

# RECURSOS LINGUÍSTICOS PARA ANÁLISE DE VOCABULÁRIO CONTROLADO: O caso do SAUSP

LINGUISTIC RESOURCES FOR ANALYSIS OF CONTROLLED VOCABULARY: The case of SAUSP

Jessica Camara Siqueira\*

## RESUMO

O vocabulário controlado é um importante instrumento de recuperação de informações, já que intervém na organização e na articulação dos pontos de acesso a fim de minimizar os principais problemas de um sistema informacional: a duplicação e a dispersão informacional. Assim, apresentam-se, além dos preceitos para a construção de um vocabulário controlado em arquivos, as relações semânticas da sinonímia, hiponímia, hiperonímia e polissemia, com o intuito de ratificar a importância da compreensão de tais relações para uma elaboração mais consistente do vocabulário controlado. Essa contribuição linguística se concentra principalmente nos procedimentos de controle de termos e expressões do vocabulário de um estudo de caso feito no contexto do Sistema de Arquivos da USP – SAUSP.

Palavras-chave: Arquivos. Vocabulário controlado. Semântica. Relações Linguísticas. Sistema de Arquivos da USP – SAUSP.

## ABSTRACT

The controlled vocabulary is an important tool for information retrieval, since it intervenes in the organization and coordination of access points to minimize the major problems of an information system: duplication and dispersal informational. Thus, we present, in addition to provisions for the construction of a controlled vocabulary in archives, the semantic relations of synonymy, hyponymy, polysemy and hypernym, in order to confirm the importance of understanding these relationships for developing a more consistent vocabulary controlled. This contribution focuses mainly on linguistic control procedures terms and expressions in the vocabulary of a case study in the context of the Sistema de Arquivos da USP – SAUSP.

Keywords: Archives. Control vocabulary. Semantic. Linguistic relations. Sistema de Arquivos da USP – SAUSP.

## 1 INTRODUÇÃO

O vocabulário controlado é um importante instrumento para a recuperação de informações, já que intervém na organização e nos pontos de acesso utilizando recursos que geram maior racionalização na recuperação informacional, minimizando os principais problemas do sistema: a duplicação e a dispersão informacional.

Tais evidências confirmam a relevância do vocabulário controlado no contexto arquivístico, tanto na organização sistemática como na recuperação da informação. Além disso, e segundo Smit e Kobashi (2003, p. 16): “é um instrumento para nomear as atividades/funções gerando confiança no sistema”.

Esse instrumento documentário é um dispositivo interlocutor e de tradução entre a linguagem do público e a linguagem documentária, estabelecida pelos sistemas de recuperação da informação. Sua principal função é “otimizar os processos inerentes à transferência de informações estocados em sistemas de informação arquivísticos” (AGUIAR, 2008, p. 17).

A utilização do controle de vocabulário (processo documentário) e o vocabulário controlado (instrumento documentário), tanto na esfera dos arquivos correntes como permanentes, contribuem para uma maior consistência terminológica, legitimação das

relações orgânico-funcionais, aguçadas pelo contexto de explosão e aumento da massa informacional, exigindo um olhar mais atento às necessidades dos usuários (AGUIAR, 2008).

Todavia, mesmo com tal pertinência pragmática, o estudo do vocabulário controlado é escasso na área Arquivística, por isso é comum se recorrer à bibliografia das áreas de Biblioteconomia e Documentação. Tal fato se justifica por duas razões: primeiro que ainda é recurso recente na área (os estudos se concentram no final do século XX); e a falta de divulgação das experiências existentes, grande parte em empresas privadas, dificulta o estabelecimento de paradigmas e metodologias de utilização do recurso.

Há ainda como entrave as divergências terminológicas e a heterogeneidade. A primeira está presente em questões pontuais, como as discutidas entre as associações arquivísticas paulista e fluminense sobre o uso e aplicação de determinado termo, que não é de consenso geral. Já a segunda é fenômeno cada vez mais presente, tanto potencializado pelo aumento documental e informacional, como pelo perfil diversificado das instituições: seus objetivos, necessidades, recursos, profissionais e usuários.

Na tentativa de colaborar na resolução dos problemas mencionados acima, há o caminho da Linguística, que resgatando recursos sintagmáticos e paradigmáticos consegue estabelecer os níveis de relações entre termos, contribuindo na construção do vocabulário controlado, utilizando os preceitos da Semântica Formal. Todavia, ainda é uma aplicação restrita, mais comumente associada à área de Biblioteconomia, tendo escassa bibliografia e ainda irrisória aplicação na Arquivística. O auxílio linguístico, além de contribuir com a normatização, garantindo maior consistência na seleção dos termos no sistema de informação, permitiu-se uma representação relacional e funcional mais consistente.

Biblionline, João Pessoa, v. 7, n. 2, p. 52-62, 2011.

A utilização de tais recursos promove importantes mudanças no sistema de recuperação de informação: a redução da ambiguidade semântica, a representação mais consistente da matéria (controle de sinônimos e quase sinônimos), além de facilitar a realização de buscas mais amplas, já que possui uma estrutura com termos relacionados semanticamente.

Diante dessa perspectiva linguística e da necessidade de se verificar sua aplicação na arquivística, elegeu-se o Sistema de Arquivos da Universidade de São Paulo (SAUSP) para a realização analítica do trabalho. A escolha foi motivada por dois aspectos: facilidade de acesso às informações estruturais e funcionais; e por ser um sistema recente, ainda adaptando-se às mudanças tecnológicas e institucionais, o que permitiu maior abertura para exposição de suas ferramentas e aplicações.

## 2 OS ESTUDOS LINGÜÍSTICOS E A CIÊNCIA DA INFORMAÇÃO

Nas últimas décadas, com o crescimento exponencial da informação e da massa documental, houve a necessidade de se criar medidas de racionalização e de otimização para a recuperação informacional. Nesse contexto, Comunicação Social, Informática e Linguística, além de outras áreas, vêm contribuindo com a Ciência da Informação, no intuito de encontrar soluções operacionais (FONSECA, 2007).

A Linguística, outrora rotulada pelo seu enfoque normativo-prescritivo, associada a doutrinas gramaticais tradicionais, normalmente não é encarada em sua faceta pós-moderna, em cuja prática tem se acentuado o dinamismo de uma pluralidade epistemológica permeável ao trânsito para outras áreas do conhecimento (MUSSALIM; BENTES, 2001).

Na área da Biblioteconomia, por exemplo, o papel linguístico é mais explícito no acesso à

informação. No final da década de 70, alguns estudos se evidenciam com os trabalhos de Pottier (1974) e Fillmore (1977), a partir de estudos de casos que contribuíram para a estruturação formal de vocabulários. É da mesma época a proposta de indexação automática do PRECIS (AUSTIN; DYSTRA 1984).

No âmbito nacional, trabalhos como os de Cintra (1983), Fujita (1985) e Cunha (1989) se aproveitaram das relações paradigmáticas e sintagmáticas da Semântica para construir um vocabulário controlado. Além disso, há os estudos da Linguística Documentária com Tálamo (1994), Almeida (1995) e Lara (1997). Houve ainda, a construção do vocabulário controlado do Sistema Integrado de Bibliotecas da USP feito por Lima, Kobashi e Imperatriz (2002). Além desses, há outros trabalhos que também contribuíram para o desenvolvimento do estudo do vocabulário controlado no Brasil.

Na Arquivística, consoante uma perspectiva tardia de estudo – final da década de 80 –, os trabalhos de Smit e Kobashi (1989) propiciaram importantes subsídios para a área, resgatando iniciativas europeias, ratificando a importância do estudo do vocabulário controlado, sistematizando a sua estruturação e combinando princípios arquivísticos e lógico-semânticos. Para entendemos a perspectiva lógico-semântica devemos primeiramente compreender o conceito de campo semântico.

Na organização de um sistema de recuperação da informação, por exemplo, em que há essa estruturação semântica, Lancaster (2002) ressalta a questão da representação documental, feita conforme a seleção de textos para um levantamento do vocabulário inicial, que deve vislumbrar desde as necessidades dos usuários às políticas da instituição, tendo como critério balizador a análise do domínio. Segundo MAI (2005), a partir da análise do domínio no processo de indexação verificam-se dois tipos de

documentos: o *Document-oriented* e o *Document-centered*. O primeiro enfatiza a descrição fidedigna da entidade, orientado para a entrada do sistema, não incluindo necessariamente o usuário. Já o *document-centered* parte de termos selecionados de acordo com as necessidades do usuário, fiando-se em uma interpretação de atributos que estabelecem significados independentemente do contexto.

Pensando na estruturação propriamente dita do vocabulário, inicialmente parte-se do levantamento de termos base, contidos nas macro-proposições semânticas. Em seguida, selecionam-se os conceitos que confrontam o vocabulário base com termos técnicos e especializados. Depois são feitas as inserções em classes, com intuito de constituir o campo semântico. A última etapa, não menos importante, ocorre com a adequação do vocabulário ao universo do usuário, procurando-se identificar os tipos-padrão de usuários potenciais, suas principais características e necessidades informacionais através da análise da linguagem utilizada pelos usuários.

Na esfera do arquivo, Smit e Kobashi (1989) propõem a divisão da construção do vocabulário em dois grandes procedimentos: micro e macro. O primeiro procura controlar os termos e expressões, focalizando, portanto, aspectos linguísticos da Semântica Lexical e Formal (nominalizações, sinonímia, homonímia, notas de escopo, etc.). Já o segundo categoriza os termos gerando uma estruturação funcional e hierárquica, responsável não só por organizar o universo documental, como também lhe atribuir sentido.

### 3 ANÁLISE LINGÜÍSTICA DO VOCABULÁRIO CONTROLADO DO SAUSP

O propósito desta análise não é apontar as falhas do sistema, e sim utilizá-las para demonstrar como a ferramenta linguística pode contribuir para a organização e

sistematização do vocabulário controlado em arquivos. Além do mais, parte do objeto analisado já vem sendo modificado, muito do que aqui será tratado talvez já tenha sido reformulado ou alterado.

Como já foi mencionado anteriormente, a faceta linguística eleita para a análise é da Semântica Formal (PIETROFORTE; LOPES, 2003) em que verificaremos a relação de significado entre as palavras, que no caso arquivístico designarão funções, atividades e documentos.

A análise do vocabulário controlado do SAUSP, no que diz respeito aos aspectos semânticos, se concentrará no procedimento micro, ou seja, com ênfase nas relações entre os termos e expressões. É claro que a questão do procedimento macro será utilizada indiretamente, já que tais termos estão inseridos nas macroestruturas do plano de classificação, índice alfabético e glossário.

O objetivo da análise é perceber como o conhecimento e o entendimento das relações linguísticas entre os termos, em especial na esfera semântica, contribuem na construção de vocabulários controlados. Certamente que se devem priorizar as necessidades do usuário da informação, porém tal conhecimento aliado a uma percepção linguística do contexto informacional consolida a estrutura e garante maior consistência e coerência na recuperação da informação.

As relações semânticas eleitas para análise são a sinonímia, a homonímia, a hiponímia, a hiperonímia e a problemática da polissemia. Cada uma será tratada separadamente e com aplicações práticas em mais de um instrumento arquivístico.

#### *Sinonímia*

É preciso retomar a noção de sinônimo e quase sinônimo. Enquanto o primeiro trata de termos diferentes, mas que são totalmente intercambiáveis num determinado contexto, o outro trata de palavras com o significado bem

próximo, mas diferenciadas operacionalmente.

A situação sinonímica é bem comum na Arquivística, por ter que conciliar a terminologia da área, a linguagem de uso local e a legislação. Portanto, para facilitar a recuperação da informação e garantir maior consistência na resposta, usam-se algumas estratégias, como a especificação com termos compostos, as notas de escopo e as remissivas.

A sigla, caso particular de acrografia, é utilizada de forma diversa no contexto do SAUSP (CÂMARA, 2000). No Plano de classificação encontramos ao menos três tipos de usos das siglas: primeiro quando aparece sozinha sem nenhum tipo de referência ao seu significado por extenso, como é caso da sigla PAE (Programa de Assistência ao Estudante); segundo, quando a sigla indica uma ideia mais genérica que é completada na sequência por outro termo especificador, como é caso do programa Programa de Educação Tutorial (PET), especificado posteriormente por “Bolsa”; e na terceira aparição temos o oposto, um termo mais genérico e a sigla com valor de especificação, como é o caso das “Relações de atividades de grupos PET/CAPEs”.

No Glossário de Espécies/Formatos e Tipos Documentais (1997) temos apenas duas ocorrências de siglas: no caso de TCC, colocado juntamente com sua designação por extenso - Trabalho de conclusão de curso - e o Informativo PROTEOS, sigla obviamente já consolidada. Já no Índice alfabético do Plano de Classificação, para facilitar a recuperação informacional, há também a utilização de siglas, só que sem sua significação por extenso, por serem expressões consolidadas no sistema. Alguns exemplos: “Avaliação CAPEs”, “Acumulação de INSS”, “CNPq”, etc.

Outro tipo de ocorrência sinonímica é quando um termo mais genérico, dentro de um contexto, é colocado como sinônimo de outro

termo mais específico. No Glossário de Espécies/Formatos e tipos documentais há o exemplo de “artigo” e “*paper*”. Enquanto o conceito de artigo refere-se a um “trabalho escrito técnico ou científico publicado ou a ser publicado, em geral resultante de pesquisa ou de reflexões”; o *paper* é definido como “trabalho científico resultado de pesquisa, reflexão ou análise”. Logo, notamos a maior especificidade deste em relação à generalização do anterior. No Plano de classificação há o exemplo da categoria “Rede Aluno”, mais genérica, que pode ser retomada mais especificadamente, tanto no Sistema Júpiter (graduação), como no Sistema Fênix (pós-graduação).

Outro exemplo de sinônimos pode ser observado na situação de termos preferidos pelo sistema, que podem até mostrar mais de uma expressão no índice ou até no Plano de Classificação, porém para efeitos de consistência, existe a necessidade de se eleger um termo preferido. Alguns exemplos que podemos observar no índice são: “Água pluvial”, expressão mais genérica, que engloba o termo eleito pelo sistema “Água e esgoto”; “Ambulatório”, que remete aos termos “atendimento médico” e “atendimento médico-veterinário”; ou até a sigla “CIPA”, que indica o termo preferencial “acidente de trabalho”.

No Glossário de Formatos e Espécies, temos vários casos, como por exemplo, o termo “Pedido” que vem acompanhado da designação USE. O termo “use” é utilizado nas relações de equivalência, quando há a seleção de um termo preferido para representar um conceito. Nesse caso indica a predileção no sistema para o uso do termo “solicitação” ao invés de “pedido”.

Quanto à questão dos termos estrangeiros e traduzidos, percebemos que há o padrão de se utilizar os termos estrangeiros sem tradução já consolidados pelo uso, como nos casos de *Layout*, *Hardware*, *Software*, etc. O único caso, no Plano de Classificação, em que

o termo em outro idioma acompanha sua tradução, numa mesma categoria é no G60702 “*As built*”- execução.

No Glossário de Espécies e Formatos ocorre a utilização de remissivas para os termos estrangeiros, no entanto há situações em que esses são os eleitos pelo sistema. Um exemplo é o caso do termo “estudo”, em que há a remissiva indicando tanto o uso do termo “*paper*” como do termo “texto”.

Há ainda os casos de termos usuais do sistema, muitas vezes de cunho mais generalista, distintos dos termos adotados pela legislação. Um exemplo também no Glossário de espécies e formatos é “documento normativo”, mais genérico, versus o termo corretamente usado pela legislação que é “ato normativo”.

Para terminar a questão da sinonímia podemos citar o caso dos Conceitos Básicos de protocolo e arquivo (SAUSP, 1997), que trazem termos com proximidade de significado, mas operacionalmente distintos, os quase-sinônimos. No que tange a esfera dos protocolos temos quatro termos: anexar, entranhamento, juntada e apensamento. No caso dos termos “anexar” e “entranhamento”, há remissiva para juntada; que é definida da seguinte maneira:

Ato de anexar formalmente um documento ou folha contendo informações ou despachos a um protocolado ou processo. Os documentos são inseridos no processo obedecendo à ordem cronológica de sua apresentação (SAUSP, 1997, p. 5).

Já o termo “apensamento” é definido da seguinte forma:

Ato de anexar um processo a outro, quando este deve servir de elementos elucidativos ou subsidiários para instrução daquele, continuando ambos com existência própria e independente. É, portanto, uma união de processos em caráter temporário (SAUSP, 1997, p. 5).

Diante das definições verificamos que mesmo com a nítida proximidade dos termos, cada qual traz um traço distintivo de significado. Enquanto a ideia de anexar está mais ligada à incorporação de um protocolado em um processo, o entranhamento refere-se à ação de anexar um documento a um processo. Essa singularidade de cada conceito também pode se estender a seus antônimos: “desanexar”, “desapensamento” e “desentranhamento”. Sendo que este último refere-se tanto ao oposto de entranhamento como de juntada que não possui um antônimo perfeito.

Aproveitando a menção aos antônimos, é ilustrativo ressaltar um exemplo dessa relação semântica no Plano de Classificação. Na classe de “Gestão de bens imóveis, espaços e instalações comuns”, a expressão “Água e Esgoto” - também recuperada no índice como “água pluvial” - reúne termos semanticamente opostos, mas que acabam neste contexto tendo uma referência comum, por conta de seu uso e aplicação.

#### 4 HOMONÍMIA

Os homônimos dizem respeito a uma mesma palavra ou expressão, mas que correspondem a atividades diferentes. Nesse caso o significativo é o mesmo, e é o significado que se altera. Um exemplo clássico são os homônimos nas listas onomásticas, problema facilmente resolvido quando se acrescentam especificadores.

Na busca do sistema PROTEOS, por exemplo, caso o nome do interessado seja igual a outro, há a possibilidade de se refinar a busca: por assunto, documento base (documento inicial que servirá de referência formal para a abertura de um processo protocolado), data de abertura do processo, setor de abertura ou mesmo o título do processo. No entanto, há casos só resolvidos pelo contexto. Um exemplo é o termo agenda no Glossário de Espécies e Formatos (SAUSP, 1997):

1. AGENDA - Disposição dos assuntos a serem tratados em uma reunião  
(Use PAUTA)

2. AGENDA - Livro ou caderno com datas registradas, impressas ou não, destinados a anotar compromissos, atividades, despesas, etc (SAUSP, 1997, p.3).

No primeiro caso, com o significado referente à disposição de assuntos de uma reunião, há uma indicação de uso do termo preferido pelo sistema “pauta”, ao invés de agenda. Entretanto na outra acepção, correspondente a livro ou caderno, notamos que é o termo eleito pelo sistema. Portanto fora do contexto de uso não há como especificarmos qual a designação do termo.

Ainda no Glossário de espécies temos o caso de “Boletim”, que para se tornar um tipo documental monossêmico deve ser acrescido de sua atividade geradora. Assim, se falarmos de boletim meteorológico, por exemplo, nos referimos à acepção “comprovação de um fato”. Porém, no caso de um boletim de sistema financeiro, o significado estará relacionado a uma “resenha noticiosa de operações”.

No Plano de Classificação, também há vários homônimos, mas que na maioria das vezes podem ser reconhecidos por sua classificação funcional determinada por seu código. No caso do termo “pesquisa”, por exemplo, teremos funções diferentes quando está subordinado à Marca e à Patente. Outro exemplo é o termo “Pagamento”, que aparece duas vezes, uma primeira subordinada diretamente à classe Gestão de Recursos Financeiros, com o código (N49000) e outra, subordinado à categoria “Serviço de terceiro” (N58000), determinado pelo código N58400.

No entanto tais ocorrências não são tão comuns no Plano de Classificação, já que há uma preocupação em evitar homonímia utilizando-se alguns recursos para especificar termos e expressões polissêmicas. Os

principais recursos são: uso de preposições, como o “de” em “Programa de pesquisa”; o parênteses, como no caso de “Liminar”, já que temos Liminar (autorização)” ou “Liminar (prorrogação)”); o hífen separando os termos compostos dos tipos de licença (Licença-saúde e Licença-prêmio); ou ainda outras especificações, como a função adjetiva em expressões compostas, como em “Auxílio funeral” e “Auxílio individual”.

Os termos compostos merecem um comentário a parte por terem a importante missão de nomear tipos documentais ou atividades desenvolvidas pela instituição, contribuindo muitas vezes para o esclarecimento de sinônimos e homônimos, evitando a polissemia (SMIT; KOBASHI, 2003). Além disso, garantem uma maior consistência para o sistema ao permitirem uma maior aproximação do vocabulário controlado à linguagem tanto dos usuários como à consolidada pela legislação.

## 5 HIPÔNIMOS E HIPERÔNIMOS

A hiperonímia e hiponímia são fenômenos derivados das disposições hierárquicas, que no caso do arquivo fundamentam a base para o Plano de Classificação. Nele o princípio estruturador é a hierarquia das atividades da instituição, expressas em classes, subclasses, grupos e séries. Assim, há uma relação entre significados englobantes e englobados de acordo com o domínio semântico de cada termo na classificação.

Diferente do Tesouro que também usa princípios hierárquicos, mas os relativiza, oferecendo mapas parciais do todo, no Plano de Classificação o conjunto das funções/atividades são segmentados em categorias autoexcludentes, o que ao contrário do primeiro dificulta sua atualização, mesmo oferecendo uma visualização do todo numa perspectiva mais simples e próxima da estrutura funcional.

Voltando à questão da hiperonímia e hiponímia, há vários exemplos bem sucedidos no Plano de classificação. Na classe Gestão de Recursos Agropecuários temos duas categorias subordinadas diretamente a ela, designadas: “Atividade Agrícola” e “Produção Animal”, hipônimos da primeira, que nesse caso, portanto é hiperônimo das duas. Mas se verificarmos a categoria “Produção Animal” veremos que há várias subdivisões: aves, bovinos, bubalinos, coelhos, etc. Nesse nível a mesma expressão tem função de hiperônimo em relação aos termos subordinados a ela, que por sua vez são hipônimos de Produção Animal.

Com esse exemplo fica claro perceber que as relações linguísticas da hiperonímia e hiponímia são relativas, já que dependem do termo que se elege como referencial na cadeia hierárquica. No entanto, mesmo mudando a perspectiva para a classificação, a hierarquia entre os termos englobantes versus englobados não se altera.

Outro exemplo que gera dúvida em relação ao nível de subordinação dos termos é a categoria “Colegiado” estar no mesmo nível que “Deliberação do colegiado”, colocados no Plano de Classificação como hipônimos da classe “Administração Geral e organização da Universidade”. Ou mesmo exemplos de hiperônimos que não têm uma relação direta com seus hipônimos no Plano, como é o caso de “Seminários e Teatro”.

Para finalizar pensemos no conceito de “documento”, genericamente, considerado como um “registro de informação que independe da natureza”, o que lhe indica num primeiro momento um caráter de hiperônimo, pois irá englobar o significado de especificações de diversas naturezas, como por exemplo, no SAUSP, os sentidos de documento base, processos, protocolados, enfim elementos que são hipônimos de “documento”. Porém se pensarmos neste termo como elemento integrante do arquivo, muda-se a relação, já que seu conteúdo

sêmico corresponderá apenas a uma parte do todo que é o arquivo, incorporando nessa perspectiva o caráter hiponímico.

## 6 POLISSEMIA

Na polissemia um único significante corresponde a vários significados, contrário dos termos monossêmicos que tem um significado para cada significante. Geralmente, nos vocabulários mais técnicos, por meio de definições construídas em seus discursos, operam modificações sêmicas para transformar palavras polissêmicas em monossêmicas, evitando assim compreensões dúbias.

Na Arquivística temos o exemplo clássico que pode gerar polissemia, a palavra Arquivo. Segundo os conceitos básicos enunciados pelo SAUSP (1997), o arquivo pode ser:

1. Conjunto de documentos que, independentemente da natureza ou do suporte, são reunidos por processo de acumulação ao longo das atividades de pessoas físicas ou jurídicas, públicas ou privadas, e conservados em decorrência de seu valor.
2. Entidade administrativa responsável pela custódia, pelo processamento técnico e pela utilização dos arquivos sob sua jurisdição.
3. Edifício ou uma das partes em que são guardados os documentos.
4. Peça de mobiliário destinada à guarda de documentos (SAUSP, 1997, p. 4)

Dessa forma, o termo por si só pode designar diferentes significados, o que nos força a perceber sua funcionalidade no contexto, espaço onde realmente poderá exprimir sua conceituação sêmica. Como as quatro acepções são bem distintas, fica mais fácil perceber quando selecionamos o uso de uma ou outra em determinada situação do discurso.

Já no caso de processo e protocolados, as acepções são próximas e às vezes podem gerar confusões. No caso de processo, existe tanto a designação de “um conjunto de documentos reunidos em capa especial, que

organicamente são acumulados no decorrer de uma ação administrativa ou judiciária”, como “o número de protocolo que registra o 1º documento com o qual o processo é aberto”. Veja que diferente de “arquivo”, o termo processo agrega designações que são usadas em contextos bem próximos e não excludentes, facilitando uma possível confusão.

No Glossário de Espécies e Formatos há também vários exemplos de termos potencialmente polissêmicos, aqui vamos comentar os casos de Agenda e bilhete. O termo bilhete pode designar tanto uma comunicação verbal em 3ª pessoa, acepção mais genérica, o que pode indicar uma variedade de tipos de documentos e suportes, ou um comprovante de pagamento, como por exemplo, o Bilhete de Seguro obrigatório. Quando há só um termo fica difícil saber o significado, mas com o auxílio do contexto e de um especificador, no caso “de seguro obrigatório”, acabam-se as dúvidas, ou mesmo se considerarmos o âmbito da Diplomática que os diferenciaria pela tipologia documental.

Já no exemplo de agenda, o termo terá duas acepções próximas que mesmo com o especificador podem trazer dúvidas. Numa primeira acepção é uma “disposição dos assuntos a serem tratados em uma reunião”, e na outra, “livro ou caderno com datas registradas, impressas ou não, destinados a anotar compromissos, atividades, despesas”. Se pegarmos, por exemplo, a especificação “Agenda mensal” ou “Agenda de consulta”, percebemos que tanto uma como a outra cabem nas duas acepções.

Uma estratégia para evitar a polissemia no Plano de Classificação é utilizando elementos de especificação. O termo “Bolsa”, por exemplo, que está ligado a várias categorias no plano, e que, portanto sozinho pode gerar dúvidas, é complementado por qualificadores, geralmente um substantivo, como no caso de “Bolsa moradia”; ou até com o auxílio de



preposição, como o “de” em “Bolsa de pesquisa”.

Outra maneira de solucionar problemas de polissemia, ou até de ambiguidade é utilizando as notas de escopo, usadas para tratar dos fatos linguísticos em que a interpretação de uma expressão depende de outra. As notas de escopo “explicitam amplitude ou entendimento atribuído a um conceito”, enquanto as notas de uso “voltam-se para uma explicitação de recomendações práticas de uso do termo” (SMIT, KOBASHI, 2003).

No Plano de classificação temos vários exemplos de notas de uso, como em “Bolsa de Pós-graduação”, que traz a nota: “documentos gerais, os demais procurar pelo nome da agência”; e nota de escopo, como em “Matrícula”, que traz a seguinte nota: “inclui reativação, trancamento, cancelamento”.

A polissemia é uma manifestação linguística muito comum na linguagem humana, já que os signos têm um caráter arbitrário por natureza, pois só ganham valor conceitual nas relações com os outros signos. Mesmo diferentes, no âmbito do vocabulário controlado é indispensável que tentemos aproximar os termos o máximo possível a um valor monossêmico, para que nosso instrumento de gestão seja o mais coerente e compatível possível.

## 7 CONSIDERAÇÕES FINAIS

No vocabulário controlado, diferente da natureza polissêmica da língua, há uma preocupação em assegurar uma interpretação unívoca. Nesse caso, a informação, ao contrário do sentido construído no discurso, deve estar o mais próxima possível do caráter monossêmico, evitando, portanto, interpretações ambíguas.

Mesmo o vocabulário controlado sendo elaborado a partir do *corpus* de um domínio,

referente a uma área especializada ou atividade, a sua constituição deve levar em conta o contexto organizacional, incorporando não só as relações hierárquicas e funcionais, mas as particularidades e usos da comunidade a que ele se destina.

Assim, a noção de estrutura linguística, base para a construção de qualquer sistema vocabular, é um ponto importante para se organizar os elementos. Contudo, a busca pela padronização persistentemente se depara com a relatividade. Sobre essa questão, Lara (2008) completa:

Controlar o vocabulário não significa apenas padronizar formas de entrada (significante) ou registrar o sinônimo; significa operar sobre a significação (significante e significado) dos termos; o que depende inteiramente de delimitações de sentido conferidas pela rede de relações lógico-semânticas amparadas por definições (LARA, 2008, p. 4).

As relações lógico-semânticas, explicitadas aqui no âmbito do arquivo, têm por base as teorias semânticas, cujos enfoques diferentes trazem em seu bojo a abordagem relacional entre significante, significado e referente extralinguístico. Dessa forma, notamos que o papel da Linguística para construir vocabulários controlados em arquivos é funcionar como uma instância de mediação entre o mundo e a consciência humana.

Quando percebemos relações de equivalência, oposições, associações ou hierarquização, na verdade estamos lidando com representações de um determinado domínio. Pensar no isomorfismo de sinônimos e antônimos, ou mesmo a organização hierárquica de vocábulos, que tanto pode se centrar na inclusão de uma herança de propriedades, quanto numa relação associativa, é utilizar a mediação semântica para estabelecer as nuances do significado.

No SAUSP, pudemos observar tais nuances de significado a partir das relações de sinônimos, homônimos, hiperônimos, hipônimos e

polissemia. As diferentes possibilidades de sinônimos- siglas, estrangeirismo, termos genéricos e específicos, além de especificações da legislação- nos permitiram perceber que a sinonímia não se limita à equivalência, já que pode tangenciar significados afins para se compor. Nessa mesma perspectiva relativa estão os hiperônimos e hipônimos, que só se constituem de forma imbricada e relacional, como no exemplo citado no capítulo anterior sobre documento. Pode tanto ser hiperônimo, caso o consideremos genericamente como registro da informação que possui distintos formatos e tipologias, quanto hipônimo, caso o encaremos como parte integrante do Arquivo.

Dessa forma, notamos que a Semântica contribui principalmente nas operações micro, pois oferece recursos que nos auxiliam na percepção das relações entre os elementos. Porém, como foi dito anteriormente, não devemos ignorar o contexto organizacional simplesmente ordenando conceitos baseados numa visão de mundo.

Portanto, conhecer as diferentes aplicações das relações lógico-semânticas contribui de fato para melhor se arquitetar um vocabulário controlado no contexto arquivístico. No entanto há situações do nível estrutural, funcional e organizacional que não são resolvidas simplesmente com o conhecimento das teorias linguísticas. Logo, é a integração de todos esses elementos que permitirá uma construção de vocabulário mais consistente e coerente.

## REFERÊNCIAS

AGUIAR, F. **O controle do vocabulário como dispositivo para organização e tratamento e recuperação da informação arquivística**. 2008 Dissertação (Mestrado em Ciência da Informação) PUC, Campinas, 2008.

AUSTIN, D.; DYSTRA, M. **PRECIS: A Manual of Concept Analysis and Subject Indexing**. United Kingdom: British Library, 1984.

CÂMARA, J. M. **Dicionário de Linguística e Gramática**. 21. ed. Petrópolis: Vozes, 2000.

CINTRA, A. M. et. al. **Para entender as linguagens documentárias**. São Paulo: Polis: APB, 1994. (Coleções Palavra Chave, 4).

FONSECA, M. O. **Arquivologia e Ciência da Informação**. Rio de Janeiro: FGV, 2007.

KOBASHI, N. Y.; LIMA, V. M. A.; IMPERATRIZ, I. M. de. Vocabulário controlado USP: desenvolvimento, implantação e gerenciamento. In: CONGRESSO INTERNACIONAL DE ARQUIVOS, BIBLIOTECAS, CENTROS DE DOCUMENTAÇÃO E MUSEUS. **Anais...**São Paulo: Imprensa Oficial do Estado, 2002.

LANCASTER, F. W. **El control del vocabulario en la recuperación de información**. 2. ed. València: Universitat, 2002.

LARA, M.L. G. Diferenças conceituais sobre termos e definições e aplicações na organização da linguagem documentária. **Ciência da Informação**, Brasília, v. 33, n. 2, 2001. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/266/234>> Acesso em: 23 mai. 2011.

MAI, J. E. Analysis in indexing: document and domain centered approaches. **Information Processing and Management**, n. 41, 2005. Disponível em: <<http://www.elsevier.com/locate/infoproman>> Acesso em: 13 mai. 2011.

MUSSALIM, F. BENTES, A. C. **Introdução à linguística: domínios e fronteiras**. São Paulo: Cortez, 2001.

PIETROFORTE, A.; LOPES, I. A semântica lexical. In: FIORIN, J.L. (Org.) **Introdução à Linguística: princípios teóricos**. 2. ed. São Paulo: Editora Contexto, 2003.

SMIT, J.W; KOBASHI, N.Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. Arquivo do Estado; Imprensa Oficial do Estado de São Paulo: São Paulo, 2003. (Como fazer. vol.10).

SAUSP. Sistema de Arquivos da Universidade de São Paulo. **Glossário de Espécies/Formatos e Tipos Documentais**. São Paulo: Universidade de São Paulo, 1997. Disponível em: <<http://www.usp.br/sausp/gestao/pdf/temp.php?codtabela=2&pg=0>>. Acesso em: 20 mai. 2011.

\_\_\_\_\_. **Conceitos Básicos de protocolo e arquivo**. São Paulo: Universidade de São Paulo, 1997. Disponível em: <<http://sistemas1.usp.br:8080/proteos/manual/glossario.htm>> Acesso em: 23 mai. 2011.

\_\_\_\_\_. **Plano de classificação**. São Paulo: Universidade de São Paulo, 1997. Disponível em: <<http://www.usp.br/sausp/gestao/plano/plano.php>> Acesso em: 23 mai. 2011.

---

#### Dados sobre Autoria

\*Bolsista CNPq (Bibliotecária), Mestrado em Ciência da Informação na Universidade de São Paulo, bolsista da Capes. Especialização em Arquivos (IEB-USP). Graduação em Biblioteconomia pela USP e em Letras pela Unimar.

Artigo enviado em junho de 2011 e aceito em outubro de 2011.