural language and its syntaxis, and on combination analyses. Application possible to manual or semi-automatized data processing, on any diawn from one are, Brazilian Transportation. Actually used by the Centro de Documentação e Publicação do Ministério de Transportes, 1974. Final format, with didatic objectives is included in the *Manual de Análise de Documentação*, ed 1975. Combines the application of controlled language through the thesaurus of Brazilian Transportation, ed. 1974 by the Centro, with the title *Pesquisa integrada na Documentação do Ministério dos Transportes, thesaurus*.