

PERSPECTIVAS NA CONTRIBUIÇÃO DA LINGUÍSTICA E DE ÁREAS AFINS À CIÊNCIA DA INFORMAÇÃO

RESUMO

Algumas das dificuldades surgidas no interrelacionamento da Ciência da Informação com a Linguística, além do mútuo desconhecimento, se devem a expectativas que a Linguística não pode e talvez não poderá preencher. Uma vista d'olhos sobre as atividades recentes de pesquisa na intersecção de ambas nos leva a constatar a existência hoje de um renovado interesse no processamento da linguagem natural, principalmente à vista da implementação de sistemas de diálogo homem-computador e do incremento da capacidade de memória computacional. Diante da continuada importância da Linguística para a Ciência da Informação, tenta-se estabelecer, com vistas à realidade brasileira, um recorte no universo da Linguística para a formação do cientista da informação. Sugere-se ainda a inclusão optativa de subáreas especializadas (Linguística Computacional, Pesquisa da Leitura ou outras), bem como áreas afins (Terminologia Teórica e Aplicada e Tradutologia). Seguem considerações acerca de cada subárea ou campo de pesquisa, com vistas à sua respectiva aplicabilidade, especialmente nos cursos de Ciência da Informação no Brasil.

*Ulf Gregor Baranow
Departamento de Letras e Linguística
Universidade de Brasília*

*Descritores: Linguística e Ciência da Informação;
Indexação e Linguística; Pesquisas Interdisciplinares
em Ciência da Informação; Formação do Cientista
da Informação no Brasil.*

1 - INTRODUÇÃO

O trabalho até hoje mais abrangente sobre Linguística e Ciência da Informação foi publicado sob os auspícios da Federação Internacional de Documentação (FID), e veio a lume há exatamente uma década. Uma das conclusões a que chegaram os autores diz respeito à dificuldade de se casarem as técnicas linguísticas com os objetivos da recuperação informacional. Apesar dessa convivência difícil, continua um relacionamento de conveniência entre ambas, evidenciado na literatura especializada, mais por interesse da Ciência da Informação do que propriamente da Linguística. Pelo menos em nosso meio, esta última tem direcionado as suas potencialidades aplicativas ao Ensino de Línguas, e mais recentemente, a problemas específicos na área

da Educação, como a alfabetização, aprendizagem e aperfeiçoamento da leitura etc.

Em nível internacional, após algumas décadas de intensa efervescência teórica, caracterizada por uma forte tendência à matematização, a Linguística parece estar voltando para águas mais tranquilas, atualmente. Abandonou-se a pretensão de a Linguística vir a ser o carro-chefe das Ciências Humanas, difundida mesmo em áreas vizinhas, onde dela se esperavam soluções ou, pelo menos, modelos explicativos para problemas próprios. Apesar de expectativas em contrário, até hoje não chegamos a uma Teoria Integrada do fenômeno linguístico, a um tipo de conhecimento básico pacificamente aceito, como ocorre em outros campos do saber humano (por exemplo, na Física). A Linguística

continua a apresentar-se com modelos teóricos os mais diversos, com um pluralismo metodológico* algo desconcertante para quem dela espera soluções viáveis para problemas concretos, especialmente da parte de cientistas da informação.

Por outro lado, observa-se atualmente um redimensionamento de objetivos entre os lingüistas, bem como a escolha de universos de pesquisa mais limitados. Ao mesmo tempo vêm-se desenvolvendo as perspectivas extralingüísticas, objetivando explicar o fenômeno lingüístico no seu contexto pragmático, sociolingüístico, psicolingüístico etc. Continuam a desenvolver-se algumas áreas essencialmente aplicativas na Lingüística e áreas afins, como a Lingüística Computacional, a Terminologia Teórica e Aplicada, a Tradutologia, a Pesquisa da Leitura e outras, com possibilidades de interface com a Ciência da Informação.

Segundo Damerau², as contribuições da Lingüística para a Ciência da Informação teriam sido menores do que aquelas provenientes de pesquisas de Inteligência Artificial e Psicologia Cognitiva. Entretanto, o mesmo autor registra, também, a grande mobilidade e as potencialidades dos estudos lingüísticos com vistas a aplicações na Ciência da Informação.

Além da já tradicional interface entre Lingüística e Ciência da Informação, abrangendo a Morfologia, a Sintaxe e a Semântica, parece que existem hoje novas possibilidades de pesquisas conjuntas, que deveriam ser objeto de um cuidadoso exame crítico. Ao mesmo tempo, se coloca a questão de uma eventual inclusão curricular de áreas de interface em cursos que se destinam à formação dos futuros cientistas da informação.

2 - INDEXAÇÃO E LINGÜÍSTICA

Antes de abordar algumas possibilidades e potencialidades de interface da Lingüística e da Ciência da Informação, gostaríamos de passar uma vista d'olhos no estágio atual de desenvolvimento da indexação automática — campo mais sensível às perspectivas interdisciplinares. Julgamos que

* Esse pluralismo teórico e metodológico, às vezes lamentado por lingüistas ciosos de uma vizinhança mais próxima com as chamadas Ciências Exatas, existe também em outras subáreas das Ciências Humanas e Sociais, como por exemplo, na Sociologia, na Antropologia, na Economia etc.

isto seja útil, à vista de novos desenvolvimentos nos centros avançados, envolvendo também a interseção da Lingüística com a Ciência da Informação.

Travis e Fidel³ fazem um minucioso relato da atual situação da área, no qual nos basearemos principalmente, devido às nossas crônicas dificuldades de acesso direto à vasta literatura em questão. Apesar de a área ter-se desenvolvido de modo inegável nos EE.UU., dever-se-ia acompanhar também o que se realiza em outros países, notadamente na França, na Grã-Bretanha e na Alemanha. Os interessantes progressos registrados na URSS vem sendo recenseados insuficientemente, mesmo em publicações especializadas americanas. Ultimamente, também o Japão vem-se destacando mais e mais nesta área, em função do desenvolvimento de software de sua poderosa indústria de computadores.

Ainda em 1978, no 399 Congresso da FID, Sparck Jones chamava a atenção para o progresso dos sistemas em linha, que estavam inaugurando uma nova fase na recuperação da informação. Esses sistemas, ainda não totalmente automatizados, utilizavam-se em muitos casos de uma indexação tradicional, isto é, manual. Naquele ano estava-se tão longe de sistemas de alta qualidade inteiramente automatizados, como nos quinze anos anteriores. E desde então, a indexação automática continua a debater-se com problemas principalmente de ordem lingüística para a sua efetiva implementação.

Recordemo-nos que as pesquisas de indexação automática visavam, de início, a imitação ou simulação de procedimentos manuais. Numa determinada linha de pesquisa predominava a análise do texto completo a ser recuperado, valendo-se os indexadores de processos e recursos fornecidos pela própria Lingüística. Numa outra linha de pesquisa se visava a aplicação de recursos e métodos da Estatística. Muitos projetos procuravam integrar ambas as linhas de pesquisas. Posteriormente, a primeira linha de estudos passou a preocupar predominantemente os pesquisadores em Inteligência Artificial. Constatamos aqui uma nova ênfase no uso da linguagem natural como meio mais adequado nos sistemas de diálogo homem-computador.

As pesquisas de indexação automática dependem, naturalmente, das disponibilidades e configurações de hardware e software, oferecidas pelas indústrias de computadores. O processamento computacional de dados lingüísticos, por sua vez, interessa a

pesquisadores oriundos de áreas as mais diversas, embora parcialmente afins, como a Informática, a Ciência da Informação, a Lingüística Computacional (ou "Informática Lingüística") e a Pesquisa de Inteligência Artificial. Neste contexto, a área científica em sentido tradicional não é relevante. Frequentemente, o interesse dos pesquisadores é centrado sobre determinados projetos e soluções, independentemente da "área" ou "subárea" em que se tenham originado.

Ao observar a evolução mais recente em Ciência da Informação, constata-se também um progressivo apagamento das fronteiras entre indexação propriamente dita e busca da informação. Como ficou demonstrado de modo evidente nos últimos vinte anos, ambas não podem mais ser separadas na teoria e na prática. Mas foi sobretudo a implantação dos sistemas de diálogo homem-computador que veio condicionar e acelerar essa fusão. Desse modo, as pesquisas mais em evidência atualmente têm como objetivos: a) o controle terminológico, b) a compatibilidade de linguagens de indexação, c) a ponderação de termos à vista de sua relevância informacional.

Enfatiza-se ainda a complementaridade de diferentes técnicas de recuperação, depois de um conturbado período de discussões na área: na primeira metade de nosso século, a discussão girava em torno de classificação vs. indexação, em seguida deslocou-se para a indexação de citações vs. indexação por assunto e, mais recentemente, a discussão vem girando em torno da busca direta em textos (free-text searching) vs. vocabulários controlados. Isto não significa que uma técnica tenha sido substituída por outra; todas continuam a coexistir, incluindo-se a indexação automática.

Na prática, a indexação vem sendo influenciada intensamente pela automação, enquanto que em nível teórico se critica a falta de uma "Teoria Integrada da Indexação". Uma teoria desse tipo também teria de esclarecer explicitamente as relações entre a Lingüística e a Indexação, e procurar incorporar em seu bojo aquelas partes da Lingüística que lhe são relevantes. Enquanto isso não se tornar realidade, continuamos com as soluções ad hoc, lingüísticas, para resolver problemas de indexação.

Em sua resenha de 1982, Travis e Fidei^B distinguem três tipos de sistemas na produção automática de índices e resumos: 1) sistemas de indexação baseados em dicionários, 2) sistemas de indexação apoiados em recursos estatísticos e 3) produção automática de resumos. De um modo geral, poucas

inovações têm sido descritas na literatura corrente, sendo que a maior parte dos sistemas de indexação automática já vem funcionando há mais tempo, ou seja antes de 1977. A difusão de outros sistemas não teria sido relatada em revistas especializadas, por não acrescentar nada de novo ao estágio de conhecimentos na área em foco.

Nos sistemas de indexação automática baseados em dicionários, são atribuídas categorias gramaticais às palavras que ocorrem num texto. Podem ser categorias usuais como nome, verbo, adjetivo ou outras categorias criadas especialmente para o programa a que se destinam. Por meio dessas categorias se efetuam as análises sintáticas que conduzem à extração de frases relevantes para a recuperação da informação.

Há pesquisas em curso, sobre indexação automática, que aproveitam as noções lingüísticas de tema-remata ou tópico-comentário que possibilitam uma aproximação da perspectiva funcional da frase. Essas noções já haviam sido desenvolvidas pela Escola de Praga e podem hoje ser consideradas "universais lingüísticos".

Ao elaborar as noções lingüísticas de tema-remata, partiu-se do pressuposto de que existe uma hierarquia informacional, tanto na frase e no período, como no texto. Por um lado, há aquilo que já se conhece, por assim dizer, o ponto de partida da comunicação ou a base temática do contexto comunicativo. Isto pode vir expresso por uma frase simples ou por um período completo, por unidades menores ou maiores. O remata por sua vez, é o que se acrescenta ao já conhecido; é a informação mais relevante no ato comunicativo. Destarte, as unidades de um texto se apresentariam ordenadas hierarquicamente, em função do tema-remata.

Existe uma problemática análoga, muito familiar para o indexador: o primeiro passo na indexação é determinar do que trata um determinado documento. É o que poderíamos chamar de 'concernência' (do verbo concernir, adj. concernente), termo mais adequado ao vernáculo do que o esdrúxulo anglicismo 'sobrecidade' (aboutness). Nos últimos anos, a perspectiva funcional da frase tem sido aplicada a unidades maiores de texto, tanto em pesquisas americanas quanto em pesquisas européias.

Não traremos maiores detalhes a respeito das intensas atividades de pesquisa desenvolvidas nos últimos anos na indexação automática, que se fundamentam

em métodos estatísticos. Um projeto brasileiro recente tem sido relatado por Robredo . Com vistas às possíveis interfaces com a Lingüística, é preciso lembrar a intensificação na aplicação de métodos quantitativos e estatísticos na Sociolingüística dos últimos anos.

Na produção automática de resumos vêm-se aplicando também soluções desenvolvidas a partir do modelo lingüístico de tema-remata que mencionamos antes. Nesses estudos tenta-se estabelecer primeiramente algoritmos capazes de identificar as partes funcionais em resumos, por meio de frases-chave. Posteriormente, esses algoritmos seriam aplicados a textos inteiros. Não houve neste campo de pesquisa maiores avanços nos últimos anos.

Os múltiplos problemas na pesquisa da indexação automática, dos quais mencionamos apenas alguns, levam-nos à questão da contribuição dos estudos lingüísticos à própria Ciência da Informação.

3 - ASPECTOS DA LINGÜÍSTICA RELEVANTES PARA A CIÊNCIA DA INFORMAÇÃO*

O recorte a ser feito no universo da Lingüística para a formação do cientista da informação deverá pautar-se pelos aspectos aplicativos exigidos em sua futura área profissional. Quanto ao enfoque didático, sugerimos partir-se da problematização de fenômenos lingüísticos relevantes para a Ciência da Informação. É desaconselhável partir-se de soluções prontas, de teorias estabelecidas como se fossem dogmas, tão a gosto de nosso ensino de Lingüística nas universidades brasileiras.

Além de se introduzir os conceitos gerais vigentes na Lingüística, sugerimos a inclusão em um curso de Lingüística para cientistas da informação, de noções de Sintaxe e de Semântica. Mais adiante, haveria um capítulo sobre teorias lingüísticas e as principais correntes no panorama teórico atual. Por outro lado, ficariam excluídos alguns capítulos que, embora essenciais na formação de um lingüista, eventualmente, não apresentam interesse para o cientista da informação. Trata-se de toda a extensa matéria da Fonética e Fonologia, bem como a Lingüística Histórica ou Diacrônica. Sugerimos

Sobre as inter-relações da Ciência da Informação com a Lingüística já existe uma série de bons artigos em língua portuguesa, onde se tenta oferecer uma visão geral da problemática ou onde se abordam problemas específicos. Deixamos de mencioná-los por falta de espaço. Há contudo, necessidade de mais pesquisas de interface, voltadas sobre a nossa realidade lingüística e documentária no Brasil.

porém, a inclusão de Noções de Lexicografia e Lexicologia que — por razões de evolução teórica na área — em geral não constam mais de manuais modernos de Lingüística. Conviria incluir também noções de Psicolingüística e Sociolingüística, sempre com vistas às eventuais interfaces com a Ciência da Informação. Por último, deverá constar um capítulo referente a "Línguas naturais e linguagens documentárias". Na diferenciação de ambas é preciso, contudo, não se perder de vista os pontos comuns, que possibilitariam também uma melhor apreciação conjunta a partir do enfoque estritamente lingüístico.

Esses "Elementos de Lingüística" introdutórios, que poderiam ser ministrados ao longo de um (01) semestre de dois ou três créditos, forneceriam as bases teórico-práticas sobre as quais se assentariam aqueles estudos lingüísticos diretamente relacionados ou relacionáveis à Ciência da Informação. Esses últimos poderiam subordinar-se ao título genérico de "Tópicos Especiais de Lingüística" ou, mais abrangente, "Tópicos Especiais em Ciências da Linguagem" (até quatro créditos), oferecendo-se optativamente:

- Sintaxe da Língua Portuguesa
- Semântica (e Pragmática) do Português
- Lingüística Computacional
- Terminologia Teórica e Aplicada
- Ciência da Tradução ou Tradutologia
- Outras subáreas ou áreas de pesquisa inovadoras.

Nas ofertas é mister não se perder de vista o caráter instrumental dos assuntos tratados em relação à Ciência da Informação. O relacionamento com esta deverá ser feito didaticamente, e a priori, pelo professor, não se deixando exclusivamente ao aluno a tarefa de estabelecer as interseções e interfaces.

3.1 O Significado e seu contexto: Semântica e Pragmática

Ao examinar a evolução dos estudos relativos à interface da Semântica e da Ciência da Informação, parece continuar válido hoje o pessimismo de Sparck Jones e Kay de há dez anos atrás, quando afirmavam ter sido o aprimoramento teórico neste campo menos alvissareiro do que se supunha. Estudos mais recentes nesta área de interface, aparentemente não vieram ao encontro de previsões mais otimistas.

Em sua resenha interdisciplinar, Walker¹³ lembrou ser essencial a informação sintática relativa aos constituintes da frase, mas que, por outro lado, não é possível resolver problemas sintáticos sem recorrer-se à Semântica. Obviamente, se uma frase está correta ou não, não pode ser decidido à revelia de seu significado. O conhecido exemplo de Noam Chomsky "Colorless green ideas sleep furiously" está correto sob o ponto de vista sintático, mas somente sob o ponto de vista semântico se pode avaliar, se a frase tem um significado ou não. Na melhor das hipóteses, a frase poderia ser interpretada metaforicamente, mas neste caso teríamos novamente apelado ao seu significado.

No processamento automático da linguagem natural tenta-se hoje estabelecer vínculos entre palavras e objetos, propriedades e fatos de um determinado universo previamente caracterizado. A aparente insolubilidade da problemática do significado levou especialistas em estudos interdisciplinares à tentativa de transpor os limites impostos pelos aspectos apenas lingüísticos da linguagem natural, passando a incluir na assim chamada representação computacional do significado — RCS (ingl. computer meaning representation) dados extra lingüísticos (extra-semânticos), a fim de facilitar a recuperação da informação.

Incorporaram-se assim em alguns modelos computacionais, destinados à recuperação da informação, além das instruções sintático-semânticas, outras que dizem respeito ao universo em questão. A RCS pressupõe a incorporação de um imenso lastro informativo, que não vem explícito em cada frase. Tal informação incorporada permitiria, por exemplo, a um hipotético sistema de informação responder a uma pergunta como: "O João caiu?", enquadrando a resposta ou num contexto relativo à força da gravidade ou num contexto, digamos, político. Note-se que muitas RCS foram elaboradas independentemente da linguagem natural ou, pelo menos, não diretamente relacionadas aos problemas levantados pela mesma. A quantidade desses sistemas hoje, nos EE.UU., é ainda incipiente, perfazendo aproximadamente 5% do que se necessitaria realmente.

De um modo geral, houve nos últimos anos, de acordo com observadores da literatura especializada, um certo progresso quanto ao emprego da Semântica ao lado da Sintaxe, nos programas que visam ao diálogo homem-computador em áreas de conhecimento limitadas. Alguns projetos na

área* introduziram "redes semânticas", representando objetos/situações e seu inter-relacionamento.

Quanto ao ensino de Semântica num Curso de Ciência da Informação, deveria ser questionada, em primeiro lugar, a conveniência ou não de se aprofundar o estudo através de uma segunda fase que acima chamamos de "Tópicos Especiais em Ciências da Linguagem". Com as noções básicas oferecidas num curso introdutório de Lingüística, evidentemente, não será possível sequer de acompanhar pela leitura o que se está realizando em centros avançados, relativamente a pesquisas de interface. Por isso, sugerimos um levantamento prévio do que se vem fazendo em recuperação da informação em nível internacional, com auxílio da Semântica, bem como sua viabilidade, aplicabilidade ou conveniência no contexto brasileiro. Só então poderiam ser traçadas diretrizes para estabelecer conteúdo e metodologia de Estudos Semânticos num Curso de Ciência da Informação. Entretanto, independentemente da questão curricular e didática, deveriam aproximar-se semanticistas e cientistas da informação, além de especialistas em Informática, para decidirem sobre possibilidades de projetos interdisciplinares, voltados sobre a língua portuguesa.

Nos últimos anos vêm-se desenvolvendo com intensidade a Pragmática, a Análise do Discurso e a chamada Lingüística do Texto ou Textologia. Na verdade, trata-se de um conjunto de subáreas das Ciências da Linguagem ainda em formação. Enquanto os americanos preferem a Análise do Discurso (discourse analysis), os ingleses e alemães se dedicam à Lingüística do Texto (textlinguistics), subáreas que, além de diferenciações teóricas, ainda não podem ser consideradas consolidadas em termos epistemológicos.

A Pragmática ocupa-se da linguagem além de seus aspectos sintáticos e semânticos intrafrasais, analisando a sua utilização (intenções e motivações dos interlocutores, por exemplo) e levando em conta o contexto extra-lingüístico da comunicação. Nessas novas áreas de pesquisa que se vêm delineando na Lingüística, a linguagem é analisada no seu contexto comunicativo ou seja, numa dimensão funcional. Retoma-se com isto uma tradição da Lingüística européia de uma temática, que as escolas estruturalistas e pós-estruturalistas americanas (com algumas exceções) haviam deixado de lado.

Veja-se entre outros WALKER, D.E. 1981 (Referências Bibliográficas).

Ao mesmo tempo, esses estudos hoje vêm recebendo fortes impulsos da Sociolinguística, que tem nos Estados Unidos, no Canadá e na Inglaterra os seus principais centros de irradiação internacional.

Os enfoques pragmáticos permitem novas soluções para problemas de ambigüidade semântica e sintática, que são de importância crucial na recuperação da informação. No afã de explicitar "como fazer coisas com palavras" (J. Searle), e pressupondo que a atividade linguística tem conseqüências e determinantes extra-linguísticas, muitos linguísticos preocupam-se atualmente com questões do tipo: quem diz o quê, a quem, quando, onde, como e por quê (who says what, to whom, when, where, how and why)¹⁷. Assim, a resposta à pergunta "Onde fica o Pão-de-Açúcar?" poderia ser, dependendo do contexto em que se insere: no Brasil; no Rio de Janeiro; na Urca; entre o Botafogo e a Praia Vermelha. Se a pergunta for: "Onde está o Pão-de-Açúcar (pão-de-açúcar)?", a resposta poderia vir a ser uma das seguintes, dependendo sempre de seu contexto extra-linguístico: atrás das nuvens (visto do Corcovado); na estante de souvenirs; no prato; na fôrma ou no molde (referindo-se ao cone de açúcar que sobrou num molde do tipo usado nos engenhos antigos). A resposta a esse tipo de perguntas sempre dependerá da habilidade do inquirido ou, no caso de um programa de recuperação informacional, de seu grau de qualidade. É claro que a ambigüidade dessa categoria de perguntas — cuja incidência em um sistema de diálogo homem-computador pode ser bastante comum — será tanto maior, quanto menor for a habilidade ou versatilidade de quem fornecerá a resposta.

As preocupações dos linguistas com os aspectos pragmáticos da linguagem nasceram da constatação de que a comunicação em linguagem natural depende de condições que vão além da linguagem propriamente dita. Dessa maneira, ao participar de um diálogo, necessita-se conhecer também o conteúdo de uma série de enunciados anteriores. Palavras soltas como "sim", "mais ou menos", "uma dúzia", "impossível", por exemplo em uma conversa telefônica entre duas outras pessoas, da qual não participamos, só adquirem significado efetivo, se a pergunta for conhecida ou a situação em que está sendo formulada. O mesmo ocorre, quando num diálogo aparecem pronomes, formulações elípticas ou formas deícticas. Os exemplos acima apenas serviram para ilustrar um sério problema que tem desdobramentos teóricos e práticos complicados na recuperação da informação. Nada mais natural, pois de se procurar subsídios em Ciência da Informação naquelas pesquisas linguísticas, onde estão sendo

problematizados esses aspectos, embora com outros objetivos.

Igualmente de interesse para a Ciência da Informação poderá ser a pesquisa linguística em torno do fenômeno das pressuposições, isto é, daquilo que é deixado implícito, não expresso, na comunicação linguística oral ou escrita e que, eventualmente, deverá ser inferido de um lastro de conhecimentos comuns às pessoas envolvidas no ato linguístico-comunicativo.

Finalmente, considerando que a Ciência da Informação continua a lidar principalmente com textos em linguagem natural, seria conveniente acompanhar os recentes progressos na Análise do Discurso, na Linguística do Texto e em campos de pesquisa afins. Como se disse acima, nesses estudos se pretende analisar por meio de metodologia rigorosamente linguística aquelas unidades que transcendem a frase e o período. Aqui interessam a constituição linguística do texto, as estratégias linguísticas de sua elaboração, suas marcas características etc. Essa área de pesquisa ainda está prescindindo de uma uniformidade de escopo e metodologia definida, razão pela qual ainda parece prematuro sugerir a sua inclusão em um curso para formar cientistas da informação. Entretanto, convém acompanhar as respectivas pesquisas, com vistas à sua aplicabilidade a sistemas de diálogo-computador e à recuperação da informação em geral. Aliás, a maior parte das pesquisas relativas ao discurso, enquanto unidade linguística maior do que a frase ou o período, situam-se na interface com os sistemas de diálogo, atualmente.

3.2 Morfologia e Sintaxe

Duas são as razões que nos levam a mencionar explicitamente a Morfologia além da Sintaxe, embora ambas estejam estreitamente ínter-relacionadas, quando analisamos a linguagem natural. Em primeiro lugar, a Morfologia, após ter sido excluída das preocupações teóricas dos linguistas pós-estruturalistas, há alguns anos novamente vem se tornando um campo amplamente pesquisado, inclusive no tocante à Língua Portuguesa. Em Linguística, a chamada "teoria lexicalista", nos últimos anos, tem ressuscitado os estudos morfológicos. A segunda razão se deve ao fato de a Morfologia ser de interesse específico para o processamento computacional da linguagem natural, justamente por ocupar-se das unidades significativas mínimas.

Em um sugestivo trabalho sobre as potencialidades de análise das estruturas lexicais com vistas à Ciência da Informação, Basílio¹⁸ mostrou como o

conhecimento de padrões morfológicos na formação de palavras pode ser relevante para a indexação. No citado trabalho, a autora trata das relações morfológicas entre verbos, adjetivos, formas nominalizadas e, naturalmente, nomes, na Língua Portuguesa. Resultados de pesquisas desse tipo sobre o Português deixam entrever importantes aplicações, tanto na indexação automática como na manual. Considerando a frequência das palavras significativas em um texto como parâmetro relevante para determinar o seu conteúdo, e possibilitando a contagem dessas palavras por meio de um programa computacional de análise morfológica, pode-se facilitar a recuperação automática da informação.

O referido trabalho é o primeiro de que temos conhecimento, surgido de um trabalho de interface no contexto brasileiro, e tendo a morfologia como objeto. A moderna abordagem morfológica nele utilizada abre uma real perspectiva de cooperação entre lingüistas e cientistas da informação.

Enfatiza-se freqüentemente em Ciência da Informação a importância de três teorias da Sintaxe — a Análise Distribucional (Harris), o Modelo Transformacional (Chomsky) e a Gramática de Casos (Fillmore). Os três modelos tiveram ampla aplicação em programas que visam à recuperação da informação; o sistema de indexação PRECIS, por exemplo, é uma aplicação da teoria de Fillmore. Poder-se-ia mencionar ainda outros modelos lingüísticos, além dos estruturalistas e transformacionalistas nesta atual fase pós-transformacional na Lingüística. Pelo menos na Europa, outros enfoques teóricos que se revelaram profícuos no passado, estão sendo retomados com vigor, por exemplo a Gramática Dependencial, em continuidade à linha preconizada pelo lingüista francês Tesnière.

As teorias lingüísticas no âmbito da Sintaxe das linguagens naturais têm como objetivo comum classificar de modo explícito os enunciados contidos em textos falados ou escritos, em corretos ou incorretos, atribuindo-lhes descrições estruturais. Quando se trata de formalizar uma gramática de modo tal a nos permitir o seu uso em computador, é necessário primeiramente um conjunto de regras, que transformam determinados símbolos em outros, de modo analítico ou sintético*. Pela síntese, que pode ser efetuada no computador, chega-se a uma cadeia final de símbolos, e, pela análise, até o símbolo de frase mais elevado na hierarquia das

Esta descrição é válida principalmente para os modelos gerativo-transformacionais.

transformações. Por meio de um desses processos, o enunciado em questão será aceito como correto ou será rejeitado pelo sistema, isto é, pela gramática explicitamente elaborada.

O processo pelo qual se tornam explícitas as relações entre as palavras que compõem uma frase ou um texto, especialmente com vistas a um programa de computador, convencionou-se chamar pelo termo emprestado ao inglês parsing; o programa que efetua essa tarefa é o parser.

Ao elaborar-se um programa computacional que opera sobre a sintaxe de uma linguagem natural, o primeiro passo será a sua descrição estrutural — ou seja, o parsing do texto. Em seguida, operam as regras de interpretação, identificando as conexões lógicas entre os elementos lingüísticos que correspondem às entradas na base de dados. Só então se processam as correspondentes operações de recuperação.

Em que pese a grande sofisticação de modelos sintáticos elaborados pela Lingüística dos últimos trinta anos, é curioso assinalar que em experiências mais recentes têm sido utilizados parsers (ou seja, programas computacionais que executam um parsing) bem mais próximos da gramática tradicional¹⁹, durante tanto tempo relegada em pesquisas nesta área.

Entretanto, os melhores resultados se obtiveram até hoje com parsers que operam sobre textos cujas estruturas gramaticais são mais simples e altamente padronizadas. Um dos alvos que se pretende atingir aqui é a geração de bases de dados, automaticamente, a partir de fontes em linguagem natural.

Uma rápida análise do estágio atual da área parece justificar que na formação do cientista da informação, a parte da Lingüística denominada Sintaxe, ou melhor, Morfosintaxe, deveria ser oferecida, com vistas à solução de problemas substanciais de indexação. É relevante que se ensine essa subárea da Lingüística em função de problemas específicos da área de formação, evitando-se a apresentação pura e simples de modelos teóricos, alheios à realidade profissional.

3.3 Lingüística Matemática e Lingüística Computacional

Trata-se na verdade de duas importantes subáreas da Lingüística que podem desempenhar um importante

papel na formação do cientista da informação, desde que ele pretenda dedicar-se à pesquisa básica e aplicada na interface da Lingüística com a Ciência da Informação.

Ao contrário do que se poderia supor, a Lingüística Matemática não trata só de métodos quantitativos (estatísticos, probabilísticos etc.) ou da descrição de fenômenos lingüísticos por meio de números naturais ou reais. Trata também, e principalmente, de métodos algébricos e da teoria dos conjuntos na construção de teorias lingüísticas. Sob este ponto de vista, pode-se falar numa crescente matematização numa vertente importante da Lingüística Moderna nos últimos trinta anos. É a época da vigência teórica de modelos estruturalistas e gerativo-transformacionais nos estudos lingüísticos.

Como vimos, constitui a linguagem humana um fenômeno empírico não-ordenado e extremamente complicado, quando colocado sob um prisma matemático-formal. Por meio de uma teoria que se constrói para explicitar esse fenômeno, o lingüista procura introduzir uma certa ordem ao universo "naturalmente" desordenado. Isto é feito por intermédio de construtos teóricos, entre os quais se podem estabelecer relações. A essas relações então, se aplicam os métodos matemáticos. O processo todo é chamado formalização. Dependerá do grau de elaboração desses construtos e de suas relações, se pudermos aplicar numa teoria um formalismo mais ou menos matematizado.

O distribucionalismo ou análise distribucional de Harris, por exemplo, fundamenta-se numa metodologia analítico-matemática. Diga-se de passagem que a análise distribucional preparou em nível teórico o advento da lingüística gerativo-transformacional de Chomsky, que foi aluno de Harris.

Cabe lembrar que o modelo gerativo-transformacional, que se tornou tão importante na Lingüística nos últimos vinte e cinco anos, também tem as suas raízes na Matemática. A propósito da terminologia: os termos **gerativo e transformacional** são usuais em Matemática, significando o termo **gerativo** nada mais do que **explícito**, diferentemente da interpretação de alguns lingüistas, que lhe atribuíram o significado de **gerar (ou produzir)**.

No presente contexto, à inclusão da Lingüística Matemática, ou mais adequadamente, de Noções de Lingüística Matemática, caberia uma função subsidiária na formação do cientista de informação.

Nesta disciplina, ele pode familiarizar-se não só com a metodologia estatística aplicada à linguagem natural, tão importante na indexação automática, como ainda adquirir o embasamento teórico necessário à compreensão de diversas teorias lingüísticas modernas. Além disso, uma escolha criteriosa de temas nesta subárea poderá servir de prolegômenos à Lingüística Computacional.

Recordemo-nos que o interesse pelo processamento computacional de dados lingüísticos inicialmente, não partiu da Lingüística, na época ainda inteiramente despreparada para tal, mas de matemáticos, lógicos e especialistas em informação. Primeiramente se elaboraram linguagens artificiais, e a pesquisa de suas propriedades formais possibilitou a sua posterior aplicação no tratamento das línguas naturais. Quando se cogitou na possibilidade de elaborar descrições do tipo de gramáticas de linguagens artificiais para linguagens naturais, começou a desenvolver-se esta área de pesquisa.

A primeira grande tentativa do tratamento computacional de dados lingüísticos resultou num imenso fracasso. Tratava-se da tradução automática, encomendada pelo governo americano*. Apesar desse saldo negativo, houve também um avanço no conhecimento, tanto em relação às possibilidades do computador na época, como em relação à natureza da linguagem natural. Descobriram-se propriedades nas linguagens naturais que até então haviam passado despercebidas a lingüistas e filólogos. Apesar de tudo, as pesquisas de Lingüística Computacional inclusive a Tradução Automática, não pararam aí. Pelo contrário, tomaram um grande surto nos centros avançados, tanto nos Estados Unidos e, principalmente, na Europa, devido a razões políticas decorrentes do Mercado Comum Europeu e órgãos supra-nacionais.

A Lingüística Computacional atinge atualmente três grandes campos de pesquisa, embora não sejam os seus domínios exclusivos, a saber: a) **Processamento automático do texto**, dedicando-se a análises morfológicas, sintáticas e semânticas, mais ou menos integradas, baseadas em gramáticas parciais ou não; possibilidades de formalização de dados lingüísticos. Produtos: concordância de textos, dicionários ou glossários automatizados, tratamento estatístico de

U.S. National Research Council. Automatic Language Processing Advisory Committee, 1966. **Language and machines; computers in translations and linguistics.** Washington, D.C, National Academy of Sciences, 1966. 124p.

textos corridos etc.— b) Recuperação automática da informação: A partir de material lingüístico armazenado em computador, sob forma de palavras, frases, textos, dicionários, bancos de dados lingüísticos em geral, procura-se obter automaticamente a informação desejada. Esse campo de pesquisas teve uma contribuição importante para o desenvolvimento de sistemas de informação. Produtos: recuperação de dados em linguagem natural, indexação automática, análise de conteúdo automatizada.— c) Sistemas automáticos de pergunta-resposta (inteligência artificial). Trata-se do diálogo homem-computador, em linguagem natural ou num subconjunto de linguagem natural. É a área que mais se vem desenvolvendo ultimamente nos centros avançados americanos e europeus²⁰

A Lingüística Computacional se encontra hoje numa interseção, na qual colaboram a Lingüística, a Lógica, a Matemática e a Informática. Devido a seu caráter interdisciplinar e, ao mesmo tempo, especializado, tentou-se introduzir em algumas universidades européias cursos de "Informática Lingüística", cujos resultados parecem ser promissores, inclusive para a Ciência da Informação.

Há uma tendência inevitável no mundo moderno industrializado de se possibilitar cada vez mais o acesso do usuário ao computador. Decorrem dessa evolução novas tarefas para o lingüista no sentido de analisar o a priori e o a posteriori na comunicação do homem com o computador.

Mesmo em sistemas modernos tão bem sucedidos como o LUNAR²¹, relacionam-se alguns dos problemas com que se debate a pesquisa em curso²²: necessidade de gramáticas mais abrangentes, capacidade de diálogo interativo, formalização mais adequada da informação semântica e regras de inferência complexas. Resumindo: ainda constituem um desafio os problemas mais gerais em linguagem natural, apesar do progresso verificado na Lingüística Computacional e no uso de procedimentos sofisticados, inspirados na análise sintática e semântica, na Pragmática e Análise do Discurso. Um passo à frente na solução desses problemas seria a inclusão do universo do usuário, isto é, seus conhecimentos, suas pressuposições e necessidades específicas. A sugestão, ainda um tanto utópica, de se oferecer ao usuário meios para fazer "várias leituras do conjunto de documentos, sempre relativas às suas necessidades pontuais", necessitará do concurso da Lingüística Computacional para a sua implementação.

Apesar da imensa utilidade da Lingüística Matemática e Computacional para a Ciência da Informação, é de se questionar, até que ponto se justificaria a inclusão de mais uma matéria num curso profissional, pós-graduado ou não, na área. Lembremo-nos que nos centros de Lingüística Computacional, o aluno se capacita ao longo de até oito semestres para tornar-se um analista de sistemas, especializado no processamento da linguagem natural. Possivelmente, num Curso de Ciência da Informação caberiam Noções de Lingüística Computacional ou noções sobre o tratamento computacional da linguagem natural, visando a facilitar o diálogo profissional entre cientistas da informação e especialistas em Informática Lingüística.

4 - ASPECTOS DE TRADUTOLOGIA OU CIÊNCIA DA TRADUÇÃO*

A Ciência da Tradução ou Tradutologia (ingl. translation Science, fr. traductologie, esp. traductología) vem-se desenvolvendo nos últimos anos em nível internacional como área de pesquisa autônoma, próxima à Lingüística, buscando integrar aspectos teóricos e aplicados do fenômeno tradutório. O objeto específico dessa área é a tradução enquanto processo e/ou resultado de um esforço humano ou artificial (por computador) para reexpressar numa língua ou linguagem B o que é oferecido numa língua ou linguagem A. Em Tradutologia não se trata portanto de produzir traduções, senão abstrair dos fenômenos e problemas específicos e individuais da tradução enquanto atividade ou produto, tentando sistematizar explicitamente seus aspectos básicos por meio de modelos teóricos. Preocupam à Ciência da Tradução questões como as seguintes: leis ou regularidades subjacentes ao processo tradutório; condições e condicionantes da equivalência interlingüística: fatores lingüísticos e extralingüísticos da tradução: possibilidades e limites da tradução; tipologia de textos à vista da tradução e métodos aplicáveis à solução de dificuldades tradutórias específicas.

Não nos referimos aqui ao interesse que o cientista da informação poderia vir a ter nesta área, considerando os problemas práticos de tradução

Associamos, via de regra, o termo "ciência" com as assim chamadas Ciências Exatas. A exemplo de outras línguas neolatinas, também em Português o campo semântico de ciência está se ampliando. Veja-se por exemplo formações como "cientista social" (social scientist), "cientista da informação" etc. Aliás, o termo Ciências (sic!) Jurídicas e Sociais já tem uma longa tradição em nosso contexto cultural-lingüístico.

com que se defronta no seu dia-a-dia, no manuseio de documentos em diversas línguas e, eventualmente, na necessidade de indexá-los em sua língua materna. O interesse poderia situar-se, tentativamente, em outro nível. Ao passarmos da linguagem natural para uma linguagem documentária, estabelecem-se relações de equivalência muito semelhantes àquelas, quando traduzimos de uma língua A para uma língua B. A explicitação dessa relação, ou seja, o processo da passagem de uma língua a outra, freqüentemente fica ao sabor das circunstâncias ou na dependência da subjetividade e cultura geral do indexador.

Ora, a equivalência entre duas línguas diferentes é um dos assuntos amplamente problematizados e refletidos em Tradutologia, podendo esta possivelmente oferecer subsídios para um melhor desempenho no processo igualmente "tradutório" do indexador. Por outro lado, o problema da equivalência surge também para o indexador, ao passar de uma linguagem documentária a para uma linguagem documentária b, bem como na elaboração e padronização de linguagens especializadas ou vocabulários controlados. De qualquer forma, gostaríamos de sugerir que sejam refletidas as possibilidades de uma aproximação entre a Tradutologia e a Ciência da Informação, em determinados campos, onde possa haver interseções.

5 - TERMINOLOGIA TEÓRICA E APLICADA

Ainda mais evidente do que a Tradutologia parece-nos a aplicabilidade da Ciência da Terminologia ou Teoria da Terminologia, ou simplesmente Terminologia Teórica e Aplicada, para os futuros cientistas da informação. Essa área de pesquisa básica e aplicada ainda não possui uma denominação universalmente aceita em língua portuguesa. Sugerimos a denominação Terminologia Teórica e Aplicada, a fim de expressar por meio do termo a duplicidade da matéria, enfatizando ao mesmo tempo o seu aspecto aplicativo.

O termo vocabulário especializado abrange certamente um âmbito mais vasto do que terminologia, uma vez que esta última pressupõe, além da especialização que o vocabulário esteja padronizado, isto é, normalizado e unívoco. Mesmo assim, feita esta ressalva, utilizaremos terminologia no presente caso, numa acepção mais ampla, incluindo também os vocabulários especializados, que prescindem de uma rigorosa padronização.

Nos trabalhos sobre Terminologia Teórica e Aplicada enfatiza-se o valor da documentação para a pesquisa

da terminologia em qualquer área do conhecimento científico ou tecnológico. Mudando de perspectiva, examinemos a seguir o valor e a aplicabilidade da Teoria Terminológica para o trabalho do cientista da informação. Sistemas de classificação em geral e tesouros em particular baseiam-se em conceitos e classes de conceitos, geralmente agrupados em ordem hierárquica. Para elaborar tais sistemas, necessita-se do levantamento prévio dos conceitos/termos ou dos objetos/coisas relevantes para a área em questão. Trata-se de uma tarefa em parte coincidente com a atividade do terminólogo. Este começa pelo levantamento lingüístico da área e do seu conteúdo, para em seguida, proceder à padronização. Esses termos padronizados, análogos aos descritores na indexação, devem cobrir uma certa área, constituindo ao mesmo tempo um vocabulário estruturado, aberto e dinâmico quanto a acréscimos e atualizações. A tarefa do indexador de determinar ou criar descritores poderia receber sugestões e impulsos efetivos a partir da Ciência da Terminologia.

As analogias entre ambas as áreas enquanto atividades práticas são evidentes. Ao proceder uma classificação terminológica numa área específica, o terminólogo introduzirá uma estruturação num universo de dados terminológicos. O produto final dessa tarefa pode apresentar-se como um dicionário sistemático ou uma base de dados terminológicos

A seguir, daremos algumas sugestões de assuntos que poderiam ser de interesse conjunto para a Terminologia Teórica e Aplicada e a Ciência da Informação, a saber:

- linguagem comum e linguagem especializada (análise das características, relacionamento entre ambas);
- conceito e termo (modelos lexicais, concatenação entre conceitos, constituintes semânticos ou sememas, estruturação e representação lingüística de sistemas conceituais;
- padronização do termo (exigências formais, processos de criação terminológica, neologia, problemas específicos como sinonímia, polissemia, homonímia);
- fundamentos da padronização terminológica;
- equivalência lingüística em geral e terminológica em especial;
- problemas terminológicos vinculados à transferência tecnológica.

As potencialidades da Terminologia Teórica e Aplicada para o cientista da informação, evidenciam-se ao examinar, por exemplo, a resenha de Travis e Fidel²⁶, Apesar da crescente disseminação de sistemas em linha, que possibilitam uma livre busca do termo desejado (free-text searching), continua imprescindível o controle dos vocabulários padronizados, por exemplo a partir de tesouros previamente estabelecidos.

Em geral, as tarefas que envolvem questões terminológicas no âmbito da Ciência da Informação têm sido tratadas de acordo com as necessidades práticas, sem diretrizes científicas explícitas. A quantidade de erros no controle terminológico tem levado alguns autores a sugerir a criação de bancos terminológicos subsidiários, que recolheriam os termos, suas definições e inter-relacionamentos. Embora seja praticamente impossível que cada sistema de informação elabore o seu banco terminológico próprio, torna-se cada vez mais necessário na documentação recorrer a bancos terminológicos já existentes. A coordenação entre tesouro e bancos terminológicos constitui um desideratum ainda longe de ser alcançado, mesmo em países que já têm uma tradição terminológica. No Brasil, a terminologia em bases científicas é ainda incipiente.

Ultimamente, se tem sugerido uma maior aproximação entre os campos de ação do terminólogo e o do indexador coordenando-se inclusive as tarefas de ambos. No Brasil, onde ainda não dispomos de bancos terminológicos de relevo, nem de uma formação institucionalizada de terminólogos, os futuros cientistas da informação parecem ser predestinados a absorver uma importante parte dessa nova área científica, que está-se firmando mais e mais internacionalmente.

A oferta de uma disciplina optativa Terminologia Teórica e Aplicada na formação dos futuros cientistas da informação e documentalistas poderia ser efetuada após minucioso exame das possibilidades de interface entre ambas as áreas.

6 - PESQUISA DA LEITURA - UM NOVO CAMPO INTERDISCIPLINAR RELEVANTE PARA A CIÊNCIA DA INFORMAÇÃO?

Vem-se constituindo, há alguns anos, um novo campo de pesquisa na Lingüística, de enfoque interdisciplinar (ou transdisciplinar) que tem como objeto a leitura, independentemente da faixa etária do leitor. Em fins de 1982, finalmente, um grupo de

trabalho sob os auspícios do INEP/MEC constituiu-se para examinar a problemática no contexto brasileiro.

Em virtude do caráter internacional do assunto, os principais impulsos para a sua análise científica têm vindo de fora, principalmente dos Estados Unidos. Embora ainda incipiente no Brasil, o campo promete uma intensificação de pesquisas nos próximos anos, uma vez que conta com recursos e a dedicação de alguns de nossos melhores lingüistas. Por isso, levando em ponderação as judiciosas observações de Anna Maria Cintra²⁸ acerca da aplicabilidade de tais estudos para a indexação, sugerimos seguir atentamente a evolução dessas pesquisas, com vistas a uma eventual aplicação e operacionalização em projetos de interface.

Enquanto no referido trabalho se faz referência à importância da leitura a partir da perspectiva do indexador, há de se considerar também o aspecto a partir do usuário. Na já referida resenha sobre a situação atual da Análise de Assunto, se menciona também esse aspecto, especialmente no tocante à avaliação de novas linguagens de indexação. Neste contexto se situa a questão sobre a reformulação de diferentes tipos de entradas pelo usuário do sistema. Qual seria, por exemplo, o efeito de conectivos, de determinadas ordens de palavras, do uso da hifenização em português etc. sobre a compreensão e a velocidade do processamento da leitura no usuário? O usuário concentrar-se-ia em palavras-chave ou lê as entradas de modo parcelado? Das soluções de tais questões, muitas das quais condicionadas culturalmente, dependeria a elaboração de determinados tipos de índices.

A pesquisa sobre a leitura poderia contribuir, também, na otimização de sistemas em linha. Atribui-se atualmente crescente importância aos aspectos comportamentais do usuário, com vistas à legibilidade de um sistema. A partir de pesquisas empíricas (de interface) poderiam advir elementos preciosos para a implementação de sistemas de maior rendimento.

7 - CONSIDERAÇÕES FINAIS

A interseção da Lingüística com a Ciência da Informação vem-se tornando objeto de reflexão científica há mais de dez anos, tendo-se evidenciado não somente em uma vasta literatura especializada, mas também a nível institucional (a criação do Comitê de Lingüística em Documentação, na FID). Atualmente, continua em pleno curso (v. supra, item 2) o inter-relacionamento de algumas subáreas

da Lingüística com subáreas da Ciência da Informação, mormente a indexação automática.

Entretanto, embora conhecidos os resultados práticos decorrentes, não se incentivou ainda, pelo menos não no Brasil, a discussão relativa a um possível balizamento da área de interseção, do que dependeriam pesquisas e projetos de interface a médio-prazo.

Outra questão, intimamente ligada à anterior, pode ser desdobrada em duas, ou seja: 19) o "recorte" a ser efetuado no universo da Lingüística e em áreas afins, para integrar o programa de formação acadêmica dos futuros cientistas da informação; 29) a metodologia a ser adotada na implementação didática daquele "recorte" para o grupo-alvo em pauta.

No presente artigo, que se desenvolveu a partir de uma palestra sobre o assunto, restringimo-nos àquelas áreas e subáreas que nos pareceram de interesse para a Ciência da Informação. Além disso, sugerimos também a inclusão eventual de áreas afins, até aqui ainda não consideradas para esta finalidade, no contexto brasileiro.

Enquanto já se publicaram alguns trabalhos no Brasil sobre a interseção da Lingüística com a Ciência da Informação (Lingüística e Documentação, Lingüística e Indexação etc.), ainda não se mencionou a Terminologia Teórica e Aplicada como área de interface em potencial.

De qualquer modo, urge enfatizar o caráter instrumental-aplicativo, a ser assumido pela Lingüística ou outra área afim, bem como a necessidade de uma atitude pragmática da parte de lingüistas interessados numa aproximação com a Ciência da Informação. Isto evitará que a cooperação desejada por ambos redunde num diálogo mutuamente ininteligível — queixume freqüente entre uns e outros. Como ainda não se elaborou até hoje uma "Lingüística Aplicada à Ciência da Informação" no Brasil, encontrar uma linguagem comum continuará a depender dos próprios especialistas envolvidos, pertencentes a uma ou outra área.

Contrariando previsões mais pessimistas, que vinham acompanhando a evolução da "informatização" das sociedades pós-industriais, a linguagem natural readquiriu hoje uma importância sem precedentes. Com a capacidade de memória computacional sensivelmente aumentada e a implementação de tecnologia de diálogo homem-computador, a

linguagem humana voltou a ocupar uma posição privilegiada nas pesquisas de interface.

Destarte, a própria evolução da tecnologia, paradoxalmente, vem contribuindo para a reavaliação de um ideal da filosofia pré-aristotélica, que nos primórdios da História do Pensamento no Ocidente reconhecera no homem a medida de todas as coisas (Protágoras de Abdera, entre 480 e 410 a.C.).

AGRADECIMENTOS

O presente trabalho foi parcialmente aproveitado em um Seminário sobre Lingüística e Indexação promovido pelo IBICT/Mestrado em Ciência da Informação, em 21 de fevereiro de 1983. Agradeço pelas críticas dos participantes daquele encontro, em especial pelas judiciosas observações das professoras Anna Maria Marques Cintra e Jandira Batista Assunção. Continuo inteiramente responsável pelas colocações e pelo caráter provisório das idéias aqui expostas, desejando antes encetar uma discussão interdisciplinar do que propriamente oferecer soluções.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 JONES, K.S. & KAY, M. *Linguistics and Information Science*. New York, London, Academic Press, 1973.
- 2 DAMERAU, F.J. Automated language processing. *Annual Review of Information Science and Technology*, 11: 107-61, 1976.
- 3 TRAVIS, I. & FIDEL, R. Subject Analysis. *Annual Review of Information Science and Technology*, 17:123-57, 1982.
- 4 JONES, K.S. The role of automatic indexing in operational on-line retrieval systems. In: *New trends in documentation and information*. Proceedings of the 39th FID Congress, University of Edinburgh, 25-8 September, 1978. FID Publication 566. London, Aslib/FID, p.33.
- 5 TRAVIS, I. & FIDEL, R., op. cit. p.143.
- 6 Idem, *ibid.* p. 143.
- 7 BARANOW, U.G. Aspectos lingüísticos de linguagens de indexação. **Anais da Conferência Brasileira de Classificação Bibliográfica**. Rio de Janeiro, 12 a 17 de setembro de 1976. Rio de

- Janeiro: IBICT; Brasília: ABDF, v. 1, p. 295-310.
- ⁸ TRAVIS, I. & FIDEL, R., op. cit. p. 139.
- ⁹ Idem, *ibid.* p. 137.
- ¹⁰ ROBREDO, J. Otimização dos processos de indexação dos documentos e de recuperação da informação mediante o uso de instrumentos de controle terminológico. *Ciência da Informação*. 11(1): 3-18, 1982.
- ¹¹ MICHELL, G. The natural language foundations of indexing language relations. *The Canadian Journal of Information Science*, 4: 99-104, 1979.
- ¹² JONES, K.S. & KAY, M., op. cit. p. 173.
- ¹³ WALKER, D.E. The organization and use of information: contributions of Information Science, computational linguistics, and artificial intelligence. *Journal of the American Society for Information Science*, 33(5): 347-63, 1981.
- ¹⁴ BECKER, D. Automated language processing. *Annual Review of Information Science and Technology*, 16:113-38, 1981.
- ¹⁵ Idem, *ibid.* p. 118.
- ¹⁶ WALKER, D.E., op. cit., p. 362.
- ¹⁷ GRICE, H.P. Logic and conversation. COLE, P. & MORGAN, J.L., ed. *Syntax and Semantics*, v. 3: Speech Acts. New York, Academic Press, 1975. p. 41-58.
- ¹⁸ BASÍLIO, M. de P. Interface linguística e ciência da informação: potencialidades na análise de estruturas lexicais. *Reunião Brasileira de Ciência da Informação*, 2. Rio de Janeiro: IBICT, 1(61):233-46, 1979.
- ¹⁹ BECKER, D., op. cit. p. 124.
- ²⁰ NOEL, J. Syntax, semantics and pragmatics in the automatic analysis of texts. In: *New trends in documentation and information*. Proceedings of the 39th FID Congress, University of Edinburgh, 25-8 September, 1978. FID Publication 566. London, Aslib/FID, p. 39-51. 1980.
- ²¹ WALKER, D.E., op. cit. p.361.
- ²² Idem, *ibid.* p.350.
- ²³ CASSIM, M.B. A informática e a evolução dos centros de informação a serviço de mudanças culturais. *Ciência da Informação*, Brasília, 11 (11:61-7, 1982.
- ²⁴ ARNTZ, R. & PICTH, H. *Einführung in die übersetzungsbezogene Terminologiearbeit*. Hildesheim, Zürich, New York, Georg Olms Verlag. 1982. 238p.
- ²⁵ Idem, *ibid.* p.193.
- ²⁶ TRAVIS, I. & FIDEL, R., op. cit. p.124.
- ²⁷ Idem, *ibid.* p.126.
- ²⁸ CINTRA, A.M.M. Elementos de linguística para estudos de indexação. Paper apresentado por ocasião do Seminário sobre Linguística e Indexação. IBICT, Rio de Janeiro, 1983, 32p. (datilografado).
- ²⁹ TRAVIS, I. & FIDEL, R., op. cit. p.133.

ABSTRACT

Some of the difficulties arisen from the interrelationship between Information Science and Linguistics, besides a mutual lack of interdisciplinary understanding, are due to expectations that Linguistics cannot now and perhaps will not ever be able to fulfil. Looking at the current research concerning Information Science and Linguistics we observe a renewed interest in natural language processing. This is mainly caused by the implementation of question-answer systems and the increase in computer memory capacity. As Linguistics is of continuing importance to Information Science, we are trying to establish those parts of it that could eventually be included in the regular Brazilian courses for information scientists. Along with this, Computational Linguistics, Reading Research, Science of Terminology and Translation Science are suggested as complementary disciplines. Each of these subareas is briefly analyzed with regard to Information Science.