

DOS OLHARES SOBRE O LÉXICO: DIFERENÇAS E SEMELHANÇAS

Ariani DI FELIPPO (PG – FCL – UNESP/Ar.)

Bento Carlos DIAS-DA-SILVA (FCL – UNESP/Ar.)

Resumo: O léxico tem sido foco de interesse em vários campos ou áreas de pesquisa, mais especificamente, na Lingüística Teórica, na Psicolingüística e no Processamento Automático das Línguas Naturais. Neste capítulo, apresenta-se o tratamento dado ao léxico nos estudos desenvolvidos nessas três diferentes áreas e ressaltam-se diferenças e semelhanças na concepção do objeto “léxico”. Por fim, enfatiza-se que a construção de léxicos computacionais pode ser auxiliada pelas pesquisas desenvolvidas nos campos da Lingüística Teórica e da Psicolingüística, o que propicia o desenvolvimento de léxico computacionais (psico)lingüísticamente motivados.

0. Introdução

A investigação sobre a natureza do léxico tem ocupado lugar de centralidade nos estudos desenvolvidos na Lingüística Teórica, na Psicolingüística e no Processamento Automático das Línguas Naturais (PLN) ou Tecnologia da Linguagem Humana (TLH) (Handke, 1995).

Na Lingüística Teórica, a natureza do léxico tem sido investigada em diferentes modelos teóricos ou gramáticas¹. Em modelos gramaticais modernos, a concepção de léxico como uma “lista de entradas lexicais” tem sido amplamente abandonada e substituída por concepção “altamente estruturada” de léxico . Conseqüentemente, as principais questões discutidas no âmbito das teorias lingüísticas são: (i) o desenvolvimento de mecanismos que capturam generalizações

¹ O termo *gramática* é tomado no seu sentido mais amplo que compreende o estudo da língua como um todo, podendo ser substituído por *teoria* ou *modelo*.

lingüísticas no léxico e (ii) a especificação da natureza e do conteúdo das entradas lexicais.

Na Psicolingüística, o léxico tem sido considerado o componente central do processamento cognitivo da linguagem. Nesse campo de pesquisa, as principais questões que estão sendo investigadas são: (i) a elaboração de estratégias de acesso ao léxico e (ii) especificação das estruturas global e interna do léxico.

No PLN, os esforços dispensados no desenvolvimento de léxico computacionais por meio de informações extraídas manual, semi ou automaticamente de obras lexicográficas demonstram o papel central que o léxico desempenha nos sistemas que processam língua natural, como: sistemas de tradução automática, sistemas de sumarização, entre outros. Nessa área, as principais questões investigadas são: (i) a especificação das estruturas global e interna do léxico e (ii) a representação formal do componente lexical.

Neste capítulo, em especial, apresenta-se o tratamento dado ao léxico nos estudos desenvolvidos nessas três diferentes áreas, buscando-se ressaltar diferenças e semelhanças na concepção do objeto “léxico”. Por fim, demonstra-se como a construção de léxicos lingüístico-computacionais pode ser auxiliada pelas pesquisas desenvolvidas nos campos da Lingüística Teórica e da Psicolingüística.

1. Do *olhar* da Lingüística Teórica

Nos trabalhos desenvolvidos pela lingüística teórica na década de 1950 e início de 1960, a sintaxe abrangia todos os aspectos regulares da língua, desde a apassivação até os processos de formação de palavras como a nominalização e a composição.

Nesse contexto, o léxico² não passava de uma lista não-estruturada de “palavras”, contendo, para cada uma delas, a especificação de sua realização fonética, de suas propriedades morfológicas, sintáticas e semânticas, além de conter todas as irregularidades, isto é, formas não previstas como o resultado da aplicação de princípios da gramática (Chomsky, 1964; Di Sciullo e Williams, 1987). A essa concepção de léxico como “componente marginalizado” da gramática e “lista desestruturada” de palavras, dá-se, neste capítulo, a denominação **concepção tradicional**³.

Nas décadas de 1960 e 1970, entretanto, a lexicografia atraiu a atenção dos lingüistas teóricos, o que desencadeou, nesse período, inúmeros trabalhos que revelavam o crescente interesse pelo léxico. Dentre esses trabalhos, citamos, por exemplo, o desenvolvimento da teoria semântica de Katz e Fodor (Katz e Fodor, 1963). Outros exemplos paradigmáticos são os trabalhos de Chomsky (1970) e Jackendoff (1975). Nesses estudos, os autores focalizam a distinção entre fenômenos sintáticos e lexicais e a identificação de generalizações morfológicas e semânticas, o que desencadeou mudanças na concepção tradicional, mais especificamente, no papel e na estruturação do léxico, tanto no âmbito da tradição gerativa como em outras correntes teóricas européias (Atkins e Zampolli, 1994).

No âmbito do gerativismo, por exemplo, a proposição da *Teoria da Regência e Ligação* (Government-Binding Theory – GB) (Chomsky, 1981) visava a restringir o poder das regras sintáticas; com isso, várias propriedades gramaticais passaram a ser vistas como projeções das propriedades lexicais, ou seja, das propriedades armazenadas no léxico. Tal idéia de centralidade do léxico atingiu seu auge com o

² Nos modelos gerativos, o *léxico* (um dos componentes do sistema abstrato que é a língua) deixa de ser o vocabulário da língua como realidade externa e passa a ser parte do conhecimento cognitivo do falante, isto é, *léxico mental* (Basílio, 1999).

³ Vale ressaltar que o léxico passa a ser concebido como um componente autônomo da gramática na Teoria Padrão de Chomsky (1964).

desenvolvimento das teorias dissidentes do gerativismo e denominadas **lexicalistas**, como a *Gramática Léxico-Funcional* (Lexico-Functional Grammar – LFG) (Bresnan, 1982) e a *Generalized Phrase Structure Grammar* (GPSG) (Gazdar et. al., 1985), que tem hoje na *Head-Driven Phrase Structure Grammar* (HPSG) (Pollard e Sag, 1994) sua principal representante. Segundo Handke (1995), nesses modelos, o léxico não só contém informações sobre as “palavras” da língua, mas também é usado como uma unidade de controle que examina a “boa formação” das sentenças geradas pelo próprio modelo gramatical.

A LFG, por exemplo, permite que o léxico contenha um substantivo conjunto de informações, como: *matriz fonética; forma lexical (ou gráfica); categoria sintática; restrições seletivas; forma semântica; funções gramaticais associadas; papéis temáticos associados; relações gramaticais do tipo função gramatical ↔ papel temático; traços semânticos como [+/- genérico], [+/- específico], [+/- definido], [+/- humano], [+/- concreto], entre outros; categorias de gênero, número, pessoa, caso, voz, tempo, aspecto e modo; traços de reflexividade e ergatividade* (Bresnan, 1982).

A *Head-Driven Phrase Structure Grammar* (HPSG) (Pollard e Sag, 1994), em especial, adota uma postura radicalmente lexicalista, pois a maior parte da informação gramatical e semântica é codificada nas entradas lexicais. Esse modelo, inclusive, amplia o conceito de *signo* de Saussure (1972) – unidade mínima independente que relaciona, de modo arbitrário, forma e significado – posto que um signo pode ser uma palavra, um sintagma ou uma sentença e constitui-se de um complexo estruturado de informações fonéticas, sintáticas, semânticas e restrições contextuais. Além disso, a HPSG parte do princípio de que (i) os signos lingüísticos podem ser representados em

sistemas de **tipos hierarquicamente organizados** e (ii) a gramática das línguas pode ser representada em termos de **sistemas de restrições**.

A utilização de hierarquias de tipos imprime ao léxico organização global e tem por objetivo reduzir parte da informação redundante no léxico (Flickinger, Pollard e Wasow, 1985; Koenig, 1999). Os tipos exemplificados, na Figura 1, são: palavras (*word*) e constituintes sintagmáticos (*phrase*); classes de partes-do-discurso (*part-of-speech*), tais como nome (*noun*), verbo (*verb*) e adjetivo (*adjective*); classes baseadas na valência, tais como transitividade (*transitive*) e intransitividade (*intransitive*).

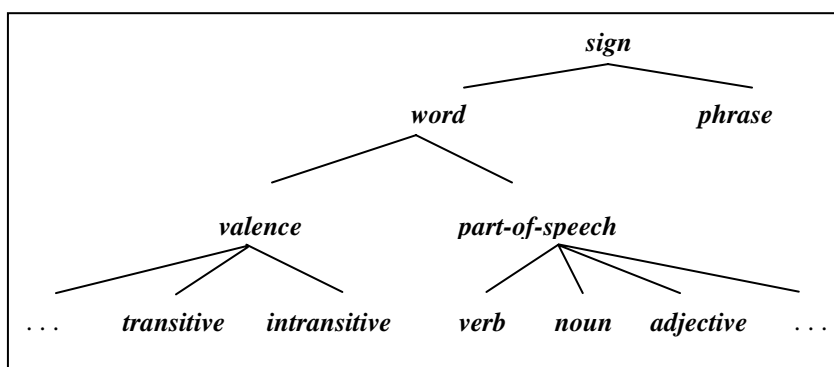


Figura 1: Representação de uma hierarquia de tipos.

As hierarquias de tipos englobam uma **hierarquia de herança múltipla** de restrições. As restrições associadas às classes particulares são herdadas por suas subclasses e, consecutivamente, por seus membros individuais. Em outras palavras, os subtipos herdam todas as restrições impostas pelos seus supertipos. As informações inferidas a partir de *sign*, por exemplo, não são especificadas novamente para *word* ou *phrase*. De um modo geral, os signos pertencem, ao mesmo tempo, a várias classes entrecruzadas de tipos, daí a denominação **herança múltipla**. Por exemplo, uma palavra pode ser membro da classe dos verbos ou dos adjetivos e, ao mesmo tempo,

pertencer às classes das palavras transitivas ou intransitivas. Com a utilização de hierarquias de heranças múltiplas, evita-se redundância no léxico, já que informações compartilhadas não são repetidas (Koenig, 1999). Na frase *Carlos comeu uma maçã*, a informação associada ao item “comer” resulta do fato de ser membro de várias classes, dentre elas, *verb* e *transitive*.

Diante de modelos gramaticais como a LFG, GPSG e HPSG, baseados na concepção de léxico como “componente central e estruturado” da gramática, observa-se que o componente lexical da língua passou a ser considerado um autêntico objeto de estudo e que as propriedades léxico-gramaticais das “palavras” e a relação entre essas propriedades passaram a ocupar lugar de destaque no cenário dos estudos lingüísticos.

Vê-se, no entanto, que não há um consenso na Lingüística Teórica a respeito da natureza do léxico. Mas existem pelo menos algumas hipóteses que parecem consensuais. Pode-se dizer, em resumo, que, diferentemente da concepção tradicional, o léxico é concebido nos modelos gramaticais modernos como uma rede de correspondência em vários níveis e o item lexical, como uma trilha de propriedades fonológicas, morfológicas sintáticas e semânticas. Dado o número elevado dos elementos do léxico mental e da complexidade combinatória resultante desse número, pressupõe-se que os itens estejam organizados de maneira funcional, para que o falante possa recuperar rapidamente não só o significado de um item, mas também todas as suas características gramaticais e usos (Mel’čuk, 1988).

2. Do *olhar* da Psicolinguística

Unindo pressupostos da Linguística e da Psicologia, a Psicolinguística estuda a existência e o funcionamento de mecanismos mentais envolvidos no processamento da linguagem humana (Saint-Dizier e Viegas, 1995).

Com o intuito de compreender, entre outras questões, como ocorre o armazenamento e o acesso aos itens lexicais de uma determinada língua, os psicolinguistas postulam a existência de um **léxico mental** (LM), definido como a parte do conhecimento lexical do indivíduo delimitada por sua língua (Bierwisch e Schreuder, 1992; Levelt, 1993).

De acordo com vários autores (Bock, 1982; Bierwisch e Schreuder, 1992; Handke, 1995; Levelt, 1992;1993; Lowie, 1998), o léxico mental ocupa lugar de centralidade no processamento cognitivo da linguagem, o qual envolve três tipos de processos: (i) **conceitualização** (especificação de conceitos); (ii) **formulação** (seleção de itens e construção de representações sintáticas e fonéticas); (iii) **articulação** (produção da fala).

No processo de produção de enunciados, por exemplo, a nomeação de um objeto perceptível envolve: (i) a identificação do objeto (conceitualização); (ii) a seleção de uma representação sintático-semântica do objeto, assim como a codificação dessa representação em termos fonológicos (formulação); (iii) a transformação da representação fonológica em realização fonética, que constitui o nome do objeto (articulação). A Figura 2 ilustra o papel do léxico mental nos processos cognitivos de produção e interpretação da linguagem.

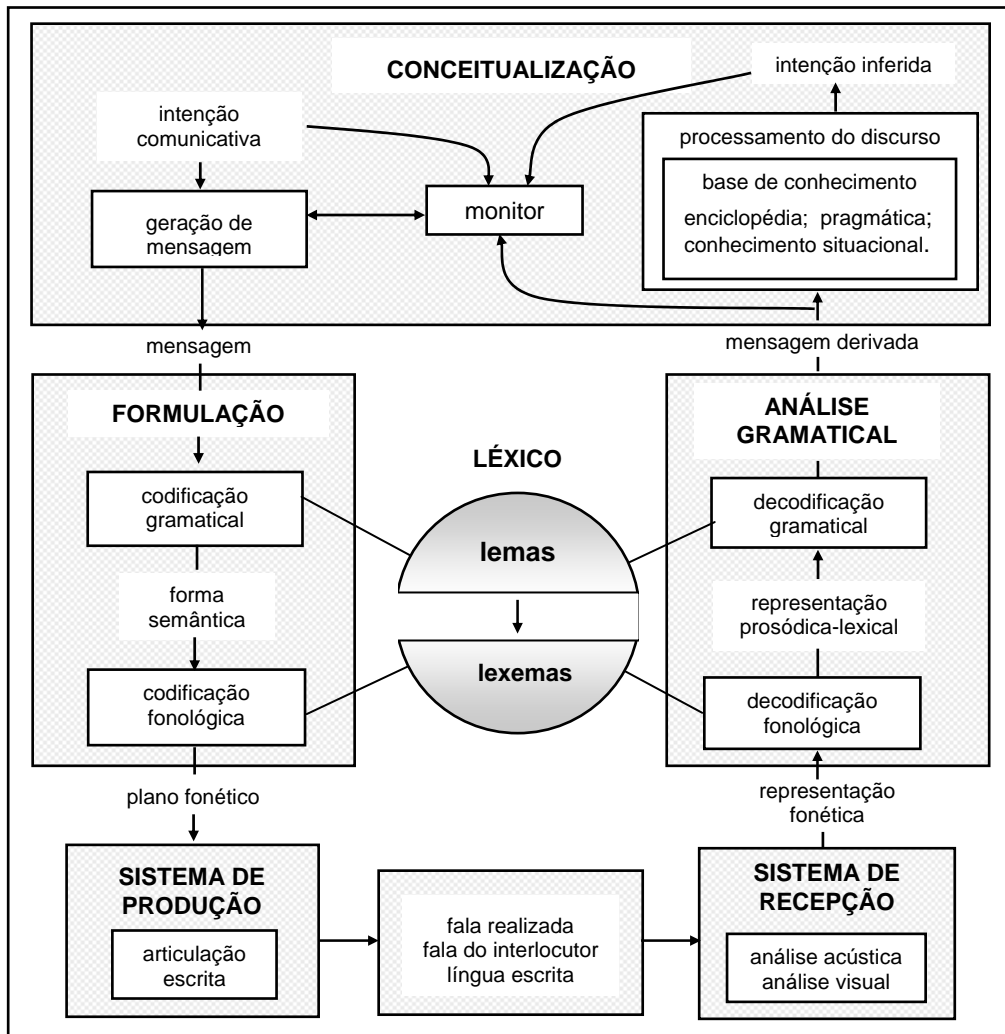


Figura 2: Arquitetura do processamento mental da linguagem baseada em Handke (1995) e Lowie (1998).

A identificação do objeto ou **conceitualização** ativa uma robusta *base de conhecimento* que contém informações extralingüísticas provenientes de diversas fontes (visual, auditiva, motora, emotiva, conceitual, entre outras), além de princípios gerais de organização conceitual (ontologia do senso comum, conceitualizações do espaço e tempo, condições gerais subjacentes ao conhecimento enciclopédico ou a sistemas de crença, etc). O processo de conceitualização gera uma *estrutura conceitual* (EC) (pré-lingüística), que é a mensagem a ser verbalizada e organizada gramaticalmente pela *formulação*, no caso, essa mensagem será o *nome* do objeto.

Já a **formulação** é responsável por transformar essa EC em um enunciado lingüístico. Essa transformação é mediada pelo LM, que, como já se disse, é a parte do conhecimento lexical delimitada pela língua do falante. Assim, a língua do indivíduo delimita o LM, que, por sua vez, media a transformação da *estrutura conceitual* em um *enunciado lingüístico*. Em outras palavras, pode-se dizer que os estímulos recebidos por um indivíduo (= estrutura conceitual) são traduzidos em itens lexicais de acordo com regras e princípios de cada língua. A essa hipótese, Glanzer e Clark (*apud* Biderman, 1981) deram a denominação **elo verbal** (do inglês, “*verbal-loop hypothesis*”).

De acordo com Bierwisch e Schreuder (1992), a conversão ou transformação de uma estrutura conceitual em um enunciado lingüístico é feita em dois estágios: *ativação* ou *seleção de lemas* e *ativação de lexemas* ou *codificação da forma*. Postula-se, assim, a existência do LM no nível lingüístico. O LM é central a todo o processamento da linguagem e contém todas as informações sobre os itens lexicais da língua, isto é, seus *lemas* e *lexemas*⁴. Dessa forma, o primeiro estágio da formulação é responsável pela seleção da representação sintático-semântica do objeto. Para tanto, é ativado, no LM, o **lema** do objeto, ao qual estão associadas informações sintáticas e semânticas que determinam, por exemplo, sua forma semântica, categoria sintática e estrutura de argumentos. De acordo com Bierwisch e Schreuder (1992), o resultado do primeiro estágio, a *formulação*, é uma *forma semântica* (FS)⁵. No segundo estágio, essa FS é transformada na *forma fonológica* (FF). Para que essa transformação seja

⁴ Cabe ressaltar que os termos *lema* e *lexema* não estão sendo empregados no sentido típico do campo da Lexicografia, isto é, uma representação canônica das entradas de um dicionário (*lema*) ou de uma unidade lexical virtual que compõe o léxico (*lexema*) (Biderman, 1999). Para Bierwisch e Schreuder (1992), *lema* é a representação das propriedades sintático-semânticas de um item lexical e *lexema* é a representação das estruturas morfológica e fonológica de um item lexical.

⁵ Vale ressaltar que há divergências quanto à postulação dos níveis EC e FS. A Bierwisch e Schreuder (1992), que defendem essa proposta, opõe-se, por exemplo, Jackendoff (1991), que propõe um nível único denominado nível da *estrutura léxico-conceitual* (ELC). Para Jackendoff, a FS não é concebida como uma entidade distinta da *estrutura conceitual*, mas parte dela.

possível, é ativado, no LM, o **lexema** do objeto, ao qual estão associadas informações fonológicas e morfológicas.

Por fim, a **articulação** opera sobre a FF, ativando programas articulatorios que produzem a realização fonética do objeto.

Tendo em vista que o acesso aos itens lexicais realiza-se nas etapas de *seleção do lema e codificação da forma* da palavra, a representação da microestrutura do LM, ou seja, da estrutura interna das entradas lexicais, subdivide-se em *unidade de acesso e especificação lexical*. A **unidade de acesso** é o item lexical propriamente dito. A **especificação lexical** de um item é a representação do seu *lema* e do seu *lexema*, os quais estão interligados por um ponteiro lexical (isto é, cada *lema* “aponta” para um *lexema* correspondente). Como já se disse, o **lema** é a representação das propriedades semânticas e sintáticas do item lexical; especifica as condições conceituais que garantem o uso apropriado do item, indicando, entre outras coisas, sua classe gramatical e seus argumentos; o **lexema**, por sua vez, é a representação das estruturas morfológica e fonológica de um item lexical.

3. Do olhar do PLN

Todos os sistemas de PLN, diferentes dos demais tipos de sistemas computacionais, manipulam o código lingüístico e, por isso, pressupõem um tipo de “arquivo” em que são armazenadas as unidades lexicais (palavras e expressões) que serão manipuladas pelo sistema durante os procedimentos de interpretação e/ou produção de língua natural. Esse “arquivo” é concebido como uma base de dados em que são especificadas, para cada unidade nela contida, informações de natureza lexical, morfológica, sintática, semântica e, até mesmo, pragmático-discursiva,

dependendo das especificidades do sistema de PLN para o qual essa base de dados é desenvolvida (Palmer, 1999). Do ponto de vista do PLN, esse tipo de “megarquivo” é definido como o “léxico” do sistema.

Teoricamente, as arquiteturas propostas para sistemas de PLN acabam por espelhar a arquitetura proposta para o sistema lingüístico (Allen, 1987). Como decorrência, um sistema de PLN deve possuir módulos autômatos, que realizam tarefas específicas e especializadas, e módulos que armazenam um modelo de conhecimento proposicional, que visa a criar simulacros de parcelas de mundo que lhe servem de referencial para interpretar os enunciados lingüísticos. Apesar da arquitetura de um sistema de PLN variar de acordo com as especificidades da aplicação, dois grupos de componentes são imprescindíveis para a implementação de qualquer sistema desse tipo: as *bases de conhecimento* e os *módulos de processamento* que atuam sobre essas bases (Dias-da-Silva, 1996). A Figura 3 ilustra esses dois grupos de componentes.

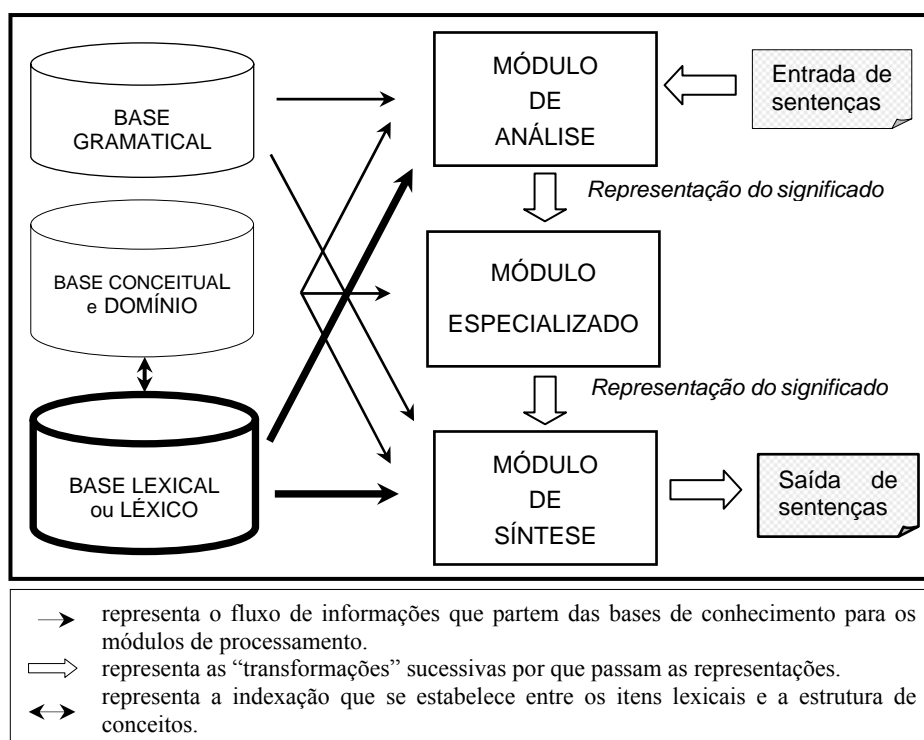


Figura 3: Arquitetura de um sistema de PLN genérico de Dias-da-Silva (1996).

Os módulos de conhecimento podem ser divididos em três módulos: o de *análise*, o *especializado* e o de *síntese*. As bases de conhecimento podem ser divididas em três bases: *gramatical*, *conceitual* e *lexical*. Com exceção do *módulo especializado*, os demais módulos de processamento e as bases de conhecimentos, embora os conteúdos possam variar em função da especificidade do sistema, possuem estrutura e funcionamento semelhantes.

O **módulo de análise** (MA) é geralmente formado pelo analisador morfológico e pelo analisador sintático (ou *parser*), além dos interpretadores semântico e pragmático-discursivo. Esse módulo é responsável pela construção de uma representação interna do significado das sentenças de entrada (no caso, digitadas via teclado).

O **módulo de síntese** (MS), por sua vez, transforma a representação abstrata gerada pelo MA em uma seqüência de “frases contextualizadas”. Ao realizar a tarefa de construção de uma representação semântica, por exemplo, o MA utiliza-se, dependendo da sofisticação do sistema de que é parte, das bases gramatical, conceitual e lexical para executar todas ou parte das análises: morfológica, sintática, semântica e, até mesmo, pragmática. Assim, cada base de conhecimento, por sua vez, fornece ao MA informações de natureza diferente (Hutchins e Somers, 1997).

A **base gramatical** fornece a representação das regras sintáticas da língua, que podem ser vistas como condições de admissibilidade de estruturas sintáticas bem-formadas; condições que servirão de referência para o módulo de análise – responsável pela construção das representações sintáticas, semânticas e pragmático-discursivas.

A **base conceitual** fornece um modelo do mundo físico e conceitual, descrevendo tipos básicos de objetos, eventos, propriedades, relações e atributos em

termos de representações hierarquicamente estruturadas, isto é, a sua estrutura consiste em uma rede de unidades conceituais interligadas em termos de relações de hiponímia/ hiperonímia, entre outras. Essa base também pode fornecer conceitos mais específicos, ou seja, conceitos referentes a domínios particulares do conhecimento ou conceitos relacionadas a tarefas específicas para a qual o módulo esteja sendo projetado.

Em particular, à **base lexical**, fica a tarefa de fornecer, aos MA e MS, a coleção de unidades lexicais, para as quais se faz necessária a especificação de conjuntos de traços morfológicos, sintáticos, semânticos e pragmático-discursivos (Boguraev, Briscoe; 1989; Briscoe, 1991; Sanfilippo, 1995; Palmer, 1999). Esse tipo de base de dados, no domínio do PLN, é definido como o léxico do sistema e recebe a denominação de **léxico tratável por máquina** (“*machine tractable dictionary*”) (Wilks, 1988).

4. Diferenças e semelhanças

Diante do exposto, observa-se que há, ao mesmo tempo, aspectos semelhantes e díspares na concepção de léxico nas áreas da Psicolinguística, Linguística e Processamento Automático das Línguas Naturais.

Ressalta-se que a principal diferença dos *olhares* dessas três áreas reside na concepção ou definição do objeto “léxico”. Sob o olhar da Psicolinguística, trata-se do léxico mental, ou seja, a parte do conhecimento lexical do indivíduo delimitada por sua língua, sob o ponto de vista da linguística, trata-se de um dos componente da gramática e, sob o ponto de vista do PLN, trata-se de um “megarquivo”.

A principal semelhança desses olhares, por sua vez, reside na função que cada um desses objetos desempenha no sistema do qual faz parte. Para os três tipos de sistemas, cognitivo, gramatical e do PLN, o léxico necessita fornecer os itens lexicais e as suas respectivas propriedades; na terminologia da Psicolinguística, as **unidades de acesso** e as respectivas **especificações lexicais**.

5. Conclusão: contribuindo com o PLN

Em função das aplicações e situações reais para as quais os programas de PLN são escritos, é premente, na construção de um tradutor automático, por exemplo, a compilação de léxicos, monolíngües e multilíngües, que sejam manipuláveis pelos programas que compõem o sistema e que sejam *robustos*, tanto do ponto de vista da quantidade de formas quanto do ponto de vista da qualidade e precisão das informações a elas associadas. São consideradas etapas essenciais desse empreendimento o planejamento lingüístico do léxico a ser construído, sua especificação lógico-estratégica e sua implementação computacional. Esse planejamento prevê, portanto, a fatoração das atividades de pesquisa em três domínios complementares: Domínio das Entidades e das Relações Lingüísticas, Domínio da Formalização e Domínio da Implementação (Dias-da-Silva, 1998).

No domínio das entidades e relações lingüísticas, os pesquisadores do PLN precisam delimitar a **unidades de acesso** e as respectivas **especificações lexicais** a serem associadas a cada entrada lexical. É claro que, como mencionado na Seção 3, o conteúdo e a estrutura do léxico podem variar de acordo com a aplicação para a qual o sistema de PLN é feito.

Dessa forma, delimitar as propriedades lingüísticas a serem associadas a cada entrada lexical depende da observação de algumas variáveis, como: a *necessidade* e o *tipo* da propriedade. Essas variáveis estão diretamente ligadas ao tipo de unidade de acesso e ao tipo do sistema, por exemplo, no caso de um “léxico de raízes” (“*root lexicon*”), em que apenas as formas que servem de base para a formação de outras formas são armazenadas, faz-se necessária a especificação de informações sobre as alternâncias ortográficas, no caso de um sistema de interpretação/produção de língua escrita, e fonológicas, no caso de um sistema de interpretação/produção de língua falada; em um léxico “de formas plenas” (“*full-form lexicon*”), em que são listadas todas as formas possíveis da língua, essa especificação não é necessária (Butterworth, 1983).

Mesmo que a delimitação da especificação lexical e das unidades de acesso dependa das especificidades dos sistemas de PLN, ressalta-se que trabalhos como os de Bock (1982), Bierwisch e Schreuder (1992), Handke (1995), Levelt (1992, 1993) e Lowie (1998) podem fornecer os subsídios para a concepção de uma entrada lexical (psico)lingüísticamente motivada, pois estão pautados em hipótese sobre o acesso ao léxico mental. Já trabalhos como os de Flickinger, Pollard e Wasow (1985), Pollard e Sag (1994), Bouma, Eynde e Flickinger (2000), Koenig (1999), entre outros, fornecem os subsídios, nas pesquisas no domínio da formalização, para a concepção de uma organização global do léxico, posto que, por meio da hierarquia de tipos e do mecanismo de herança múltipla, é possível capturar generalizações e reduzir redundâncias no léxico.

Por fim, pode-se dizer que, com a ajuda das pesquisas lingüísticas e psicolingüísticas, os pesquisadores do PLN podem desenvolver léxico computacionais (psico)lingüísticamente motivados.

7. Referências Bibliográficas

- ALLEN, J. F. **Natural language understanding**. Menlo Park: Benjamin Cummings, 1987.
- ANDERSON, S. **A-morphous morphology**. Cambridge: Cambridge University Press, 1992.
- ATKINS, B.T.S.; ZAMPOLLI, A. **Computational approaches of the lexicon**. Oxford-New York: Oxford University Press, 1994.
- BASÍLIO, M. Introdução: questões clássicas e recentes da delimitação das unidades lexicais. **Palavra**, n.5, p.9-18, PUC/Rio, 1999.
- BIDERMAN, M. T. C. A estrutura mental do léxico. In: **Estudos de Filologia e Lingüística - Homenagem a Isaac Nicolau Salum**. São Paulo: Editora da USP; T. A. Queiroz, p. 131-45, 1981.
- _____. Conceito Lingüístico de Palavra. In: Basílio, M. (ed.) **Palavra**. Departamento de Letras da PUC-Rio, pp. 81-97, 1999.
- BOCK, J. K. Towards a Cognitive Psychology of Syntax. **Psychological Review**, 89, p. 1-47, 1982.
- BOGURAEV, B., BRISCOE, T. (Eds.). **Computational Lexicography for Natural Language Processing**, London: Longman, 1989.
- BOUMA, G.; EYNDE, F. van. FLICKINGER, D. Constraint-based lexicons. In: Eynde, F. van (Ed.). **Lexicon Development for Speech and Language Processing**. ELSNET, Leuven, 2000.
- BRESNAN, J. (Ed.). **The mental representation of grammatical relations**. Cambridge, Mass: The MIT Press, 1982.
- BRISCOE, T. Lexical issues in natural language processing. In: Klein, E.; Veltman, F. (Eds.). **Natural language and speech**. Springer-Verlag, p.39-68, 1991.
- BUTTERWORTH, B. (Ed.) **Language production volume 2: development, writing and other language processes**. London: Academic Press, 1983.
- CHOMSKY, N. **Aspects of the theory of syntax**. Cambridge: MIT Press, 1964.
- _____. Remarks on nominalizations. In: Jacobs, R. A.; Rosenbaum, P.S. (Eds). **Readings in English Transformational Grammar**. Ginn and Company, 1970.
- _____. **Lectures on government and binding**. Dordrecht: Foris, 1981.
- DIAS-DA-SILVA, B. C. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais**. Araraquara, 1996. 272p. Tese

(Doutorado em Letras) - Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara.

_____. Bridging the gap between linguistic theory and natural language processing. In: **Proceedings of the 16th international congress of linguistics**. Oxford: Elsevier Sciences, n. 16, p. 1-10, 1998.

DI SCIULLO, A-M; WILLIAMS, E. **On the definition of word**. Cambridge: MIT, Press, 1987.

FLICKINGER, D.; POLLARD, C.; WASOW, T. Structure sharing in lexical representation. In: **Proceedings of the 23rd annual meeting of the association for computational linguistics**. Morristown, N.J.: Association for computational linguistics, 1985.

GAZDAR, G. et al. **Generalized phrase structure grammar**. Cambridge, Mass.: Harvard University Press, 1985.

HANDKE, J. **The structure of the Lexicon: human versus machine**. Berlin: Mouton de Gruyter, 1995.

HUTCHINS, W. J.; SOMERS, H. L. **An introduction to machine translation**. London: Academic Press, 1997.

JACKENDOFF, R. Morphological and semantic regularities in the lexicon. **Language**, v.51, 1975.

_____. **Semantic structures**. Cambridge: Mass.: The MIT Press, 1991.

KATZ, G.G.; FODOR, J. A. The structure of a semantic theory. **Language**, v.39, 170-210, 1963.

KOENIG, J-P. **Lexical relations**. Stanford, CA: 1999.

LEVELT, W .J. M. Accessing words in speech production: stages, processes and representations. **Cognition**, 42, p.1-22, 1992.

_____. **Speaking: to intention to articulation**. Cambridge, Mass.: The MIT Press, 1993.

LOWIE, W. **The acquisition of interlanguage morphology: a study into the role of morphology in the L2 learner's mental lexicon**. Groningen: University Library Groningen, 1998.

MEL'CUK, I. **Dependency Syntax: theory and practice**. The SUNY Press, Albany, N.Y, 428p. 1988.

PALMER, M. Multilingual resources – Chapter 1. In: Hovy, E. et al. (Eds.) **Multilingual Information Management: Current Levels and Future Abilities**,

1999. Disponível em <<http://www.cs.cmu.edu/~ref/mlim/>>. Acesso em 1 abril de 2005.

PERINI, M. A. **Gramática descritiva do português**. São Paulo: Editora Ática, 1998.

POLLARD, C.; SAG, I. **Head-driven phrase structure grammar**. Chicago: University of Chicago Press, 1994.

SAINT-DIZIER, P., VIEGAS, E. **Computational lexical semantics**. Cambridge: Cambridge University Press, 1995.

SANFILIPPO, A. Lexicons for constraint-based grammars. In: Cole, R. A (Ed). **Survey of the state of the art in human language technology**. Oregon: Graduate Institute, p. 118-121, 1995.

SAUSSURE, F. de. **Curso de lingüística geral**. 2a. ed. Tradução de A. Chelini et al. São Paulo: Cultrix/Editora da USP, 1972.

WILKS, Y. D. et al. Machine tractable dictionary as tools and resources for natural language processing. In: **Proceedings of Colling'88**, p. 750-55, 1988.