

4

A proposta do Léxico Gerativo e a análise semântica dos compostos nominais

Neste capítulo, serão apresentados a Teoria do Léxico Gerativo (Pustejovsky, 1995), assim como os desdobramentos dessa proposta presentes em trabalhos desenvolvidos no âmbito da Linguística Computacional e em um estudo mais recente de Beatrice Warren (2003). Em seguida, mostrar-se-á como os termos técnicos são tratados dentro do modelo do léxico gerativo.

4.1 A Teoria do Léxico Gerativo

A Teoria do Léxico Gerativo, cujo principal pesquisador é James Pustejovsky, surgiu na década de 1990 como uma tentativa de cobrir vários fenômenos semânticos de que as teorias das décadas de 1960/70 não deram conta. Segundo Chishman (2002:51), “[l]evando-se em conta que o método tradicional de organizar uma descrição linguística de um léxico prevê uma enumeração estática, a proposta de se explorar a geratividade do léxico já é merecedora de atenção”.

Pressupor que um léxico é gerativo é justamente ir contra toda uma tradição linguística de que os significados das palavras são obtidos através de técnicas enumerativas do significado. Segundo Pustejovsky (1995), há três argumentos básicos que mostram a inadequação do modelo de enumeração semântica: o uso criativo das palavras (o significado se dá no contexto); a permeabilidade dos sentidos (o significado das palavras remete a outros significados e palavras), e a expressão das múltiplas formas sintáticas (um único significado pode ter várias formas sintáticas).

Sendo assim, os principais problemas para a semântica lexical são:

- 1- explicar a natureza polimórfica da língua;
- 2- caracterizar a semanticidade dos enunciados da língua natural;
- 3- capturar o uso criativo das palavras em novos contextos;
- 4- desenvolver uma representação semântica cocomposicional mais rica.

A Teoria do Léxico Gerativo tem como objetivos capturar a criatividade lexical e a extensão de sentido, através de uma representação formal; descrever como as expressões das línguas naturais têm conteúdo e como este conteúdo pode ser modificado e modulado em novos contextos. Sendo assim, Pustejovsky explora duas linhas de investigação: (1) Como podemos acionar um número finito de palavras em nossa língua em um número ilimitado de contextos?; (2) As informações lexicais e suas representações na composição de significados são separáveis do nosso conhecimento?

Para dar conta de explicar esses fenômenos, é necessário um modelo lexical que associe relações semânticas com mecanismos que capturem a relação entre essas representações e a expressão sintática. Segundo Pustejovsky (1995:8), “[i]t is the goal of any lexical semantic theory to adequately classify the lexical items of a language into classes predictive of their syntactic and semantic expression”³⁹.

O autor propõe, então, um sistema contendo ao menos quatro níveis de representação semântica (1995:58):

the notion of argument structure, which specifies the number and type of arguments that a lexical item carries; an event structure of sufficient richness to characterize not only the basic event type of a lexical item, but also internal, subeventual structure; a *qualia* structure, representing the different modes of predication possible within a lexical item, and a lexical inheritance structure, which identifies how a lexical structure is related to the other structures in the dictionary, however it is constructed⁴⁰.

A interpretação composicional das palavras dentro de um contexto é obtida a partir da associação (através de dispositivos gerativos) dos quatro níveis citados acima.

Dentre os quatro níveis, a estrutura *qualia* parece ser a mais adequada para se estudar a semântica dos nominais. Dessa forma, atendo aos propósitos deste trabalho, será interessante focar a atenção apenas nesta estrutura.

“What *qualia* structure tells us about a concept is the set of semantic constraints by which we understand a word when embedded within the

³⁹ “o objetivo de qualquer teoria semântica lexical é agrupar adequadamente os itens lexicais de uma língua em classes que indiquem sua expressão sintática e semântica”.

⁴⁰ “a noção de uma estrutura argumental, que especifica o número e o tipo de argumentos que um item lexical carrega; uma rica estrutura de evento para caracterizar não apenas o tipo de evento básico de um item lexical, mas também uma estrutura subeventual interna; uma estrutura *qualia*, que representa diferentes modos de predicação possíveis dentro de um item lexical, e uma estrutura de herança lexical, que identifica como uma estrutura lexical está relacionada a outras estruturas no dicionário, não importando como seja construída”.

language”⁴¹ (1995:86). Ainda, “[the qualia] structure provides the jumping off point for operation of semantic reconstructions and type change, which in turn contribute to our overall goal of characterizing a natural language as polimorphic”⁴². (1995:77).

A estrutura *qualia* especifica quatro aspectos essenciais do significado de uma palavra:

- Constitutivo: a relação entre um objeto e suas partes constituintes (material, peso e partes, ou elementos);
- Formal: distinção de um objeto em um domínio maior (orientação, magnitude, formato, dimensão, cor e posição);
- Télico: propósito e função de um objeto (o objetivo do agente ao executar uma ação, função ou objetivos inerentes que especificam determinadas atividades);
- Agentivo: fatores envolvidos na origem de um objeto (criador, artefato, tipo natural e sequência causal).

De acordo com Pustejovsky (1995:85), os quatro fatores acima guiam a interpretação básica de um objeto ou de uma relação no mundo. Este método foi baseado, em parte, no trabalho de Moravcsik (1975), que viu os fatores gerativos (*aitia*) de Aristóteles como um sistema de compreensão construtiva e de inferência.

O próprio Pustejovsky afirma que, à primeira vista, a estrutura *qualia* pode parecer uma simples listagem de papéis casuais associados a itens lexicais; no entanto, ela é muito mais do que isso, uma vez que revela as coerções semânticas por meio das quais compreendemos uma palavra em uma língua. “*Qualia* not only structure our knowledge of words, but also ‘suggest’ interpretations of words in context”⁴³ (1995:87).

A palavra “livro”, por exemplo, pode ter diferentes significados, ou melhor, diferentes papéis (estruturas *qualia*), dependendo do contexto em que está

⁴¹ “O que a estrutura *qualia* nos revela sobre um conceito é o conjunto de restrições semânticas pelo qual compreendemos uma palavra quando embutida em uma língua”.

⁴² “a estrutura *qualia* representa o ponto de partida para operações de reconstruções semânticas e mudanças de tipo, o que, por sua vez, contribui para o objetivo geral de caracterizar uma língua natural como polimórfica”.

⁴³ “As *qualia* não só estruturam nosso conhecimento em relação às palavras, como também ‘sugerem’ interpretações das mesmas dentro de um contexto”.

inserida. Se dissermos “O livro está na estante”, estaremos nos referindo a “livro” como objeto situado em um domínio maior, em que a preocupação é situar o objeto “livro”. Dentro das funções *qualia*, a palavra “livro”, neste contexto, estaria inserida na categoria formal. Já se dissermos “Estou lendo um livro”, estaremos nos referindo à informação e ao fato de que livros servem para ler. A palavra “livro”, então, estaria inserida, respectivamente nos papéis constitutivo e télico. Se dissermos “João escreveu um livro”, a ênfase está no criador do livro, portanto, no papel agentivo.

A cocomposição das palavras torna a relação entre os itens lexicais mais clara, possibilitando a criação de novos significados, de acordo com as combinações lexicais. “[C]onhecer um objeto envolve não somente a capacidade de identificar ou referir, mas também a de explicar como um artefato vem a existir ou para que ele é utilizado” Chishman (2002:57). Essa nova forma de pensamento vai contra toda uma tradição que acredita que o nome, por exemplo, serve basicamente para se referir a ou identificar algo no mundo real.

É importante ressaltar que a interpretação correta da palavra “livro” em determinado contexto depende também dos outros elementos da sentença. É por isso que a interpretação semântica é considerada composicional. Em “Estou lendo um livro”, a interpretação de livro é bidirecional. Não só o verbo “ler” seleciona a palavra “livro”, como também o objeto “livro” seleciona o verbo “ler”.

O autor defende que os quatro sentidos da palavra “livro” mostrados mais acima são relações inerentes e fazem parte do significado da palavra “livro”. Essa capacidade de um item lexical agregar múltiplos sentidos é chamada por Pustejovsky e Anick (1988) de Paradigma Conceitual Lexical ou Paradigma Léxico-conceitual (PCL), cuja denominação em inglês é *Lexical Conceptual Paradigm* (LCP). Nas palavras de Pustejovsky (1995:91), “the intuition behind the notion of an LCP is that there is something inherent in the semantics of a noun such that it is able to project any of three separate senses of the noun in distinct syntactic and semantic environments”⁴⁴. A noção de PCL foi posteriormente mais bem desenvolvida por Pustejovsky e Boguraev (1993), de modo a projetar o

⁴⁴ “a intuição por trás da noção de um PCL é que haja algo de inerente na semântica de um substantivo, de modo que seja possível projetar qualquer dos três sentidos separados de um substantivo em diferentes ambientes sintáticos e semânticos”.

comportamento sintático a partir da semântica dos substantivos caracterizados como um PCL.

Os PCLs são úteis para capturar as ambiguidades sistemáticas nas línguas. Além disso, eles possibilitam caracterizar um item lexical como uma metaentrada. A metaentrada é uma representação que permite dar conta das várias possibilidades de sentido que um item lexical pode assumir em diferentes contextos sintáticos e semânticos. Esta seria extremamente útil na tarefa de capturar as ambiguidades presentes na linguagem. Por exemplo, no caso de “livro”, ilustrado anteriormente, não teríamos significados distintos, mas sim expressões lógicas com diferentes aspectos em relação à metaentrada para a palavra “livro”. Os PCLs projetam os significados das palavras, sem alterar o número de entradas.

4.2 O estudo dos compostos nominais influenciado pela Teoria do Léxico Gerativo

4.2.1 Sob a ótica da Linguística Computacional

A partir da Teoria do Léxico Gerativo de Pustejovsky, foram desenvolvidos vários trabalhos, principalmente na área da Linguística Computacional, com o intuito de mostrar e explicar a aplicação desta teoria na geração de compostos nominais. Um dos objetivos das pesquisas desenvolvidas nesta área é tentar reproduzir o processamento da linguagem natural no computador, a fim de aprimorar os programas de tradução automática, os sistemas de buscas, e a recuperação de informação.

A estrutura dos grupos nominais complexos e sua interpretação semântica também desafiam os sistemas de processamento da linguagem natural. Segundo Johnston, Boguraev e Pustejovsky (1995:1), “[c]omplex nominals present one of the greatest challenges in semantics to date, particularly when viewed from the concerns of polymorphic expressiveness in natural language”⁴⁵. Os principais problemas enfrentados são identificar o composto dentre outros termos em um

⁴⁵ “Os grupos nominais complexos representam um dos maiores desafios à semântica, principalmente em relação à expressividade polimórfica nas línguas naturais”.

texto; interpretar as relações semânticas subjacentes e analisar os compostos sintaticamente (Lauer, 1994).

Sendo assim, a fim de dar conta da semântica dos nominais, é necessária a criação de um modelo que limite a necessidade de listar os compostos no léxico e que dê conta de explicitar a relação semântica entre os termos. Segundo os pesquisadores da área da Linguística Computacional, o modelo do léxico gerativo parece ser o mais adequado, uma vez que atrela as representações da semântica lexical a mecanismos que capturem a relação entre essas representações e a expressão sintática.

Até então, pesquisadores como Jespersen (1942), Marchand (1970), Levi (1978) e Warren (1978) foram bem-sucedidos em enumerar os diferentes subtipos de compostos nominais, embora não tenham conseguido encontrar ferramentas para desambiguar esses subtipos. Ademais, é importante frisar que a abordagem adotada por esses pesquisadores, segundo os autores, pressupõe a existência de uma semântica plenamente articulada para cada uma das palavras que compõem o composto nominal.

Enquanto Levi defende uma abordagem que mostre as relações semânticas na língua a partir de um modelo que assume regras transformacionais aplicadas a uma estrutura profunda que contenha primitivos semânticos, Johnston & Busa (1996-1997), com base em Pustejovsky (1995), defendem um tratamento composicional das construções do composto que limite a necessidade de listá-las no léxico. Se, por um lado, os CNs de Levi são gerados a partir da supressão ou da nominalização de predicados, os *noun compounds* (referidos assim por Johnston & Busa em *Cross-Linguistic Semantics for Complex Nominals in the Generative Lexicon*), sob a ótica da Teoria do Léxico Gerativo, são obtidos através da análise componencial dos termos e da utilização de esquemas estruturais. Vejamos um exemplo.

Se a teoria proposta por Levi, já apresentada no capítulo 2 fosse adotada, *bread knife* ‘faca para cortar pão’ seria o resultado da supressão do predicado FOR. Na estrutura profunda, haveria algo do tipo “a knife FOR (cutting) bread”. Segundo Levi (1978:98), na maioria das vezes, “the predicate FOR indicates only this: that a relationship of intent or purpose exists (or is believed to exist) between

the objects named by its two underlying arguments”⁴⁶. A autora afirma “believed to exist” porque os CNs formados a partir da supressão do predicado FOR são considerados vagos, e isso se deve à opacidade da preposição FOR. Sendo assim, o predicado FOR não pode fornecer um detalhamento maior do tipo de propósito envolvido na relação entre o núcleo e o modificador. Tal detalhamento, segundo ela, só poderia ser deduzido a partir de conhecimento dos dois elementos envolvidos no composto, além do conhecimento extralinguístico, que nos permite diferenciar uma estrutura plausível da não plausível dentro de determinado contexto.

Para mostrar a opacidade do predicado FOR, a autora utiliza os seguintes exemplos: *bug spray* ‘inseticida’ = *spray for bugs [to harm them]* e *pet spray* ‘spray antipulgas’ = *spray for pets [to help them]*. Como se pode ver, apesar da supressão do mesmo predicado, a interpretação dos CNs é oposta.

Levando-se em conta a pouca especificidade do predicado FOR, incluir o verbo *to cut* na paráfrase de *bread knife* seria apenas uma dedução, baseada no conhecimento que se tem da relação entre os dois elementos e do contexto.

Considerando a Teoria do Léxico Gerativo, as estruturas *qualia* dos substantivos de um composto fornecem a estrutura relacional que possibilita a interpretação composicional da modificação do núcleo pelo pré-modificador. Em relação a *bread knife*, a metaentrada do composto é representada da seguinte maneira⁴⁷:

$$\left[\begin{array}{l} \text{bread knife} \\ \text{TYPESTR} = [\text{ARG1} = x \text{ artifact_t} \text{ ool}] \\ \text{ARGSTR} \left[\begin{array}{l} \text{D - ARG1} = y \text{ bread} \\ \text{D - ARG2} = w \text{ human} \\ \text{D - ARG3} = z \text{ human} \end{array} \right] \\ \text{EVENSTR} = \left[\begin{array}{l} \text{D - E1} = e_1 \text{ transitio n} \\ \text{D - E2} = e_2 \text{ process} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = x \\ \text{CONSTITUTI} \quad \text{VE} = \{\text{blade, handle...}\} \\ \text{TELIC} = \text{cut_act} \quad (e_2 \cdot w \cdot z \cdot x) \\ \text{AGENTIVE} = \text{make_act} \quad (e_1 \cdot z \cdot x) \end{array} \right] \end{array} \right]$$

⁴⁶ “O predicado FOR indica apenas que uma relação de intenção ou propósito existe (ou acredita-se que exista) entre os objetos nomeados pelos dois argumentos subjacentes”.

⁴⁷ A metaentrada de *bread knife* foi reproduzida de Johnston & Busa (1997:9).

Esta metaentrada fornece as seguintes informações: na estrutura *qualia* de *bread knife*, tem-se CUT-act (e2, w, x, y). Isso indica que o ato de cortar é um evento do tipo 2, ou seja, é um processo. A ação de cortar envolve três argumentos: w, x e y, que são, respectivamente, *creature*; *artifact-tool*; *bread*.

Segundo Johnston & Busa (1996:8), o tipo de modificação presente em *bread knife* é denominado télica, uma vez que

the modifying noun relates to the purpose of the head noun. The preferred interpretation of this compound is that it is a knife which is used to cut bread. The fact that a knife is an object whose inherent purpose is to cut things is encoded by the predicate **cut_act** in the TELIC role. The function of the modifier bread is to specify the third argument of the **cut_act** relation⁴⁸.

Comparando-se *bread knife* a *butcher knife* ‘faca usada pelo açougueiro para cortar carne’ ou ‘faca de açougueiro’⁴⁹, por exemplo, percebe-se que, ao invés de se ativar um dos argumentos de cortar (no caso de *bread knife* é o pão), em *butcher knife* o que é especificado é o agente de cortar – o açougueiro.

Vejamos outros exemplos. Dentro da perspectiva da semântica gerativa, o composto *lemon juice* ‘suco de limão’ seria o resultado da supressão do predicado FROM (*juice from lemon*). Esses tipos de compostos geralmente têm o modificador como a fonte do núcleo, e, na maioria das vezes, essa fonte é um objeto “natural” (fruta, legume ou animal).

Sob a ótica da Teoria do Léxico Gerativo, a modificação presente no composto *lemon juice* seria classificada como agentiva dentro da estrutura *qualia*, uma vez que o modificador menciona a origem de seu objeto. A estrutura *qualia* de *lemon juice* é representada como:

$$QUALIA = \left[\begin{array}{l} \text{FORMAL} = x \\ \text{AGENTIVE} = \text{squeeze_act} (e_1, y, x) \end{array} \right]$$

⁴⁸ “O substantivo modificador se refere ao propósito do substantivo núcleo. A interpretação mais adequada para esse composto é a de que a faca é usada para cortar pão. O fato de a faca ser um objeto cujo propósito inerente seja o de cortar coisas está codificado pelo predicado **cut_act** no papel TÉLICO. A função do modificador “pão” é a de especificar o terceiro argumento da relação **cut_act**”.

⁴⁹ Repare que se quisermos dizer “faca do açougueiro”, ou seja, que determinada faca pertence a determinado açougueiro, a estrutura *qualia* ativada é a constitutiva, e a construção em inglês muda (*butcher’s knife*).

Enquanto *glass door* ‘porta de vidro’ dentro da teoria de Levi seria classificado como tendo a supressão do predicado MAKE (*door which is made of glass*), sob a perspectiva da Teoria do Léxico Gerativo, o mesmo composto teria um papel constitutivo, que determinaria parte da denotação do núcleo ou do material do qual é feito. Eis a representação da estrutura *qualia* do composto:

$$QUALIA = \left[\begin{array}{l} \text{FORMAL} = \text{hold}(y, x) \\ \text{CONSTITUTIVE} = \{x : \text{glass}\} \\ \text{TELIC} = \text{walk_through_act}(e_2, w, y) \\ \text{AGENTIVE} = \text{make_act}(e_1, z, x, y) \end{array} \right]$$

Em relação aos compostos resultantes do processo de nominalização do núcleo, do modificador ou de ambos, caracterizado por Levi, pode-se fazer um paralelo com o que Johnston & Busa (1996)⁵⁰ chamam de *telic event modifiers*. Os autores mencionam exemplos de modificadores que denotam eventos dentro do papel télico. Para uma análise mais completa, é necessário especificar a caracterização desse evento: se é uma atividade, um estado ou um resultado obtido.

Há ainda uma importante distinção, que é entre o processo de uma atividade e o resultado da mesma. De acordo com Abney (1987:74), “[p]rocess nominals denote actions/events and result nominals denote objects.”⁵¹ Há substantivos que indicam ou um processo ou um resultado e há outros que indicam ambos, como *examination*. Em [*examination of the students*] *will take several hours*, o substantivo indica uma ação, o ato de examinar. Já em [*the examination of the students was printed on pink paper*] indica um objeto; neste caso, uma prova.

Johnston & Busa (1996) afirmam que, para exemplos com o modificador deverbal, o papel télico dentro da estrutura *qualia* é bem mais complexo, uma vez que as estruturas télicas do núcleo e do modificador geram (individualmente) o papel télico do composto. Por exemplo, em *hunting rifle*⁵² ‘rifle de caça’, cujo

⁵⁰ Os autores não mencionam exemplos em que o núcleo seja o substantivo deverbal.

⁵¹ “os nomes que indicam processo denotam ações/eventos e os nomes que indicam resultado denotam objetos”.

⁵² Os autores propõem a seguinte interpretação para este composto: “a rifle which is used in its typical capacity (i.e. firing) for the purpose of performing the activity of hunting”.

modificador é um substantivo deverbal que caracteriza um processo, a estrutura *qualia* vem representada da seguinte forma:

$$QUALIA = \left[\begin{array}{l} \text{FORMAL} = x \\ \text{TELIC} = \text{activity_lcp} \\ \text{TELIC} = \text{hunt} (e_2^p, x, z) \\ \text{AGENTIVE} = \text{fire} (e_1^2, x, y) \end{array} \right]$$

Já Levi (1978) diria que esse mesmo CN seria primeiro o resultado de uma nominalização (*rifle for hunting*). Em um ciclo mais tardio, as regras do RDP usariam a nominalização como o *input* para a formação do CN através da supressão do predicado FOR. Dessa forma, em *hunting rifle* teríamos previamente uma nominalização do verbo *to hunt* e depois a supressão do predicado FOR [*rifle (used) for hunting*].

É importante mencionar que, assim como visto em Levi (1978), a Teoria do Léxico Gerativo considera a estrutura dos compostos recursiva, ou seja, no caso da última, a composição permite que os compostos sejam expandidos, formando estruturas mais complexas. Johnston & Busa (1996) afirmam que o que antes era considerado um fenômeno idiossincrático hoje pode ser capturado por simples esquemas composicionais.

Embora a Teoria do Léxico Gerativo seja munida de mecanismos que capturem a recursividade dos compostos nominais, Girju *et alii* (2005:480) apontam que “the interpretation problem becomes significantly more complicated for larger noun sequences (...) since both the modifier and the head noun can form noun compounds generating structural ambiguities”⁵³. Para interpretar os compostos com múltiplos modificadores, o primeiro passo seria utilizar o alinhamento à esquerda ou à direita; entretanto, os autores ressaltam que escolher o alinhamento mais provável não é tarefa fácil, uma vez que o agrupamento das palavras é governado por razões semânticas, mas não sintáticas. Quando há a possibilidade de haver o alinhamento à esquerda ou à direita, a desambiguação se dá apenas no contexto. Para o composto *car radio equipment*, há duas

⁵³ O problema da interpretação passa a ser consideravelmente mais complicado no caso de seqüências mais extensas de substantivos [...] uma vez que tanto o modificador quanto o núcleo podem formar compostos nominais, gerando ambiguidades estruturais.

configurações possíveis: [*car [radio equipment]*] ou [[*car radio*] *equipment*]. O primeiro teria a leitura “equipamento de rádio para o carro” e o segundo “equipamento para o rádio do carro”.

Para um falante não nativo, essa sutileza passa despercebida; no entanto, para fins tradutórios, as nuances são bastante importantes, principalmente em textos técnicos. Os compostos com múltiplos modificadores serão mais bem explorados nos capítulos 6 e 8.

Um outro fator importante e que mostra certa convergência entre a semântica lexical enumerativa e os modelos inspirados na Teoria do Léxico Gerativo é a constatação de que um mesmo núcleo não garante a mesma relação com o modificador (por relação subentende-se a utilização dos mesmos predicados de Levi ou “papéis participativos” de Warren), nem o mesmo modificador modifica igualmente diferentes núcleos, conforme já citado no capítulo 3.

4.2.2 A classificação mais recente de Warren (2003)

Nesta subseção, será apresentada a proposta mais recente de Beatrice Warren (2003), em que a autora aproveita alguns conceitos do modelo do léxico gerativo. Sendo assim, a análise proposta por ela é um pouco diferente da apresentada em sua tese de doutorado (1978) e já discutida no capítulo 3.

Em seu artigo intitulado *The role of links and /or qualia in modifier head constructions* (2003), Warren afirma que o grande interesse pelo modelo do léxico gerativo nos últimos tempos é um possível indício de que a teoria tenha dado grandes contribuições à linguística.

Talvez o grande sucesso do modelo proposto por Pustejovsky (1995) resida no fato de que ele tenha ratificado a opinião de muitos, porém sempre dando um passo à frente, ou seja, sempre trazendo informações novas. Assim como muitos semanticistas, o autor acredita que o significado seja negociável no contexto, e, no entanto, que a polissemia seja governada por regras. Além disso, ele afirma que o significado das palavras é componencial, mas rejeita a ideia de que o léxico seja uma simples lista de palavras. No lugar dessa simples lista, Pustejovsky propõe uma representação complexa, já apresentada brevemente na seção 4.1.

Segundo Warren (2003), a estrutura *qualia* parece atender os princípios da categorização humana, uma vez que os princípios básicos incluem: (i) características físicas (estruturas *qualia* CONSTITUTIVA e FORMAL); (ii) ontogenia (estrutura *qualia* AGENTIVA) e (iii) papéis potenciais (estrutura *qualia* TÉLICA).

Em relação aos estudos sobre a semântica dos compostos nominais, Warren aponta que todos eles apresentam as relações nominais através de “ligações implícitas”, ainda que elas sejam divergentes e representadas de maneiras diferentes (verbos, locuções verbais, preposições). Além disso, pode-se dizer que há determinadas relações padrão, ou seja, que aparecem em quase todos os estudos, como a ideia de constituição, pertença, propósito e local. Quando os compostos não se encaixavam nas relações propostas pelos autores, ou eles eram excluídos (como fez Levi em inúmeras situações), ou se estendia demais o escopo das relações (como fez Warren).

Com a Teoria do Léxico Gerativo, além de se pensar em relações padrão, também é possível pensar em relações não previstas e que surgem em um determinado contexto, já que a teoria prevê que, quando combinamos palavras, as estruturas representacionais e os mecanismos gerativos são ativados, de modo a gerar significados implícitos e não listados no léxico. (Warren, 2003:235). Vejamos um exemplo fornecido por Warren. Na sentença “Bob finished the novel”, há dois sentidos regulares: “Bob terminou de escrever o romance” ou “Bob terminou de ler o romance”. Na primeira leitura, a função agentiva é ativada na estrutura *qualia* (o romance existe porque alguém o escreveu) e, na segunda, a função télica é ativada (a finalidade dos romances é a leitura).

Fazendo um paralelo entre a classificação clássica de Warren (1978) e a influenciada pelas ideias de Pustejovsky, é possível identificar algumas mudanças. Na publicação de 1978, Warren previa seis grandes categorias, com inúmeras subdivisões. No artigo publicado em 2003, a autora também prevê seis grandes categorias, porém com algumas diferenças. O nome de algumas categorias mudou. A categoria CONSTITUTE passou a ser chamada de COMPOSITIONAL; BELONGING TO passou a ser denominada POSSESSIVE; LOCATION passou a ser LOCATIVE. A autora, entretanto, manteve os nomes das categorias PURPOSE e RESEMBLANCE. Já a categoria ACTIVITY foi

excluída e acrescentou-se a CAUSATIVE. Os “papéis participativos” passaram a ser chamados de “ligações implícitas”.

Diferentemente do que fez em seu primeiro trabalho, a autora também analisa as relações semânticas em compostos adjetivais e afirma que todas as relações encontradas nos adjetivos *bona fide*⁵⁴ também foram identificadas em compostos do tipo N + N (*moon surface* ‘superfície da lua’) ou Adj. não predicativo + N (*polar bear* ‘urso polar’). Já o contrário não pode ser dito. Ainda, a autora constatou que as relações encontradas em compostos em que o modificador era um adj. não predicativo (Adj. não predicativo + N) eram iguais às dos compostos N + N, o que comprova a função nominal desse tipo de adjetivo e corrobora a classificação dos grupos nominais complexos proposta por Levi e já discutida no capítulo 3.

As ligações implícitas dos adjetivos *bona fide* têm uma função diferente da dos substantivos. Segundo a autora, as ligações dos adjetivos são partes convencionalizadas do significado deles; elas não dependem do contexto toda vez que são ativadas. Por outro lado, parece que as ligações implícitas dos substantivos não são convencionalizadas e que se referem ao composto como um todo e não à parte dele. Além disso, essas ligações geralmente não são passíveis de serem listadas previamente, pois dependem do contexto.

Comparando-se a estrutura *qualia* proposta por Pustejovsky (1995) e as categorias semânticas contidas em Levi (1978) e Warren (1978, 2003), pode-se dizer que apenas quatro papéis da *qualia* dão conta da maioria das relações propostas pelas duas autoras, como mostra a tabela abaixo:

⁵⁴ Os adjetivos *bona fide* são os que funcionam como atributos (*sad girl*); como predicativos (*the girl is sad*); apresentam variação de grau (*very sad girl*) e permitem inserções (*a sad, anxious girl*). Exemplos retirados de Warren (2003:238).

PUSTEJOVSKY (1995)	LEVI (1978)	WARREN (1978)	WARREN (2003)
AGENTIVE	CAUSE	LOCATION	CAUSATIVE
	USE	LOCATION	— ⁵⁵
	FROM	LOCATION	LOCATIVE
CONSTITUTIVE	HAVE	BELONGING TO	POSSESSIVE
	MAKE	CONSTITUTE/ LOCATION/ ACTIVITY	COMPOSITIONAL
	ABOUT	CONSTITUTE	—
FORMAL	BE (metafórico)	RESEMBLANCE	RESEMBLANCE
	BE/ MAKE	BELONGING TO/ ACTIVITY	—
	IN	LOCATION	LOCATIVE
TELIC	FOR	PURPOSE	PURPOSE

Tabela 3 - Quadro comparativo estrutura *qualia*/ categorias propostas por Levi (1978) e Warren (1978; 2003)

A autora afirma que há uma relação entre as ligações implícitas e a estrutura *qualia*; resta saber se a estrutura *qualia* dá origem às ligações e estas, quando ativadas, se transformam em significados abstratos, que podem ser memorizados e se tornam independentes da estrutura *qualia*, ou se elas têm funções distintas, porém a mesma base conceitual.

Warren conclui que ambas revelam a relação que torna o modificador e o núcleo apenas uma unidade de referência. Além disso, as ligações e a estrutura *qualia* são ativadas quando as palavras são combinadas.

Em relação aos papéis da *estrutura qualia*, o papel FORMAL é o que mais dá margens a dúvidas. Conforme já destacado, o papel FORMAL remete à dimensão, ao tamanho ou ao formato de algum objeto. Por ser bastante abrangente, sua capacidade descritiva é um tanto questionável; sendo assim, o papel FORMAL merece maior atenção dos estudiosos. Para fins do presente trabalho, será tomada uma posição diferente da de Warren (2003) e considerar-se-á que o papel FORMAL cobre as noções de semelhança quando na comparação de uma propriedade física; de localização, por distinguir um objeto em uma dimensão maior e de especificação de um subtipo (modificador) em um domínio

⁵⁵ Não é possível afirmar que a autora não preveja as relações de USO, TÓPICO e de GÊNERO-ESPÉCIE. No entanto, em seu artigo de 2003, tanto essas relações quanto compostos que expressassem tais não foram mencionados. Dessa forma, optou-se por não classificá-las aqui.

maior, o núcleo (Exs: *tuna fish, pine tree*). *Tuna* é um tipo de peixe (atum) e *pine* é um tipo de árvore (pinheiro).

Warren (2003) reconhece que a Teoria do Léxico Gerativo abriu caminhos para explicar as regularidades semânticas encontradas em construções de modificação; para se distinguir entre a polissemia listável e a não listável e para fazer uma diferenciação entre a polissemia que surge quando se combinam palavras, e quando estas são relacionadas com referentes contextuais. No entanto, a autora aponta que ainda falta dar conta dos significados não prototípicos e literais não composicionais, ou seja, das metáforas.

Chishman (2003) também apresenta críticas de alguns teóricos ao modelo de Pustejovsky. Segundo a autora, por motivos diferentes, Fodor e Lepore (1998) e Françoise Gayral (1998) criticam a versão de composicionalidade. Para os dois primeiros,

a geratividade e complexidade lexical não são fatores necessariamente indissociáveis. Ao se manifestarem sobre a questão mais diretamente ligada à geratividade — a composicionalidade —, eles são categóricos: cada constituinte contribui com o seu conteúdo, e o efeito de um constituinte sobre os demais é absolutamente independente do contexto. (2003:64)

Já Gayral (1998) *apud* Chishman (2003), ao contrário de Fodor e Lepore, considera o princípio de composicionalidade adotado por Pustejovsky fraco demais, uma vez que a situação de enunciação e a intenção comunicativa são relevantes para se abordar a inferência. A autora acredita que as inferências são mais bem explicadas pragmaticamente do que através de modelos lógicos.

4.2.3 Os termos técnicos sob o olhar da Teoria do Léxico Gerativo

A análise dos compostos que figuram como termos técnicos merece destaque neste trabalho, uma vez que o *corpus* foi extraído de um texto técnico da área de engenharia elétrica/eletrônica. Sendo assim, nesta seção será apresentada uma caracterização dos compostos nominais técnicos realizada por pesquisadores da área da Linguística Computacional, baseada na Teoria do Léxico Gerativo.

Johnston, Boguraev e Pustejovsky (1995) afirmam que a terminologia técnica ocupa um lugar especial dentro da categoria dos compostos nominais, já que ao mesmo tempo em que é extremamente produtiva, aglomera conceitos

fundamentais de determinado domínio técnico em uma estrutura bastante concisa. Além disso, apesar de, em geral, os compostos nominais serem formados gerativamente, os CNs “técnicos”, ou seja, que denotam um termo técnico, normalmente não o são, por serem compostos de palavras técnicas, das quais os dicionários convencionais não dão conta.

Sendo assim, os autores afirmam que, apesar de os trabalhos de Levi (1978) e Warren (1978) serem bem-sucedidos em enumerar diferentes subtipos de compostos nominais, eles não oferecem procedimentos para tornar esses subtipos menos ambíguos. Ademais, “theses approaches presuppose the existence of a fully-articulated semantics for each of the words in a complex nominal, a situation that is not readily available for technical vocabulary”⁵⁶ (Johnston, Boguraev e Pustejovsky 1995:1). Dessa forma, os autores percebem a importância de se desenvolver e aplicar técnicas que possibilitem determinar a interpretação dos compostos nominais dentro do âmbito técnico, sem haver a necessidade de supor um domínio do conhecimento ou uma semântica pré-definida dos elementos que formam os compostos.

Os autores propõem uma análise componencial dos compostos nominais técnicos, cujo significado deve ser definido pelas relações e propriedades da composição. Tais relações e propriedades representariam parcialmente a estrutura *qualia* de um composto nominal técnico (TCN)⁵⁷.

O objetivo da análise dos compostos técnicos, segundo os autores, é capturar a estrutura interna dos compostos; a operação semântica da composição (a relação semântica); a ontologia local (que será explicada mais à frente) na qual os compostos técnicos se encaixam e a classificação cruzada das estruturas através de suas relações.

A composição entre um constituinte e outro envolve inúmeras relações semânticas. Por exemplo, no composto *hard disk* ‘disco rígido’, a função do modificador é especificar uma propriedade formal do núcleo. Já em *start-up disk* ‘disco de inicialização’, percebe-se que o modificador indica uma relação de propósito do núcleo. As operações semânticas podem ser restringidas pelo contexto.

⁵⁶ “essas abordagens pressupõem a existência de uma semântica totalmente articulada para cada uma das palavras em um grupo nominal complexo, uma situação que ainda não está prontamente disponível para o vocabulário técnico”.

⁵⁷ Sigla em inglês para *technical complex nominal*.

Para analisar os TCNs dentro de um *corpus* fechado⁵⁸, seria interessante separá-los em o que os autores chamam de “ontologias locais”. Ontologias locais são hierarquias que capturam subdomínios dos objetos relacionados, tendo em vista o conteúdo de um determinado texto. A função das ontologias é agrupar os TCNs que se referem a objetos ou entidades relacionadas. Por exemplo, em um manual de informática contendo os compostos *floppy disk* ‘disco flexível’ ou ‘disquete’, *hard disk*, *start-up disk*, *RAM disk* ‘disco de memória’, o tipo mais básico de ontologia seria *disk*. Os compostos listados acima teriam um significado mais especializado que *disk* (pois seriam hipônimos do mesmo) e, portanto, teriam um uso mais restrito. A discussão sobre ontologias será retomada no capítulo 8.

Conforme já mencionado na subseção 4.2.1, antes de identificar a operação semântica entre os constituintes do composto, é necessário determinar a estrutura interna, ou seja, é importante saber quem modifica o quê. Quando se trata de um composto com dois constituintes, a tarefa é fácil. Em geral, o primeiro elemento modifica o segundo. Já quando há múltiplos modificadores, é necessário estabelecer a distribuição dos termos. Por exemplo, se em um texto técnico houver *key chain access code* ‘código de acesso da sequência de chaves’ figurando como um único termo e *access code* como outro, supõe-se que *access code* forma um termo e *key chain* outro, e que ambos compostos são unidos para formar um outro composto mais específico (*key chain access code*).

Identificada a estrutura interna dos compostos, é possível agrupá-los em ontologias locais. Os CNs *floppy disk*, *hard disk*, *RAM disk*, por terem como núcleo *disk*, seriam categorizados na ontologia local *disk*. Após a identificação do núcleo, estabelece-se a estrutura *qualia* de *disk* dentro do universo dos textos técnicos, que fornece o seguinte conjunto: *access*; *erase*; *name*; *initialize*; *use*; *test*; *save file on*; *repair*⁵⁹. Este conjunto define o que todos os discos têm em comum e, portanto, caracteriza *disk* no contexto técnico e forma o que se denomina *core qualia*. A *core qualia* mostra as relações comuns entre todos os membros de uma ontologia local.

Além da *core qualia*, há ainda uma estrutura mais especializada, a *discriminant qualia*, que elencaria as características que diferenciam um CN do

⁵⁸ Um exemplo de *corpus* fechado seria um manual de instrução. Um *corpus* fechado teria artifícios bem definidos, de modo a fornecer um texto com intenções previamente delineadas, diferentemente do *corpus* aberto, que poderia ser alterado livremente.

⁵⁹ acessar; apagar; nomear; iniciar; usar; testar; salvar arquivo em; consertar.

outro dentro da ontologia local. Por exemplo, *floppy disk* teria em sua estrutura *discriminant qualia* as qualidades: *eject*, *protect*, *insert*, *into disk drive*, *lock*, *unlock*⁶⁰, que *RAM disk* e *hard disk* não teriam.

Para determinar a semântica de *floppy disk* e *hard disk*, por exemplo, comparam-se os conjuntos de relações (tanto as comuns quanto as divergentes) entre um composto e outro. Após elencar as similaridades e diferenças entre as relações de um composto e de outro, representam-se as diferenças entre os membros da ontologia local através da estrutura *qualia* dos modificadores *floppy* e *hard*. São essas características diferentes que distinguem um disco do outro no caso de *floppy disk* e *hard disk*. As formas a seguir, reproduzidas de Johnston, Boguraev e Pustejovsky (1995), fazem parte da entrada para *disk* e fornecem as interpretações *floppy disk* e *hard disk*, respectivamente.

$$\begin{array}{l}
 \textit{QUALIA} = \left[\begin{array}{l} \text{modifier - lcp} \\ \text{dq} = \{ \text{eject, protect,} \\ \text{insert into, diskdrive,} \\ \text{lock, unlock} \} \end{array} \right] \\
 \\
 \textit{QUALIA} = \left[\begin{array}{l} \text{modifier - lcp} \\ \text{dq} = \{ \text{attach to printer,} \\ \text{reinstallsystem} \\ \text{softwareon,} \\ \text{copyprogramdiskto} \} \end{array} \right]
 \end{array}$$

No presente trabalho, pretende-se utilizar a noção de ontologia de modo a verificar se núcleos ou modificadores iguais implicam relações semânticas iguais. Foge ao escopo da pesquisa definir a *discriminant qualia* de cada composto inserido em uma mesma ontologia, por exemplo.

⁶⁰ ejetar; proteger; inserir; na unidade de disco; bloquear; desbloquear.