

Universidade Federal de Santa Catarina
Centro de Comunicação e Expressão
Programa de Pós-graduação em Linguística

RODRIGO PANCHINI AK FERNANDES

A Integração dos Papéis *Qualia* para Redes Semânticas

Volume Único

Florianópolis
2006

RODRIGO PANCHINIAK FERNANDES

A Integração dos Papéis *Qualia* para Redes Semânticas

Volume Único

Dissertação apresentada
como requisito parcial para
a obtenção de grau de
Mestre em Linguística, pela
Universidade Federal de
Santa Catarina, Centro de
Comunicação e Expressão,
Programa de Pós-graduação
em Linguística

Orientador: Prof. Dr.
Heronides Maurílio de Melo
Moura

Florianópolis
2006

Ficha catalográfica elaborada por: Fernanda de Sales, CRB 14/643.

F363i FERNANDES, Rodrigo Panchiniak.

A Integração dos papéis *qualia* para redes semânticas / Rodrigo Panchiniak Fernandes. –
2005.
97 f. il.

Orientador : Heronides Maurílio de Melo Moura.

Dissertação (Mestrado em Linguística) – Centro de Comunicação e Expressão – Universidade
Federal de Santa Catarina, 2005.

Bibliografia: f. 79-85.

1. Estrutura *qualia*. 2. Redes semânticas. 3. Regras de inferência.

I. Heronides Maurílio de Melo Moura. II. Universidade Federal de Santa
Catarina. Centro de Comunicação e Expressão. III. Título.

CDU 801.541.2

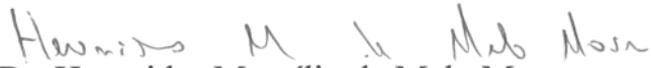
INTEGRAÇÃO DOS PAPÉIS QUALIA PARA REDES SEMÂNTICAS

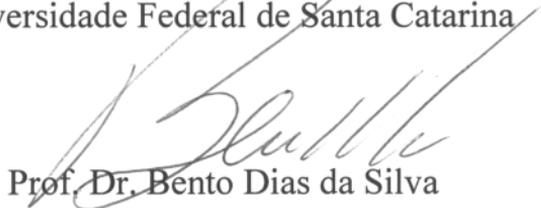
RODRIGO PANCHINIAK FERNANDES

Esta dissertação foi julgada adequada para obtenção do grau de Mestre em Linguística e aprovada em sua forma final pelo Curso de Pós-graduação em Linguística da Universidade Federal de Santa Catarina.

Prof. Dr. Fábio Luiz Lopes da Silva
Coordenador do Programa de Pós-graduação em Linguística da Universidade
Federal de Santa Catarina

BANCA EXAMINADORA:


Prof. Dr. Heronides Maurílio de Melo Moura
Orientador
Universidade Federal de Santa Catarina


Prof. Dr. Bento Dias da Silva
Universidade Estadual Paulista "Júlio de Mesquita Filho"

Prof. Dr. Celso Reni Braidá
Universidade Federal de Santa Catarina

Florianópolis, dezembro de 2005

À Izidora Panchiniak Fernandes (in memoriam)

Agradecimentos

Muitas das idéias expressas neste trabalho começaram a ser pensadas na graduação em filosofia na Universidade Federal de Santa Catarina, e foram lapidadas durante o mestrado em lingüística. Por isso não podemos nos abster de mencionar, inicialmente, a felicidade de termos contado com a orientação dos professores Luis Felipe Belintani Ribeiro e Heronides Maurílio de Melo Moura. Ao primeiro devem-se vários semestres de aprendizado do idioma grego, cujos labirintos gramaticais ajudaram a despertar o interesse sobre os fonômenos da linguagem. Ao segundo os golpes de lapidação. Também gostaríamos de agradecer às idéias trocadas em várias conversas amistosas com os professores Celso Braidá e Marco Rocha. Muito do entusiasmo necessário para a conclusão deste trabalho veio da contribuição informal destes professores.

Dos colegas de pós-graduação na lingüística teremos sempre a lembrança da calorosa acolhida, plena de generosidade intelectual representada, sobretudo por Magdiel Medeiros de Aragão Neto e Maria Leonor dos Santos. Deve-se a eles a possibilidade de nossa, menos traumática, imersão ao complexo mundo da lingüística.

Também é indispensável mencionar a colaboração, ainda que tolhida pela distância oceânica que nos separa, da professora Palmira Marrafa (Universidade de Lisboa), pelas suas sugestões e comentários quando da defesa da qualificação do projeto que levou à presente dissertação. E, mais recentemente, da professora Christiane Fellbaum (Universidade de Princeton), pela espontânea solicitude ao nos enviar importantes referências bibliográficas. O professor Bento Dias da Silva (Universidade Estadual de São Paulo) nos fez a bondade de prontamente ceder uma amostra de *synsets* da Wordnet, em curso, do português brasileiro.

Olavo L. Silva Filho (UNB) nos brindou com o envio de *Inferências Lexicais e Interpretação-rede de Predicados*, cujo programa teórico nos parece tão promissor quanto

desejamos que venha a ser o presente trabalho.

Gostaríamos de agradecer, ainda, às pessoas que participaram do seminário Integração das Relações Lexicais, no qual expusemos algumas das idéias centrais desta dissertação. À secretária do Programa de Pós-graduação em Linguística, Simone, pela não-exigível constante simpatia e ao professor coordenador do programa de pós-graduação Fábio Lopes.

Os amigos Gabriel Garcia e Carlos Eduardo Rodrigues não apenas fizeram a gentileza de ouvir algumas das esotéricas idéias deste trabalho como, também, sugeriram outras. Assim, pudemos nos aproveitar de seus valiosos conhecimentos em informática e matemática.

À Verinha e à família, o apoio incondicional.

A CAPES prestou, parcialmente, suporte financeiro por meio da bolsa de mestrado.

Como é praxe dizer, quaisquer defeitos ou incorreções não se remetem à contribuição das pessoas acima mencionadas, mas são de exclusiva responsabilidade do autor.

A forma mais baixa de pensamento consiste no mero reconhecimento do objeto. A mais elevada, na intuição compreensiva do homem que vê todas as coisas como parte de um sistema.

Platão

Resumo

Este trabalho se desenvolve ao redor de dois eixos: o teórico, no qual se discutem alguns temas de ontologia filosófica e semântica lexical, como a individuação e a polissemia, o nominalismo e o realismo; e o técnico-prático, no qual se estabelece um conjunto de regras de inferência para ontologias computacionais, que chamaremos de *QualiaNet*, cujo objetivo é a integração das relações da estrutura *qualia* do léxico gerativo às árvores de inferência ou hierárquicas. São propostas regras como a seguinte: o constitutivo do formal de X é formal do constitutivo de X. Por fim, aplicamos esta mesma regra a uma parte das relações da Wordnet 2.0, por meio das fontes em PROLOG, e demonstramos que a *QualiaNet* não apenas incrementa coerência entre as relações semânticas como, também, pode ser utilizada para gerar novos arcos hiperonímicos, os quais são, como se sabe, a espinha dorsal de qualquer rede semântica.

Palavras-chave: estrutura *qualia*, redes semânticas, regras de inferência.

Abstract

This work is developed under two main axes: theoretical, where we discuss some relevant aims of philosophical ontology and lexical semantics, as individuation and polissemey; end the technical-practical, where we establish a set of inference rules to computational ontologies, which will be called *QualiaNet*, whose aim is the integration of generative lexicon *qualia* structure relations to inheritance or hierarchical semantic trees. The proposed rules are similar to the following: the constitutive of the X formal is the formal of X constitutive. At last, we apply this same rule to a subset of Wordnet 2.0, through PROLOG sources, and have demonstrated that *QualiaNet* does not only increase the cohesion between semantic relations as, too, may be utilised to generate new hyper/hyponymic arrows, which are, as is well known, the main component of any semantic net.

Key-words: *qualia* structure, semantic net, inference rules.

SUMÁRIO

INTRODUÇÃO	11
1 ONTOLOGIA	13
1.1 OS LIMITES ENTRE A ONTOLOGIA E A FILOSOFIA	13
1.1.1 O principal interesse da ontologia: o que há?	22
1.1.2 Nominalismo e realismo.....	23
1.2 CATEGORIAS	26
1.3 O SENTIDO GRAMATICAL DOS PAPÉIS <i>QUALIA</i>	32
1.4 THESAURI	35
1.4.1 <i>Thesaurus</i> de Roget.....	40
1.4.2 As Wordnets	42
1.4.2.1 Pressuposições teóricas da Wordnet.....	43
1.4.2.2 Organização dos <i>synsets</i> e outras questões de semântica relacional.....	44
1.5 A ONTOLOGIA NA COMPUTAÇÃO: A SEMANTIC WEB E O SISTEMA UNL	57
2 REDES SEMÂNTICAS	61
2.1 RELAÇÕES SEMÂNTICAS	61
2.2 OS PAPÉIS <i>QUALIA</i> COMO RELAÇÕES SEMÂNTICAS E AS SÉRIES PROPORCIONAIS	64
2.3 INTEGRAÇÃO DAS RELAÇÕES SEMÂNTICO-CONCEITUAIS	69
2.4 PRODUÇÃO AUTOMÁTICA DE RELAÇÕES NA WORDNET 2.0 POR MEIO DE QF1	72
2.5 O PROBLEMA DA COMPATIBILIDADE ENTRE CO-HIPÔNIMOS: O CASO DO CÃO POLICIAL	74
CONSIDERAÇÕES FINAIS	76
APÊNDICES	78
1.1 APÊNDICE I: TABELA DE CATEGORIAS.	78
1.2 APÊNDICE III: CATEGORIAS DE ROGET.	78
1.3 APÊNDICE IIII: CATEGORIAS DE ROGET (CONT.).....	78
1.4 APÊNDICE IIII: PARES DE HIPÔNIMOS/HIPERÔNIMOS.....	78
1.5 APÊNDICE IIII: ANÁLISE E CONTRA-EXEMPLOS.	78
REFERÊNCIAS	79

Introdução

Nesta dissertação abordaremos temas de interesse para as grandes áreas da lingüística, da filosofia e da ciência da computação. De um modo específico, trataremos de questões de semântica lexical e léxico gerativo, ontologia e filosofia da linguagem, sistemas especialistas, inteligência artificial e processamento automático de linguagens naturais. Esses temas se entrelaçam assim como, em alguns momentos, se entrelaçam as três grandes áreas citadas.

Nosso principal objetivo é discutir e apresentar um constructo teórico que chamaremos de *QualiaNet*, o qual integra relações lexicais utilizadas em sistemas de herança lexical, em especial nas Wordnets (Cap. 2, seção 3).

A *QualiaNet* consiste na integração das quatro relações lexicais, estrutura *qualia*, do léxico gerativo, e é desenhada para ser implementada juntamente ou paralelamente a sistemas de conhecimento lexical estáticos como as Wordnets (Cap. 1, seção 4.1) ou a *Universal Networking Language* (UNL), ou sistemas dinâmicos como as ontologias associadas à *Semantic Web* (Cap. 1, seção 5). Na hipótese de ser implementada juntamente a estes sistemas, a *QualiaNet* pode ser utilizada simplesmente para enriquecer o repositório de informações semânticas das redes. No caso da implementação paralela, pode prover independentemente um sistema especialista orientado a agentes inteligentes, na construção de inferências para a simulação lingüístico-cognitiva. Exploraremos apenas a primeira hipótese.

Além do objetivo técnico de prover um conjunto de regras de inferência para o tratamento das informações semânticas das linguagens naturais (Cap. 2), esta dissertação trata de discussões teóricas acerca dos assuntos de interesse: filosofia da linguagem, ontologia e semântica lexical (Cap. 1). Neste sentido, discutiremos questões de filosofia da linguagem, principalmente sob o viés da filosofia analítica. No que diz respeito à ontologia, buscaremos aproximar os *problemas*

ontológicos gerais às questões sobre a construção computacional de ontologias (Cap. 1, seção 2); no viés semântico, iremos investigar o problema da compatibilidade entre co-hipônimos e a contribuição que a estrutura *qualia* pode dar a este problema (Cap. 2, seção 5).

Do ponto de vista do método nos utilizaremos freqüentemente da prática, usual em semântica lexical e filosofia da linguagem, que consiste em avaliações da semanticidade e gramaticalidade de expressões lingüísticas, com base no julgamento subjetivo. Ao invés de nos utilizarmos de um *corpus* convencional, ao final apresentamos exemplos de relações semânticas automaticamente extraídas da Wordnet 2.0, a qual, de certa forma, também é um *corpus*, pois consiste em um repositório expressivamente grande de informações lingüísticas.

1 Ontologia

Este capítulo abordará algumas das várias discussões a respeito da ontologia. Na seção 1, questionamos o significado da ontologia conforme sua abordagem mais tradicional, ou filosófica. Neste aspecto iremos frisar que a palavra “ontologia” muitas vezes se confunde com “filosofia”, uma vez que ambas, segundo certos pontos de vista, se direcionam ao mesmo objeto de pesquisa, a saber, o ente. Outros debates são levantados, como o nominalismo vs. realismo (seção 2), conforme o roteiro que adotamos sugerido em Moravcsik (1998). Na seção 4 discutiremos os *thesauri* e o lugar destes diante da ontologia tradicional, ou filosófica. O que os *thesauri* representam diante da *pergunta ontológica*? É a questão que pretendemos abordar. Na seção 5, levantaremos aspectos da recente utilização dos conceitos ontológicos pela ciência da computação, enfatizando projetos como a Semantic Web, as Wordnets e a UNL.

1.1 Os limites entre a ontologia e a filosofia

Talvez a definição mais trivial de ontologia, no campo dos estudos filosóficos, seja a de “estudo do ente” (ABBAGNANO, 1998, p. 720) com a inevitável cacofonia que a expressão produz. Neste sentido, a ontologia é generalíssima assim como o é a própria filosofia segundo a visão pré-moderna, que de algum modo persiste até hoje, e de acordo com a qual todos os campos de conhecimento são filosofias. *En passant*, a efetividade histórica desta classificação do conhecimento deixa-se ver em títulos como o Princípios de Filosofia Natural, do paradigmático tratado de física de Isac Newton, e das Gramáticas Filosóficas do séc. XIX (Cf. BARBOSA, 1871).

Do ponto de vista da discussão acadêmica contemporânea, podemos falar de duas escolas ontológicas: aquela de cunho metafísico-histórico, e a de tendência lógico-positivista (BEZIAU, 2003). Estes dois rótulos não estão isentos de imprecisões e simplificações possivelmente

tomadas como injustas para os adeptos das duas correntes, no entanto, têm a sua utilidade. Do ponto de vista metafísico-histórico, são assuntos de interesse ontológico a dicotomia entre ser e não-ser, ou simplesmente o nada; o tempo; o discurso; a razão ou o *logos*. A fenomenologia e o existencialismo são escolas de base desta corrente, e Husserl, Heidegger, Sartre, autores de destaque. Do ponto de vista lógico-positivista a ontologia preocupa-se com questões acerca da possibilidade e necessidade; da realidade; da verdade e da universalidade. O formalismo, o intuicionismo e o empirismo lógico são as escolas de destaque e Frege, Russell, Carnap, alguns autores paradigmáticos. Novamente aqui procedemos com simplificação, mas aceitaremos a indulgência do leitor. De um modo específico, esta separação entre duas correntes não é precisa, porque os assuntos de uma e de outra escola se entrelaçam segundo o ponto de vista de um observador razoavelmente imparcial. Um ponto de vista que tenda à imparcialidade parece ser conveniente tanto quanto os adeptos de uma ou de outra escola, comumente, tenham dificuldade em aceitar a legitimidade filosófica dos membros da escola opositora. Uma outra diferença, não de conteúdo, mas de método, pode ser a chave desta incompatibilidade de gênios. Tratamos disto logo abaixo.

De um lado a corrente metafísico-histórica utiliza-se da análise filológica como meio de justificação para suas proposições, de outro lado os lógico-positivistas sentem-se endossados pelas formalizações lógico-matemáticas.

Um exemplo trivial da análise filológica empregada pelos metafísico-históricos é a interpretação de determinado conceito expresso por uma palavra em língua portuguesa, digamos, *razão*, à luz dos conceitos correspondentes, que outras línguas, preferencialmente clássicas, e em especial o grego, expressam por palavras de diferente polissemia. Desta forma se faz uma genealogia dos conceitos, ou, no mínimo, se buscam inspirações em outras línguas para problemas de significado da língua materna. Esta prática também é importante para a chamada

história dos conceitos, uma vez que busca elucidar, comparar e *incorporar* determinadas interpretações dos termos empregados por pensadores e povos antigos aos termos habituais contemporâneos. Ainda que isto possa parecer impróprio para olhares puristas, há uma constante preocupação com a diacronia dos conceitos filosóficos. Assim, a palavra grega correspondente à razão, *logos*, é também a palavra grega correspondente a discurso e linguagem. Podemos interpretar a própria categoria ontológica da razão como uma mescla entre as categorias do discurso e da linguagem, algo como, razão-discurso-linguagem, em algum dos sentidos contemporâneos de razão. Em geral, este método pretende a síntese de vários sentidos primitivos para a clarificação de um sentido composto e corrente.

Na outra margem, um exemplo da análise lógico-matemática é a empregada por Russell (1872 – 1970), na decomposição das descrições definidas. Russell, seguindo Frege (1848 – 1925), irá decompor uma expressão, digamos, “bela rosa vermelha”, como uma referência aos objetos que são simultaneamente belos e vermelhos e que recebem o predicado de serem rosas. Isto, segundo a definição para o termo, ou o conjunto de sinônimos, mais apropriados para a palavra rosa. Sendo que o conceito de definição, utilizado neste raciocínio, é idêntico ao conceito de definição lexicográfico. Por exemplo, como definição de rosa teríamos “flor da roseira”¹. O importante, porém, é que as relações entre as classes, “bela”, “vermelha” e “flor da roseira” estejam suficientemente inequívocas por meio da utilização de operadores lógico-matemáticos, como os quantificadores, as conjunções e as disjunções. Assim, o significado das palavras é interpretado como uma relação entre o mundo e os conceitos, segundo determinados modos, ou caminhos pelos quais se dá esta ligação (o sentido).

Esta análise, passando da composicionalidade à teoria das descrições definidas, é utilizada para solucionar paradoxos semânticos (FERNANDES, 2004), causados por expressões que se

refiram a objetos inexistentes, como “O atual rei do Brasil é careca”. Na decomposição lógica desta expressão, teríamos: $\exists x (Rx \wedge Cx)$. Conforme um modelo segundo o qual R corresponde ao predicado unário “é rei do Brasil, atualmente”, C corresponde ao predicado também unário “é careca” e x corresponda a uma variável individual a ser substituída pela entidade que satisfaça as condições impostas pelos predicados. Deste modo, a negação desta proposição faria a negação do quantificador existencial². Somam-se a esta análise as minúcias técnicas da semântica de tabelas veritativas dos operadores lógicos, e da sintaxe para o estabelecimento de fórmulas bem formadas (*well formed forms*, wffs), que não nos interessam no momento. Apenas convém notar que este tipo de análise lógica da linguagem, desenvolvida em parte pelo sonho da Conceitografia de Frege, pretende limitar os objetos do mundo científico aos valores de verdade, ou os sentidos dos predicados aos conceitos.

Isto é o que representa o referente de uma proposição como “Florianópolis é uma ilha” ser o verdadeiro. Obviamente o esvaziamento de conteúdos da linguagem, mesmo da linguagem científica, não a torna capaz de lidar com outras matérias que não as relativas, exclusivamente, às ciências formais. E ainda neste campo surgem problemas.

A prática metafísico-histórica nos interessa menos do que a sua opositora para os fins desta dissertação. Afinal, ela se dá menos ao desenvolvimento de uma técnica de lingüística computacional do que o cálculo de predicados. A corrente metafísico-histórica é essencialmente hermenêutica. No entanto, tais reflexões sobre a linguagem estão longe de serem despropositadas,

¹ FERREIRA, Aurélio Buarque de Holanda. **Minidicionário Aurélio**. Rio de Janeiro: Nova Fronteira, 1988. p. 451.

² Existe uma diferença de tratamento para a negação de expressões que possam pressupor a existência de objetos inexistentes, entre Frege e Russell. Russell admite a possibilidade da negação do quantificador existencial, o que impede o pressuposto de existência (RUSSELL, 1974). Por outro lado, Frege nega esta possibilidade através do conceito de representação que cabe às expressões da literatura, entre outras (FREGE, 1978 [1884]). Transferindo o pressuposto de existência de objetos imaginários, em termos atuais, para a esfera das possibilidades do contexto pragmático (cf. MOURA, 2003).

inúteis, ou sem qualquer valor filosófico-científico³. Elas contêm não apenas intuições que tentam clarificar diacronicamente o sentido de determinados termos de interesse, mas, também, eventualmente, argumentos críticos, contra ou a favor, mas geralmente contra, as tomadas de posição teórica alcunhadas de positivistas.

Já a análise frege-russelliana, ainda que tenha representado um avanço paradigmático em relação às teorias precedentes, que não caminhavam muito além do silogismo aristotélico, esbarra em pelo menos uma dificuldade muito árida. A saber, o *problema da reprodução de descrições definidas ad infinitum*. Este problema está relacionado ao apelo a meta-meta-linguagens, igualmente *ad-infinitum*, quando se trata do estabelecimento dos fundamentos formais das linguagens⁴. O *problema da reprodução das descrições definitas* repousa sobre a abstenção da abordagem lógico-positivista em analisar os conceitos por outro viés que não o estritamente semântico (em oposição às discussões que levam em conta os aspectos pragmáticos). Esta abordagem orgulha-se de ser “desprovida de conteúdo”, universal e, portanto, livre das contingências próprias do mundo atual.

Os operadores matemáticos são, no entanto, entidades tanto puras quanto estéreis, dirão os críticos. Em alguma medida esta crítica se revela acertada. A análise de “bela rosa vermelha” esbarra no problema do denotacionismo, ou no de apelar ao denotacionismo, quando nos perguntamos sobre o significado de um dos termos da proposição. Tal que “bela” significa sua extensão, o conjunto de objetos que são designados por este termo. Outro problema é a substituição de uma expressão simples em linguagem natural por outra expressão, certamente mais complexa, na mesma linguagem, como uma paráfrase da primeira expressão. Ora, esta

³ Esta corrente nos inspira à adoção da postura filosófica de Silva Filho e Cabrera (2004), quanto à importância científica das chamadas lógicas divergentes, e lingüística de Moura (1999), quanto à conciliação semântico-pragmática.

⁴ Este problema é conhecido desde, pelo menos, a formulação do antigo tropo chamado de Terceiro Homem, (LAERTIUS, V 1977 [séc. IV]). Atualmente, o trabalho de Silva Filho e Cabrera (2004) estabelece uma posição em

segunda expressão não será outra que não mais uma descrição definida, ainda que fosse uma descrição definida o que desejássemos analisar inicialmente. Por outro lado, o denotacionismo implica o sempre inconveniente apelo ao mundo exterior, na linguagem do cartesianismo.

A outra dificuldade é a da produção de meta-linguagens de meta-linguagens *ad infinitum*. Isto fica claro quando se trata da axiomatização das teorias científicas. A escola lógico-positivista, de um modo geral, enxerga uma teoria científica como uma linguagem bastante imprecisa acerca de determinado domínio do conhecimento humano, embora tenha um rigor obviamente maior do que a linguagem do dia-a-dia. Assim, fala-se em axiomatização da física, da matemática e de outros campos de conhecimento (cf. KRAUSE, 2002). Sucintamente, esta axiomatização consiste em estabelecer o léxico, a sintaxe e a semântica apropriados como um modelo que sirva para descrever os fenômenos conforme vistos pela lente da teoria objeto. A linguagem pode conter regras de inferência básicas, como, no caso da física newtoniana, as três leis da mecânica. No entanto, do ponto de vista lógico-lingüístico, o modelo científico em questão não determina quais regras subjacentes à sua própria área estão sendo utilizadas. Um modelo de física clássica irá se utilizar, para ser construído como uma linguagem formal, não-ambígua, de um modelo aritmético-matemático. Este, por sua vez, irá ser construído sobre uma determinada lógica, que dividirá espaço com uma teoria de conjuntos ou, alternativamente, uma mereologia. Ora, tanto o modelo aritmético-matemático quanto o lógico-dedutivo não se encerram sobre si mesmos, e exigem, também, linguagens subjacentes que venham ao socorro de lacunas formais. E, diga-se de passagem, estas tentativas formalistas, por assim dizer, de reduzir um campo de conhecimento de elevado índice pragmático a outro campo de conhecimento mais semântico, *more geometrico*, não têm sido bem sucedidas sequer na área de conhecimento na qual se esperava grande facilidade para tal procedimento. A aritmética, segundo o teorema de

semântica filosófica que supomos capaz de superar estas dificuldades recorrentes.

Bernays-Goedel, não pode ser reduzida à lógica. Em termos técnicos, a linguagem da aritmética não é um teorema da linguagem do cálculo de predicados. Em outras palavras, nenhuma linguagem formal pode conter todas as definições de si mesma — lembrar do Paradoxo de Eubúlides —, o que exige a existência de uma segunda linguagem, uma meta-linguagem. No entanto, a meta-linguagem também não se definirá, exigindo uma meta-meta-linguagem e assim sucessivamente.

A importância destes problemas para o presente trabalho consiste na possibilidade de, senão resolvê-los, pelo menos amenizá-los, por meio de um instrumental teórico e prático que favoreça a resolução de inferências lexicais, do tipo:

(1)

“Herder assistiu pessoalmente aulas de Kant;

Kant nunca saiu de Königsberg;

Portanto, Herder esteve alguma vez em Königsberg”⁵.

As Wordnets ou outras ferramentas baseadas em semântica relacional, bem como a constituição de um léxico estruturado, podem ser vistas como uma valiosa ferramenta para a automatização da resolução destas chamadas inferências lexicais. Habitualmente, estas operações são consideradas materiais, pragmáticas e, portanto, “fora do escopo da lógica” (cf. Op. cit). O que se pauta, portanto, é a investigação da possibilidade do incremento do poder de informatividade semântica das redes conceituais.

Detenhamo-nos sobre este ponto.

Para explicitar estas questões podemos nos valer dos seguintes raciocínios. Assim como uma máquina para produzir tortas não pode, ela mesma, ser feita de tortas; uma máquina cujo

⁵ (SILVA FILHO; CABRERA, op. cit., p. 10).

objetivo fosse produzir conhecimentos sobre botânica não poderia, a princípio, constituir-se de tais conhecimentos. Mas seria razoável que fosse programada, desde o início, com conhecimentos acerca das possibilidades de aquisição do conhecimento natural, ou seja, como observar, como coletar, como classificar, etc. Funcionaria analogamente à máquina de fazer tortas, a qual precisaria estar apta a manipular a massa e acrescentar ingredientes. Este é o mesmo raciocínio que nos leva a aceitar a proposição de que um técnico em estatística precisa ter conhecimentos técnicos em matemática, mas, por outro lado, um técnico em matemática não precisa ter conhecimentos técnicos de estatística. Mesmo antes de Leibniz, considerado o pai da inteligência artificial, (Cf. BITTENCOURT, 1998, p. 53), são feitas tentativas de reduzir as formas de conhecimento a elementos tão anteriores que pudessem, bem arrajados, produzir os demais conhecimentos. Desde Os Elementos de Euclides estamos familiarizados com este tipo de análise, a tentativa de reduzir elementos complexos a elementos mais simples, com o objetivo de alcançar a compreensão criativa para a elaboração dos conhecimentos complexos. Esta é uma maneira de compreender a tentativa da lógica matemática de reduzir seu escopo às operações puramente formais. Ainda que se considere a lógica como disciplina nascida da argumentação, é razoável supor que tal disciplina não poderia ater-se sobre cada um dos possíveis objetos de argumentos. Mas, antes disso, que deveria se ater sobre as formas gerais de quaisquer argumentos, acerca de quaisquer conteúdos. O que levou Aristóteles (384 – 322 a.C.) à exaustiva classificação das formas dos silogismos. Dito isto, podemos entender porque a lógica matemática, clássica ou não-clássica, viu-se na tentativa de extrair qualquer nódoa de conteúdo dos procedimentos analíticos. O que levou, inclusive, à controversa assertiva do positivismo lógico, pautado no primeiro Wittgenstein (1889 – 1951), de reduzir o significado das expressões aos procedimentos de verificação das mesmas.

No entanto, a direção programática da pura analiticidade parece fadada a um certo

esterilismo: que se sintetiza na rejeição à consideração de quaisquer fatores pragmáticos. Para avançar nesta discussão, retornemos ao exemplo (1), de inferência lexical, apresentado acima.

Ab initio, percebemos que esta inferência depende de pelo menos uma premissa não explicitada. A saber, a de que Kant não se bilocava enquanto lecionava. Com efeito, a luz atinge vários pontos simultaneamente, sem que precise sair de um para estar no outro. Se Kant dispusesse de tal poder, estaria apto a lecionar fora da Europa, sem, entretanto, ausentar-se de Königsberg. Ainda que não cogitemos ceder tamanha capacidade ao autor de *Crítica da Razão Pura*, e a impossibilidade de bilocar-se de Kant pareça óbvia, não é tão óbvia ou necessária quanto as tautologias das Leis de Morgan (cf. MORTARI, 2001). Vejamos outro exemplo:

(2)

“Hiraldó é um mamífero.

Portanto, Hiraldó é um vertebrado.”⁶

Este exemplo é muito mais simples do que o de Herder, uma vez que é necessário pensar apenas na hiperonímia. No entanto, mantém a mesma relação de inferioridade epistemológica com as tautologias que a inferência sobre Herder mantém, em comparação com as Leis de Morgan.. Para usar a expressão de Leibniz, assim como podemos imaginar um *mundo possível* em que Kant podia se bilocar ao mesmo tempo em que lecionava, também podemos imaginar um mundo possível, ou mesmo um outro planeta, em que haja mamíferos invertebrados. E alguém lembraria dos terráqueos ornitorrincos, que são mamíferos ovíparos?

Mas, desde que a língua seja um sistema de valores que se opõem, um sistema de diferenças (cf. SOUSSURE, 1967), podemos assumir que, possa ou não Kant bilocar-se, sejam ou não os mamíferos vertebrados, dadas certas relações, outras relações se deduzem. As

⁶Idem p. 20.

possibilidades de estruturação de um encadeamento de relações constituem um dos principais temas da presente dissertação.

Apenas para não deixar dúvidas, a inferência (1) pode ser vista como uma dedução de certas propriedades dos elementos das premissas: ((Herder, Kant) -> (pessoas) -> (objetos físicos)); ((assistir) -> (predicado binário, aula) -> (evento espacio-temporal)); ((Koinigsberg) -> (espaço, etc.)). A maneira de organizar estas informações é altamente sofisticada, (cf. HARABAGIO; MOLDOVAN, 1998), e depende da forma como são constituídas as relações entre os termos. No entanto, parece inegável que as propriedades semântico-lexicais podem ser utilizadas para a extração de conhecimento por inferências.

1.1.1 O principal interesse da ontologia: o que há?

É possível detectar em ambas as correntes discutidas na seção anterior, analítica e hermenêutica, o pensamento de que a ontologia é a filosofia primeira em oposição a outros aspectos dos estudos filosóficos, como aqueles concernentes ao comportamento humano como ética ou filosofia política. Neste sentido, a ontologia é a atividade primordial da filosofia ou a filosofia propriamente dita, uma vez que não apresentaria aspectos de ciência aplicada. Daí o interesse tradicional entre os filósofos, como Kant (1724 – 1804), em classificar as áreas de conhecimento, mas não praticar nenhuma delas em especial. Porém, igualmente a outras classificações apresentadas acima, esta não pode ser tomada como absoluta. O mais natural, aparentemente, é que haja uma relação de determinações recíprocas entre estes campos, sejam de natureza mais aplicada ou de natureza mais pura.

Assim como tomadas de decisão teórica, acerca dos conceitos fundamentais da ontologia filosófica, podem direcionar posicionamentos em ética para este ou aquele caminho, o contrário

também é possível. De fato, estes dois fluxos de influências ocorrem⁷.

A expressão “estudo do ente” pode parecer demasiadamente hermética. Uma forma alternativa de introduzir o conceito de ontologia é apresentando a pergunta “o que há?”. No entanto, como resposta não basta a mera enumeração das entidades⁸. O interesse ontológico está em organizá-las conforme categorias. Se isto já não fosse tarefa suficientemente hercúlea, a filosofia contemporânea tem se dedicado mais à tentativa da resolução do problema da indecidibilidade do que à construção de uma táboa de categorias propriamente dita.

A indecidibilidade diz respeito às seguintes perguntas: quais são os critérios pelos quais se faz a ligação entre mundo e conceito? Como decidir a que categoria pertence determinado objeto? Este problema teve um grande impacto na tradição, a qual se voltou, desde a segunda metade do sec. XIX, quase que, exclusivamente, à resolução de questões dos fundamentos da lógica e da matemática.

1.1.2 Nominalismo e realismo

O debate entre nominalistas e realistas tem um paralelo não muito distante com aquele apresentado no Crátilo de Platão (428? – 348? a.C.), e por isso representa uma discussão cujas conseqüências metodológicas nos interessam. No diálogo citado questiona-se a natureza arbitrária, convencional, do signo lingüístico em oposição à possibilidade da sua gênese natural, espontânea. Há alguma relação de causalidade, uma motivação, entre o signo e o objeto significado? Ou esta relação é acidental? Assim se pode estabelecer um paralelo entre o nominalismo e o convencionalismo e entre o realismo e o naturalismo lingüístico. O nominalismo

⁷ Ver, por exemplo, a proximidade entre o utilitarismo ético de Stuart Mill (1773 – 1836) e o empirismo epistemológico de David Hume (1711 – 1766).

⁸ “Alguém pode querer simplificar a pergunta ontológica [O que existe?] respondendo-a com longas listas de tempos e objetos específicos. O filósofo, por outro lado, não deseja listas infinitas de causas primeiras, mas as categorias gerais da realidade”, (MORAVSIC, 1980, p. 3, tradução nossa).

Embora não pretendamos nos comprometer com o peso da expressão “reality”, o trecho acima dá o tom de nossa discussão.

de Okham (1300 – 1350) afirma que não há existência objetiva dos universais. Universais são os conceitos enquanto extrapolam a existência dos indivíduos particulares. Para usar o termo de Platão, enquanto há diversos homens particulares, Crátilo, Teeteto, Trasímaco, há, por outro lado, a idéia (*eidós*) de homem, a *hominidade*, que não se encerra em nenhum de seus particulares. Onde estão os universais? No “mundo das idéias”, segundo Platão, cuja cosmologia estabelece as características das regiões supralunares, infralunares e o tártaro (ou subterrâneo). Embora uma tal cosmologia possa nos parecer, à primeira vista, completamente anacrônica, a ontologia do senso comum, atualmente, parece dispor de entidades muito semelhantes as do mundo antigo, porém pelo intermédio de outra nomenclatura. No linguajar de Chomsky, as propriedades lingüísticas residem na/no mente/cérebro (cf. CHOMSKY, 2000), o que é uma expressão da diferença entre os aspectos imaterias e materias envolvidos no fenômeno da linguagem.

O empirismo, que nega a existência de outros fatos que não os observáveis direta ou indiretamente, não é aceito com unanimidade. Além disso, excetuando-se as extravagâncias míticas de Platão, seu “mundo das idéias” é a metáfora para uma coleção abstrata de arquétipos conceituais: científicos, estéticos e morais. Em uma palavra, no mundo das idéias residem as formas perfeitas, ideais, para as quais tendem as imperfeitas e atuais. Isto pressupõe uma telicidade determinista para as ações humanas e para a natureza — o *kosmós* (a ordem) em oposição ao caos — e uma teoria positivista do progresso científico, traduzindo-se para os termos da epistemologia contemporânea.

A tese nominalista coincide com a afirmação de que não há realidade extralingüística a que se refiram os nomes comuns. Ou seja, não há referências para os universais, logo, estes são apenas palavras. Isto somado ao solipicismo dos primeiros momentos das Meditações de Descartes, dá ensejo à tese antagônica por trás da famosa metáfora de Searle, da Sala Chinesa, apud. Holmes (1997) e Bittencourt (2001): é possível entender uma linguagem sem estabelecer

relações entre os signos e objetos, mas por meio das atitudes, gestos ou ações, em um jogo, correspondentes a signos. O jogo imaginário consiste em uma pessoa que encerrada em uma sala sem saber nada de chinês, dispõe apenas de manuais para a formação de sentenças nesta língua. E, ao receber perguntas, em chinês, por baixo da porta, é capaz de respondê-las por meio da consulta aos manuais. Se, por um lado, este experimento mental pode ser usado para denegrir a possibilidade do pensamento automático — pois não tenderíamos a acreditar que o jogador na sala chinesa de fato conhece o idioma, ainda que produza respostas inteligíveis para os falantes do chinês — por outro lado, o mesmo experimento mental nos leva a refletir sobre a possibilidade de conceber a linguagem sem a existência de um mundo exterior, de natureza ontológica dubitável, para a qual os signos da linguagem apontem. A contrapartida da metáfora da Sala Chinesa induz a que a aquisição da linguagem, ou o entendimento lingüístico, depende da relação entre os signos e entre estes e seus usuários, mais do que entre os signos e o mundo, as coisas ou a realidade.

A relação entre o naturalismo lingüístico e o realismo, tese defendida pelo personagem Crátilo, no diálogo platônico citado, pode ser estabelecida por meio da afirmação de telicidade ontológica ou, como podem querer os metafísicos, cosmológica. Se os signos não são arbitrários em relação aos objetos significados, então há um plano natural, uma racionalidade onomatúrgica (do grego, *onoma* = nome + *turgós* = fazedor, i.e., fazedor de nomes). Se os universais não são mais do que meras palavras, mas se respaldam em entidades, sejam sociais (cérebros), sejam de outra categoria mais infável (mentes)⁹; e se estas entidades são os arquétipos aos quais estão voltados os indivíduos particulares, então o realismo mostra-se como a contra parte, talvez fundante, do naturalismo lingüístico, ficando estabelecidos os paralelos entre realismo e

⁹ Relacionamos o conceito de cérebro à esfera social em oposição ao conceito de mente, que, segundo a tradição cartesiana, traz um sentido subjetivo, individual e de algum modo, solipscista. Em uma sociedade de cérebros os indivíduos necessariamente se relacionam, têm lugar e tempo determinados; uma sociedade de mentes pode levar a

naturalismo lingüístico e entre nominalismo e empirismo.

Em que esta discussão interfere na questão das relações semânticas? A resposta repousa sobre a motivação analítica com que iremos sugerir as regras de inferência *qualia* (Cap. 2, seção 7). Portanto, trata-se de uma questão de método.

1.2 Categorias

Na seção anterior discorreremos sobre alguns temas de interesse para a ontologia filosófica contemporânea. Feito isto, vejamos quais categorias têm sido pressupostas pelos filósofos¹⁰.

As categorias podem ser vistas como critérios de individuação (cf. MORAVSCIK, op. cit., p. 234), uma vez que são especificações e subespecificações de tipos de entidades. Nesta seção iremos avaliar algumas das opções de categorização ontológica apresentadas na Tabela de Categorias (Apêndice i), tendo em vista nossa finalidade de integração das relações lexicais, ou semânticas, que será exposta mais a frente. Cabe frisar que trataremos os papéis *qualia* como se tratássemos das relações semânticas ou conceituais, ou ainda léxico-conceituais, habituais da semântica relacional utilizada nas Wordnets¹¹.

que seus integrantes jamais se encontrem: as mentes, ou eventos, para usar o termo de Donald Davidson (2001), não possuem determinação espácio-temporal.

¹⁰ Cf. Apêndice i: Tabela de Categorias.

¹¹ Com efeito, este paralelismo é feito por Miller (1998, p. 35-37), o qual avalia considerações de Wierzbicka (1984) e Pustejovsky (1991) a respeito da hiperonímia e de Pustejovsky (1991) a respeito da meronímia. São consideradas as relações de *hiperonímia puramente formal* (ex. Um tordo é uma ave) e *hiperonímia puramente télica* (ex. Um adorno é uma decoração). Nós simplificaremos esta nomenclatura adotando, simplesmente, que a hiperonímia corresponde somente à relação formal, assim como, em Miller (1998, p. 37), a quale formal corresponde somente à relação hiperonímica. A diferença fundamental entre a relação formal ou propriamente hiperonímica e as demais relações expressas nas causas aristotélicas é a necessidade. Um tordo não pode deixar de ser uma ave, já, um adorno pode ter as mais diversas utilidades, como a de servir de calço para uma mesa de pés desiguais.

Também consideramos conveniente considerar que, se a escolha de termos recaísse sobre o uso de expressões como hiperonímia puramente formal e hiperonímia puramente télica, estaríamos levando em conta os significados *ipsis literis* dos termos hiperônimo e hipônimo e teríamos que estender este uso para as demais relações, tal que cheríamos, também, aos termos hiperonímia puramente meronímica e hiperonímia puramente agentiva (além da possível hiperonímia puramente instrumental). Com efeito, toda relação de subordinação e superordenação é uma relação de hiponímia e hiperonímia. No entanto, tal escrutínio terminológico não foi praticado nos documentos da Wordnet, nos quais a hiperonímia corresponde, simplesmente, a hiperonímia puramente formal.

De antemão percebemos que algumas categorias são recorrentes. Por exemplo, “substância” aparece várias vezes e sob palavras diferentes. “Ser” em Platão, sem nos preocuparmos com preciosismos exegéticos, pode ser visto como sinônimo de substância em Aristóteles, “Deus” em Wolf provavelmente é um tipo de substância. Também são tipos de substância as categorias “coisas naturais” e “alma humana”. Outra categoria recorrente é a de relação, que aparece em 5 dos 10 filósofos estudados.

Vamos nos ater à coluna 2 da tabela, onde estão as categorias de Aristóteles. É possível conciliar esta tábua com a famosa Teoria das Quatro Causas exposta pelo mesmo autor em sua *Metafísica*¹². Inicialmente, experimentemos concordar com Wolf quanto ao papel primordial da categoria “relação”. Aliás, esta categoria também foi dialogicamente questionada por Platão: “Creio que admites que, de alguns dos entes, se deve dizer que são unicamente por si, enquanto, de outros, que estão sempre em relação com outros”¹³. Já, Kant colocará a relação em posição inferior à do sujeito cognoscente: “Se suprimíssemos nosso sujeito ou mesmo apenas a natureza subjetiva dos sentidos em geral, toda a natureza, todas as relações entre os objetos no espaço e no tempo, aliás, o espaço e o tempo mesmo desapareceria”¹⁴. Sobre a afirmação de Platão, podemos, sem muito compromisso, imaginar um ponto de vista conciliador entre os dois extremos. Aquele segundo o qual não há entes absolutamente somente por si, nem há entes sempre em relação com outros. Mas, ao invés disso, os entes variam conforme variam os contextos nos quais são interpretados e nos quais existem ou atuam. Em outras palavras, haveria momentos em que algum ente está apenas por si, e momentos em que está em relação a outros. Um exemplo disso aparece no que diz respeito às finalidades.

O fundamental preceito kantiano de que tudo deve ser considerado um fim em si

¹² *Metafísica*, 1013b36 – 1014a8

¹³ *Sof.* 255 c-d., apud Abagnano, op cit.

¹⁴ *Crit. R. Pura*, ¶ 8, apud Iden *ibidem*.

mesmo¹⁵, especialmente os seres humanos, que advém de sua ética deontológica, implica que de todos os possíveis fins de um ente X, X seja o único fim necessário. Ora, se, e enquanto fica aberta a questão acerca da existência de outras finalidades, propriamente instrumentais, que não de o ente ser uma finalidade em si mesma, a única finalidade estabelecida é aquela que ou (i) não liga o ente a nenhum outro ente ou (ii) o liga apenas a si mesmo. Se a relação télica não é aliorelativa (Cf. QUINE, 1980), ou seja, não é tal que X não possa estar ligado a X, então, (ii) sendo verdadeiro, ainda que o ente seja um fim em si mesmo, esteja “unicamente por si” (*autá kath autá*), ainda assim haverá a relação presente entre o ente e a si mesmo. Ou seja, o conceito de relação permite a relação entre iguais. Por outro lado, se a relação télica é aliorelativa, tal que X não pode ter como finalidade X, assim como ninguém pode ser irmão de si mesmo, então o ente não pode “ser por si mesmo”. Portanto, parece estar demonstrado que, seja por si, seja para outro, sempre existe uma relação acompanhando os entes, do ponto de vista de sua telicidade. Mas vejamos a afirmação de Kant quanto à primazia do sujeito cognoscente.

Esta nos parece como a questão dos ovos e da galinha. No humanismo de Kant, sem o sujeito as relações desaparecem. Mas parece convincente o pensamento de que sem as relações o sujeito não é capaz de conceber a si mesmo e às próprias relações.

Portanto, ainda que estejamos submetidos às variações do contexto para o fim de decidir se, e quando, há relações não-aliorelativas de finalidade de um ente para consigo mesmo, ou, de constitutividade, no caso dos átomos no pensamento de Demócrito (460 – 370 a.C.), ou, de agentividade, no caso do motor imóvel de Aristóteles, ainda assim teremos claro que a relação, qualquer que seja ela, permanece como um *a priori* ontológico. Deste ponto em diante podemos nos perguntar se e como as demais categorias da coluna 2 da Tabela de Categorias (Apêndice i) são tipos de relações. E, obviamente, que tipos são estes. Por tal esforço obtivemos a seguinte

¹⁵ (Cf. KANT, 1998 [1785]).

classificação:

Quadro de Categorias

relação

substância = causa formal (*quale* formal)

ter = causa material (*quale* constitutivo)

qualidade

quantidade

lugar

posição

tempo

agir = causa final (*quale* télico)

sofrer = causa eficiente (*quale* agentivo)

Isto não é mais do que dizer que as categorias referentes às Quatro Causas Aristotélicas, são (1) sinônimas das categorias, *substância*, *ter*, *agir* e *sofrer* da coluna 2 da Tabela de Categorias (como indicado pelo sinal = no quadro acima); (2) são tipos de relações e que, (3) *qualidade*, *quantidade*, *lugar*, *tempo* e *posição* são tipos de causas materiais.

Que a substância, *ousía*, seja sinônimo de causa formal em Aristóteles é uma tese controversa, assim como tantas outras em relação às interpretações dos termos da filosofia antiga. Mas não entraremos aqui em detalhes acerca deste ponto. Estabelecemos esta sinonímia principalmente como um procedimento de simplificação da tábua de categorias, ainda que se pudesse defender conclusivamente o nexos entre os dois termos por diversas características tradicionalmente atribuídas a ambos. Um argumento pode ser encontrado nos exemplos de substâncias aos quais Aristóteles se remete: “homem ou cavalo”. Para sermos mais precisos estes são exemplos de substâncias segundas. Pois a diferença entre substância primeira, ou substância propriamente dita, e substância segunda, em Aristóteles, diz respeito à diferença entre particulares e universais¹⁶. Portanto, as substâncias variam conforme o grau de particularidade do objeto referido. Ora, a transição entre o mais particular e o mais genérico, ou universal, é,

justamente, o que reflete a noção de causa formal: uma expressão que diferencie um objeto por superordenação. A causa formal de um objeto/termo X é um objeto/termo Y, que doa suas características a X, pelas quais X se diferencia de qualquer outro objeto/termo. Portanto, substâncias, como “homem ou cavalo” são, também, causas formais em relação a outras substâncias, mais particulares, como brasileiro ou marchador.

Já a sinonímia entre as categorias ter (*éxein*) e a causa material é menos controversa. No entanto, que *qualidade*, *quantidade*, *lugar*, *posição* e *tempo* são tipos de causas materiais, isto sim pede argumentação.

A qualidade expressa por adjetivos como, por exemplo, em “Sócrates é inteligente”, pode ser parafraseada com a proposição tipicamente constitutiva: “Sócrates possui inteligência”. Ou, “parede branca” / “parede cuja cor é o branco” / “parede que possui a cor branca” / “parede que possui a brancura”. O mesmo para as demais categorias dominadas pela causa material:

Quantidade:

“quatro reais”/
“reais cuja quantidade é quadrupla”/
“reais que possuem a quadrupla quantidade”/
“reais que possuem a quadruplicidade”.

Pode-se substantivar o numeral e, menos estranhamente, parafrasear:

“Reais que possuem *o quatro* como quantidade”.

Lugar:

“O Centro de Comunicação e Expressão está na Universidade” /
“O Centro de Comunicação e Expressão possui a característica de sua construção ter sido espacialmente coincidente com a construção da Universidade” /
“O lugar do Centro de Comunicação e Expressão é a universidade” /

Posição:

“O aluno está sentado” /
“A posição do aluno é sentado” /

“O aluno possui a condição de estar na posição a que se chama “sentado””.

Tempo:

“Cheguei ontem” /

“Ontem foi o tempo da minha chegada”/

Ou, substantivando o advérbio:

“O ontem foi ou é o tempo da minha chegada”.

Como se pode perceber, a idéia que está por trás destes exemplos é a da possibilidade de substantivar qualquer classe gramatical, e da constitutividade, própria do caso genitivo, expresso em língua portuguesa pela preposição “do”.

Já as semelhanças entre causa final e causa eficiente com o agir e o sofrer, nos parecem não exigir maiores explicações. Cabe apenas notar que há uma sutil variação entre os sentidos nominais e verbais dos papéis agentivo e télico.

Do ponto de vista nominal, o télico é o substantivo resultante da ação do agente. Por exemplo, a mesa como télico do marceneiro. Do ponto de vista verbal, poderíamos reproduzir a expressão dando-lhe uma ênfase sentencial: o fazer a mesa como o télico do marceneiro. A mesma observação cabe ao papel agentivo, ou à causa eficiente, caso se queira usar a denominação clássica ou moderna. Do ponto de vista nominal, o agente responsável pela derrubada da madeira é o lenhador. Do ponto de vista verbal, é a ação do lenhador sobre a madeira, que, passivamente, se transforma no “ser cortada” da madeira. Em outras palavras, a ação sofrida é a ação responsável por uma mudança de estado. No caso específico das ações agentivas ou eficientes, é a ação que diz respeito ao vir-a-ser do objeto. Embora se possa considerar que toda a ação, enquanto transformadora de estados, produz necessariamente um vir-a-ser do objeto, o qual é alterado de uma condição, inicial, para outra condição, final. Assim, por exemplo, enquanto o agentivo (nominal) da cadeira é o artesão, o agentivo da cadeira suja, ou da

sujeira sobre a cadeira, pode muito bem ser a criança. Uma vez que a criança tenha alterado o estado da cadeira, será a responsável pela sua nova condição¹⁷.

Observação: se toda a ação modifica o estado de um objeto O, i.e., o transfere da condição A para a condição B, então toda a ação é eficiente do objeto O(B), ou seja, O no estado B. A ação é, do ponto de vista de quem age, um télico deste agente. No entanto, a mesma ação é, do ponto de vista de quem a sofre, a ação eficiente. O télico verbal do lenhador, cortar a madeira, é a causa eficiente verbal na voz passiva do pinus de corte, o ser cortado pelo lenhador¹⁸. Esta é uma das várias relações lexicais que se formam entre os quatro componentes da estrutura *qualia*, e que produzem as regras de inferência lexicais que veremos com detalhes no Capítulo 2.

1.3 O sentido gramatical dos papéis *qualia*

Até aqui temos tratado dos papéis *qualia* mais como rótulos para relações semânticas, do tipo das descritas em Cruse (1986), do que como papéis temáticos, segundo a orientação gerativista descrita em Miotto et. all (2004). Vejamos alguns detalhes sobre esta variação e no que o conceito de verbos causativos télicos (cf. MARRAFA; MOURA, 2005) contribui para esta discussão.

Em primeiro lugar, precisamos considerar os conceitos de s-seleção e c-seleção. Segundo Miotto et. all. (2004), uma infração da gramaticalidade quanto à c-seleção é cometida no seguinte exemplo: A Maria beijou que o Pedro saiu. Obviamente, neste caso, a agramaticalidade se localiza na alteração do lugar dos componentes. Sendo que a expressão se corrigiria em: A Maria

¹⁷ Poderíamos evocar em nossa defesa o princípio da conservação de Laplace, “Nada se cria, nada se destrói, tudo se transforma”. No entanto, uma fonte ainda mais antiga se dispõe ao nosso socorro: pelo menos em relação ao sentido da agentividade, Aristóteles concordaria com o Laplace da citação acima: “é causa [eficiente] aquilo que dá início à mudança ou ao repouso: v.g., o autor de uma decisão é causa dela, o pai é causa do filho e, em geral, **o que produz a mudança é causa da mudança**” (apud. Abagnano, 1998 p. 125, grifo nosso).

¹⁸ Poderíamos apresentar um exemplo menos anacrônico se sugeríssemos a relação de agentividade entre um trator, destes que comumente varrem o solo amazônico, e as árvores derrubadas. No grau de especificidade em que estamos trabalhando isto seria aceitável. Porém, existe uma relação entre os contemporâneos trator e assalariado da madeireira e os anacrônicos serrote e lenhador. A saber, a relação de instrumento. Como a causa instrumental foi

beijou o Pedro que saiu. Por outro lado, um caso de restrição quanto à s-seleção (a letra ‘s’ diz respeito ao termo semântica), ocorre no exemplo: A pedra beijou o Pedro. Embora seja possível imaginar diversos contextos em que tal proposição seria aceitável, como no uso metafórico de beijar como “acertar em cheio a cabeça de”, usualmente ou por padrão, esta proposição infringe a s-seleção. Isto porque a pedra, não possuindo boca, não pode beijar. Trata-se, portanto, de uma condição ontológica e universal. Em qualquer que seja a língua humana, a pedra não possui boca para beijar. Por outro lado, a c-seleção, como se restringe a uma questão de ordem dos componentes (syn, gr. com, taxe, gr. ordem), poderia apresentar variação de uma língua para outra. É possível imaginar uma língua humana em que “A Maria beijou que o Pedro saiu” corresponda exatamente à ordem das palavras na sentença que exprime aquilo que em português expressamos por “A Maria beijou o Pedro que saiu”¹⁹. Dito isto, podemos apresentar o seguinte quadro de seleções:

	s-seleção permitida	s-seleção impedida
c-seleção permitida	A Maria beijou o Pedro	A pedra beijou o Pedro
c-seleção impedida	A Maria beijou que o João saiu	A pedra beijou que o João saiu.

Tabela 1: Restrições seletivas

Assim, podemos aceitar as seguintes definições:

s-seleção é a seleção de argumentos permitidos pelo tipo semântico.
c-seleção é a seleção de argumentos permitidos pelo tipo sintático ou morfológico.

Entretanto, embora Miotto et. all. (op. cit., p. 52) afirme: “As informações relativas à s-seleção codificam o que é chamado na teoria gerativa de papel temático ou papel theta” é patente a dissonância entre as restrições de tipo causadas pela oposição mineral/humano em relação à

proposta pelo menos por Plotino (cf. PETERS, [s.d.]), preferimos abstermo-nos de utilizá-la.

¹⁹ “A c-seleção de um predicado também pode variar de uma língua para outra, embora a s- seleção se mantenha estável. Se traduzíssemos uma sentença com *gostar*, por exemplo, para o inglês, veríamos que o objeto do verbo que é um PP em português deve ser um DP no inglês, como mostra (9):

a. O João gosta [PP da Maria]

b. John likes [DP Mary].

Enquanto as informações categoriais de um argumento podem estar em variação, as semânticas parecem ser universais, ontológicas.” (MIOTTO et. all., 2004, p. 88).

oposição entre **do** mineral/**para** o humano, ou simplesmente, **do X/ para** o Y. Enquanto o primeiro par configura variações de s-seleção, o segundo par configura variações de papéis temáticos ou casos nominais, segundo a nomenclatura tradicional. Ou seja, enquanto o primeiro par representa a oposição de categorias ontológicas, o segundo par representa oposição entre os casos da gramática tradicional genitivo e dativo, ou, na nomenclatura lingüística, papéis temáticos fonte e benefactivo. Segundo Moura (comentário em sala de aula: UFSC, 2002), as categorias ontológicas e os papéis temáticos não se correspondem, mas aquelas têm a capacidade de determinação sobre estas. Em outras palavras, do conhecimento de uma categoria ontológica se poderia aceitar ou rejeitar este ou aquele papel temático, mas o contrário não aconteceria. Para reforçar este pensamento analisemos os seguintes exemplos:

O João cortou o cabelo²⁰/
O João cortou a árvore

Vs.

A Maria beijou o João/
A pedra beijou o João

A variação do primeiro par é de carácter diferente da variação do segundo par. Enquanto a variação do primeiro diz respeito às declinações: João na primeira sentença é genitivo, ou tema, enquanto na segunda sentença é nominativo, ou agente. Por outro lado, a variação no segundo par não diz respeito a estas classificações. Pois, tanto a Maria quanto a pedra são nominativos. O que há, neste caso, é uma chamada restrição de tipo, a qual se refere às categorias ontológicas, mas não morfológicas (no caso do grego: Sokrathes ékopsa ton déndron) ou sintáticas (no caso do inglês: John has cut the tree) e/ou preposicionais ou sintático-preposicionais, no caso do

²⁰ Com o significado de, se João não é cabelereiro, “O cabelo do João foi cortado (provavelmente por outra pessoa que não ele mesmo e pelo seu próprio desejo)”. O que também pode ser novamente parafraseado em: “Alguém, muito provavelmente um (a) cabelereiro (a) cortou o cabelo de João segundo o desejo deste”. Ou simplesmente, “O cabelereiro cortou o cabelo de João”. Se pretendermos preservar a ordem original, e apenas marcar os itens que

português²¹.

1.4 *Thesauri*

Os filósofos, na maioria das vezes avessos a trabalhos de implementação, dedicaram-se, sobretudo, à reflexão acerca do significado das categorias fundamentais. A questão da realidade das categorias também desempenhou um papel preponderante nas teorias e nas querelas filosóficas. Geralmente este interesse sobre os fundamentos da ciência não se converte para a atividade científica propriamente dita. O que, aliás, se diz diferenciar a atividade científica da atividade filosófica. Um caso de teoria que não se converte em prática pode ser visto na teoria das categorias filosóficas, que não levou diretamente ao desenvolvimento de uma categorização geral dos entes, mas apenas à reflexão sobre os fundamentos de uma tal categorização²².

Por este viés é que podemos compreender as diferenças entre as iniciativas filosóficas e os *thesauri* de cunho, digamos, filológico universalista. Da mesma maneira, ainda que diversos filósofos gregos e latinos tenham refletido sobre as categorias gramaticais de seus idiomas, e também, comparativamente entre dois ou mais idiomas²³, não foram eles que desenvolveram as gramáticas, enquanto instrumentos úteis ao ensino das línguas. Outra manifestação científico-cultural que parece ser o desenrolar de um intuito filosófico, é a do dicionarista. Um dicionário parece conter um inventário total das entidades do mundo, algo que um filósofo apreciaria, mas,

sofreram elipse com parênteses, temos: “(d) O João (alguém) cortou o cabelo.”

²¹ Uma aplicação particular do conceito de telicidade é utilizada por Marrafa e Moura (2005), na proposição dos verbos causativos télicos como uma classe de verbos evidenciados por alternâncias sintáticas e semânticas como: O treinador preparou o atleta (para a competição)/ O treinador humilhou o atleta. Segundo os autores, na primeira proposição preparar seleciona “para a competição” como um papel télico inerente ao significado do verbo. O que não acontece na segunda. Embora não seja nosso objetivo explorar este tipo de alternância com mais detalhes, apenas mencionamos que o fenômeno em questão pode ser interpretado como um resultado do estímulo à memória para a seleção de termos que façam parte do mesmo campo semântico. Treinador e atleta selecionam competição. Quanto mais próximos os termos dentro do campo semântico, maior a possibilidade de eles se selecionarem reciprocamente (c.f POSNER, 1980).

²² A incorporação dos papéis qualia aos *Top Concepts* (TCs) da EuroWordnet, (cf. VOSSSEN, et. all., 1998), pode ser vista como um resultado prático de uma teoria filosófica. Porém, não se lançaram mão de argumentos filosóficos para a adoção das causas aristotélicas. Os papéis qualia foram implementados na EuroWordnet sob a justificativa de que correspondiam empiricamente aos dados semânticos observados frequentemente nas línguas participantes.

²³ (Cf. AGOSTINHO, 1973 [345 - 430]).

os dicionários não costumam ser elaborados pelos filósofos, embora haja inexpressivas exceções. O movimento enciclopedista francês, (cujos grandes nomes são Diderot (1713 – 1784) e Rousseau (1712 – 1778)), promotor do espírito da Revolução Francesa, de fato chegou a produzir uma enciclopédia, que pretendia abarcar o conhecimento humano a partir de várias ciências. Outro exemplo, mais recente, é o do dicionário de Wittgenstein, o qual foi produzido para crianças. Cabe também citar Leibniz, cujo grande objetivo era o de produzir uma *lingua universalis*, meta que, entretanto, o próprio Leibniz nunca chegou a por em prática. Portanto, podemos dizer que os dicionários, as enciclopédias e os *thesauri* possuem uma inspiração filosófica, que é a de abarcar a totalidade do conhecimento (seja considerando o conhecimento lingüístico (*de dicto?*), ou o conhecimento de uma forma geral, (*de re?*)). Mas, impelidos pela necessidade de abarcar um gigantesco número de itens lexicais, conceitos, relações e definições, os dicionaristas e enciclopedistas tradicionais jamais poderiam se dar ao luxo de uma detida reflexão filosófica acerca de cada item. Um dicionário propriamente filosófico muito provavelmente não sairia da letra "A".

Então, se os dicionaristas não produzem seus verbetes segundo reflexões filosóficas, os produzem segundo uma intuição informal acerca da natureza dos conceitos. E esta intuição não deve ser chamada simplesmente de senso comum, uma vez que normalmente os dicionários não são produzidos a partir de pesquisas objetivas acerca do que pensam as pessoas em relação aos conceitos. Mas são realizados a partir de uma pesquisa subjetiva: o que pensa o dicionarista acerca do que pensam as pessoas acerca do conceito. Este método, assim como as próprias doutrinas filosóficas, produz um resultado extremamente controverso. Não apenas cabem perguntas quanto ao conteúdo das definições e relações entre os conceitos propostos pelo dicionarista, como cabem questões acerca da forma com que tais definições e relações entre conceitos são expressas. Se já não fosse tarefa suficientemente complicada a de determinar o

sentido das palavras, o dicionarista ainda precisa ater-se sobre o problema de quais palavras terão seus sentidos determinados. Normalmente a tradição pesa nestas decisões, e um novo dicionário acaba sendo uma compilação ou atualização de dicionários anteriores²⁴.

Uma alternativa a esta postura subjetivista e tradicionalista é aquela de índole empirista. Um dicionarista adepto do pensamento cético de Locke (1632 – 1704) irá tentar recorrer ao uso efetivo das palavras para, apenas daí, extrair seus sentidos. De outro lado, irá recorrer à frequência das mesmas no uso efetivo, para decidir-se por quais serão e quais não serão elencadas no dicionário. Então, a frequência das palavras, a co-ocorrência, e a frequência dos sentidos das palavras passam a ser os critérios para as decisões na construção do léxico. No entanto, esta postura também guarda suas fraquezas. A primeira delas que podemos elencar é a de cunho epistemológico: o empirismo, de que se serve a chamada lingüística de *corpus*, supõe poder derivar todo o conhecimento da experiência. Imediatamente nos sentimos instigados a perguntar com base em que experiência o empirismo assume sua metodologia empirista. Obviamente a experiência por si só não traz à luz enunciados verificáveis — ainda que sem pretensões à verdade — sem o intermédio da intuição do pesquisador acerca da forma como tais enunciados devem ser extraídos da experiência. No caso da lingüística de *corpus*, as frequências das palavras e os índices de concordância nada diriam, e sequer poderiam ser atestados ou mesmo existiriam, sem os mecanismos matemático-formais da estatística utilizados para tratar os dados puros dos *corpora*.

Outra crítica diz respeito às pretensões socio-antropológicas da lingüística de *corpus*. Que no *corpus* haja conhecimento, deste fato podemos não duvidar; que no *corpus* haja o único conhecimento que nos interessa, isto parece temerário. Por dois motivos: primeiro: se assim fosse, ao radicalizarmos a noção empirista de produção de conhecimento, este se tornaria preso

²⁴ Um verdadeiro manual prático para a construção de dicionários convencionais é Svensen (1993).

em um círculo sobre si mesmo²⁵. Se não há conhecimento válido fora do *corpus*, ou seja, fora do já dito, então não há conhecimento válido fora do estritamente passado. Com isso, se impediria a noção de evolução científica. O segundo motivo está necessariamente relacionado ao anterior: que no *corpus* haja o conhecimento praticado majoritariamente, disto não temos dúvida. Mas podemos não estar interessados no conhecimento praticado majoritariamente. Podemos não estar interessados nos usos mais freqüentes ou estatisticamente relevantes de determinadas construções lingüísticas. Um dicionário, estritamente baseado em *corpus*, também não poderia conter definições, glosas, dos termos, mas apenas exemplos de usos dos mesmos. E isto parece não apenas não ser eficiente para os fins da construção de dicionários para humanos, como, também, veremos no parágrafo seguinte, não se tem obtido um meio eficaz de sistemas de desambiguação exclusivamente por meio de *corpus*.

Os sistemas de desambiguação computacional de linguagens naturais são testados com base na desambiguação manual ou humana de linguagens naturais (cf. RIGAU; et. all., 2001). Vamos imaginar a seguinte experiência. Primeiro um grupo de humanos lê dois textos, A e B, e determina os sentidos de cada palavra nos dois textos. Então uma máquina, um algoritmo, é *ensinada* pelo texto A, o qual lhe é mostrado com a revelação dos sentidos determinados pelos humanos para cada palavra. Depois disto, a máquina é exposta ao texto B, sem que lhe sejam reveladas as anotações semânticas feitas pelo grupo de humanos. Com base nas regularidades estatísticas encontradas entre o texto A e as anotações semânticas feitas pelos humanos neste mesmo texto, a máquina procurará inferir as melhores alternativas para as anotações semânticas do texto B. Este tipo de tarefa, experimento ou exercício (*task*), na prática, sofre algumas alterações, como as que dizem respeito ao tipo de avaliação que a máquina sofre. Mas estes detalhes não nos interessam no momento. O importante é que a máquina será avaliada em função

²⁵ De modo geral, a crítica da autoreferência que se pode fazer ao pirronismo também se aplica ao empirismo.

do grau de semelhança entre as suas anotações semânticas sobre o texto B e as anotações dos humanos sobre o mesmo texto. Deste ponto em diante, nos dirigimos ao ponto crucial. Primeiro, que os resultados não ultrapassam cerca de 60% de acerto (são feitas diferentes medições, que aferem diferentes capacidades de dedução dos sistemas). E, infelizmente, o resultado de 60% faz com que tais sistemas não satisfaçam as necessidades práticas de desambiguação. Principalmente porque um erro leva ao aumento da probabilidade de ocorrerem erros seguintes. Porém, mais grave do que este baixo desempenho dos sistemas avaliados é o seguinte *colapso metodológico*: nem sequer entre dois grupos de humanos há concordância maior do que a obtida pelo sistema computacional. Dois grupos de humanos que não se comuniquem, farão anotações semânticas cerca de 60% diferentes uma em relação a outra²⁶! Aqui estão em cheque dois princípios metodológicos. O primeiro, no que concerne à possibilidade de extrair o valor semântico dos itens lexicais exclusivamente pela sua companhia (os itens lexicais que o cercam a esquerda e à direita, dentro do *corpus*). O segundo princípio metodológico que podemos questionar, que acompanha o primeiro como uma fragilidade da fé empirista, é que não há concordância alta sequer entre os grupos de humanos. Do que podemos questionar que as regularidades e todo o conhecimento que se deseja estejam exclusivamente depositados no uso real e cotidiano das formas lingüísticas. Nossa proposta não negará a provável utilidade do método da lingüística de *corpus* para o fim do processamento automático da semântica das linguagens naturais, no entanto, diante destes argumentos, nos investimos de motivação para o desenvolvimento de um constructo teórico que se dedica a observar regularidades apriorísticas e não empíricas dos fenômenos semânticos.

Carnap (1950), diz, inclusive, que o empirismo é um tipo de ceticismo.

²⁶ Para ser fiel aos detalhes, com o objetivo de participar do SENSEVAL-2, em 2001, Rigau et. all. obtiveram apenas 64% de concordância entre as anotações de dois grupos de humanos!

1.4.1 *Thesaurus* de Roget

Peter Mark Roget (1779-1869) foi um médico inspirado pelo ideal iluminista do século XVIII. Não se notabilizou pelos seus trabalhos em medicina, nem pelos seus trabalhos nas várias sociedades científicas de que participava em Londres, mas por um lazer tomado a sério do qual se ocupou durante sua aposentadoria. Roget era um colecionador de sentidos. Fez ao longo de sua vida um inventário de palavras, que ele relacionava por meio dos conceitos de sinonímia, antonímia, e hiperonímia.

Na confecção do *Thesaurus*, seu objetivo era o de prover uma tábua de idéias ou conceitos mais do que uma classificação de palavras. Pretendia que seu trabalho desse ao consulente a oportunidade de, tendo uma idéia, encontrar as palavras relacionadas, e não, como acontecem nos dicionários alfabéticos convencionais, ter primeiro uma palavra para então localizar sua definição, ou idéia. Havia nisto um sentimento de universalidade. Roget pretendia que seu *Thesaurus* fosse a expressão não do idioma inglês no tempo e local em que ele próprio vivera, mas que fosse a expressão da própria possibilidade de organização dos entes no mundo. Pretendeu que seu trabalho viesse a desempenhar algum papel na união cultural dos povos de línguas diferentes, a caminho de uma era de ouro, proporcionada pela "união e pela harmonia" ²⁷.

Na Introdução ao seu livro, reproduzida na 4ª edição do *Internationals Rogets Thesaurus*, Roget indica a existência de um *Thesaurus* indiano, muito anterior ao seu. No entanto, este trabalho, a que não tivemos acesso, era inapropriado, segundo Roget, por conter classificações como as de “deuses”, “demônios” e “fogo”, certamente elementos importantes para a visão de mundo ou concepção metafísico-ontológica dos indianos que o produziram. Assim, Roget propôs sua própria classificação, segundo um critério subjetivo, sem ter por anteparo qualquer treinamento filosófico ou filológico, mas, ao contrário, um sentido prático.

Portanto, o *thesaurus* consiste em uma coleção de palavras relacionadas entre si, organizadas segundo categorias ontológicas ou, mais especificamente, no caso do de Roget, segundo categorias da causa formal. As relações são de antonímia, sinonímia e hiperonímia. Outra forma de enxergar o *thesaurus*, segundo uma perspectiva menos lingüística e mais epistemológica, é que ele consiste não em uma coleção de palavras relacionadas entre si, mas, primordialmente, em uma coleção de categorias às quais acidentalmente se encaixam palavras da língua inglesa. Ou seja, as palavras e expressões não são o essencial no *thesaurus*, mas as categorias nas quais elas se organizam.

Roget tomou a iniciativa de separar os conceitos segundo seis classes principais, que pretendiam abarcar áreas disjuntas. São elas: relações abstratas; espaço; matéria; intelecto: o exercício da mente; volição: o exercício da vontade; emoção, religião e moralidade. A idéia era que estes campos abarcassem qualquer conceito imaginável. Os conceitos, por sua vez, são representados pelas palavras. Portanto, cada palavra da língua com determinado sentido poderia ser encaixada em uma e apenas uma destas seis classes. Estas, por sua vez, se subdividem em seções, e as seções em cabeças (*heads*) que se distribuem em positivas, negativas e neutras, correspondendo às antonímias²⁸. Podemos perceber uma série de paralelos entre as classes de Roget e as categorias do Apêndice i. Por exemplo, Relações abstratas, com Relação em Aristóteles, Espaço com Lugar, Matéria com um sentido restrito de Causa Material Física, e as demais classes podendo ser coincidadas com sentidos de Agir e Sofrer. Outras identificações entre as categorias de Roget e as categorias filosóficas poderiam ser encontradas, sobretudo se descêssemos nas subdivisões. Por exemplo, enquanto Tempo, para Roget que é uma Relação Abstrata, para Aristóteles, segundo nosso Quadro de Categorias acima, é uma Causa Material, que, por sua vez, é um tipo de relação. No entanto, até esta predomionância da categoria relação

²⁷ (Cf. ROGET, 1977. p. xxx-xxx).

proposta no Quadro citado, e presente em diversos filósofos, pode ter um paralelo no *thesaurus* de Roget, se considerarmos que a ordem em que as seis categorias estão expostas estabelece uma relação da mais abstrata e genérica à mais concreta, particular ou de conteúdo determinado. Portanto, Roget também permite à "relação" uma colocação de destaque frente às demais categorias.

1.4.2 As Wordnets

A palavra Wordnet servia, a princípio, para designar o banco de dados de conhecimento lexical (cf. MARRAFA, 2001), desenvolvido na universidade de Princeton, por um grupo liderado por George Miller, sob a justificativa teórica de experimentos psicolinguísticos realizados nas décadas de 60 e 70, (cf. COLLINS; QUILLIAN, 1968 e MILLER, 1998).

Atualmente, fala-se em Wordnets, e é preciso distinguir entre a "mãe de todas as Wordnets" (cf. FELLBAUM, 1998), aquela desenvolvida na universidade de Princeton, e as demais Wordnets, como a Wordnet.PT, ou a India Wordnet, que diversas instituições ao redor do globo têm desenvolvido para as respectivas línguas nacionais²⁹.

Uma Wordnet, a exemplo da primeira, não é apenas um *thesaurus* eletrônico, uma vez que sua arquitetura computacional possibilita uma série de aplicações em engenharia da linguagem, que outra concepção de *thesaurus* eletrônico, mais fechada, não permitiria. Outro aspecto da Wordnet de Princeton, bem como das demais Wordnets, é que não se pretendeu criar simplesmente um produto, mas se pretendeu a elaboração de um experimento linguístico, psicolinguisticamente motivado, que pudesse ser utilizado gratuitamente (Miller, 1998).

Este banco de dados de conhecimento lexical foi construído com base, principalmente, no

²⁸ Cf. Apêndice ii.i.

²⁹ Para uma lista das Wordnets, ver a página da Global Wordnet Association: <http://www.globalwordnet.org/gwa/index.htm>. Acesso em: 10/10/2005.

Brown *Corpus* (Standard *Corpus* of Present Day Edited English) e no International Rogets *Thesaurus* (ROGET, op. cit.). Do Brown *Corpus* retiraram-se as palavras, a fonte lexical, e do *Thesaurus* os conjuntos de sinônimos. Mas, com o passar do tempo, diversas outras fontes foram utilizadas.

A palavra Wordnet nasceu em 1984 com uma rede de 45 nomes, atualmente em sua versão 2.1, comporta cerca de 200.000 palavras e *synsets* entre todos os conjuntos abertos de categorias lexicais (substantivos, verbos, adjetivos e advérbios). A documentação relativa à Wordnet vs. 2.1 pode ser obtida em <www.Wordnet.princeton.edu>. Acesso em: 10/10/2005.

1.4.2.1 Pressuposições teóricas da Wordnet

Algumas pressuposições teóricas foram tomadas para a realização do projeto de Miller. A primeira é o que podemos chamar *hipótese da separabilidade*. Ou seja, o léxico pode ser estudado independentemente dos outros componentes gramaticais. Isto é motivado, em parte, pelo fato de que as pessoas continuam adquirindo informações lexicais ao longo da vida, mesmo depois de estabelecida a linguagem adulta. A segunda pressuposição é a *hipótese do padrão*. Pode-se afirmar a existência de padrões mentais que governam os significados lexicais. Estes padrões se revelam pela recorrência de fenômenos como o reconhecimento de relações lexicais (cf. cap. 1, seção 2).

A Wordnet pretende-se baseada em uma perspectiva de semântica lexical relacional, ao invés de uma semântica lexical componencial. Esta foi desenvolvida no intuito de se estabelecerem primitivos semânticos que compusessem os significados dos itens lexicais. A abordagem componencialista não apresentou resultados satisfatórios, uma vez que não se chegou ao consenso de qual fosse a lista destes primitivos (cf. SCHANK, 1972). A semântica lexical relacional, por outro lado, assume que o significado de um item lexical não é formado com base em primitivos, de partes constituintes que não seriam, elas próprias, itens lexicais, mas que o

significado de um item no léxico deriva dos significados dos demais itens, de acordo com certas possibilidades de relações ao longo da árvore semântica. Porém, ainda que "no claim is made that the meaning of furniture is a component of the meaning of table", (MILLER, 1998), veremos, mais adiante (Cap. 2), que a perspectiva tomada pelas Wordnets não tem porque rejeitar o encapsulamento de sentidos no entendimento de Cruse (1986). Uma vez que a adoção do conceito de encapsulamento não altera a arquitetura ou organização da Wordnet, mas, apenas, enfatiza a característica do léxico mental de, mais facilmente, selecionar hiperônimos a partir de hipônimos. Uma vez que, necessariamente, há menos hiperônimos diretos para um determinado termo (normalmente um, e eventualmente dois ou três) do que hipônimos deste mesmo termo.

1.4.2.2 Organização dos *synsets* e outras questões de semântica relacional

Os principais tijolos da Wordnet são os chamados *synsets*, que são os conjuntos de sinônimos. Cada *synset* corresponde a um conceito, e é explicitado por uma glosa a exemplo dos dicionários convencionais. Isto se dá pela conclusão de que a mera observação dos sinônimos não era capaz de tornar claro o conceito em questão, quanto mais aumentava o número de nós na árvore. Portanto, a sinonímia é uma relação interna aos *synsets*, e liga as formas de palavras, e não a expressão dos conteúdos. O mesmo acontece com os antônimos, que são relações, no dizer de Cruse (1986), morfo-lexicais. Já, a relação mais importante entre os conceitos, e por isso, talvez a relação mais importante da Wordnet (excluindo-se a inevitável necessidade da sinonímia) é a de hiperonímia. Esta relação se estabelece não entre as formas de palavras, quaisquer que sejam os *synsets* em que estas ocorram, mas entre os próprios *synsets*. As relações de hiperonímia estabelecem uma ligação entre os sentidos, dos mais particulares aos mais gerais. E foi construída com base em 25 categorias principais, para os substantivos, derivadas a partir de

variações lingüísticas atestadas na combinação entre nomes e adjetivos³⁰.

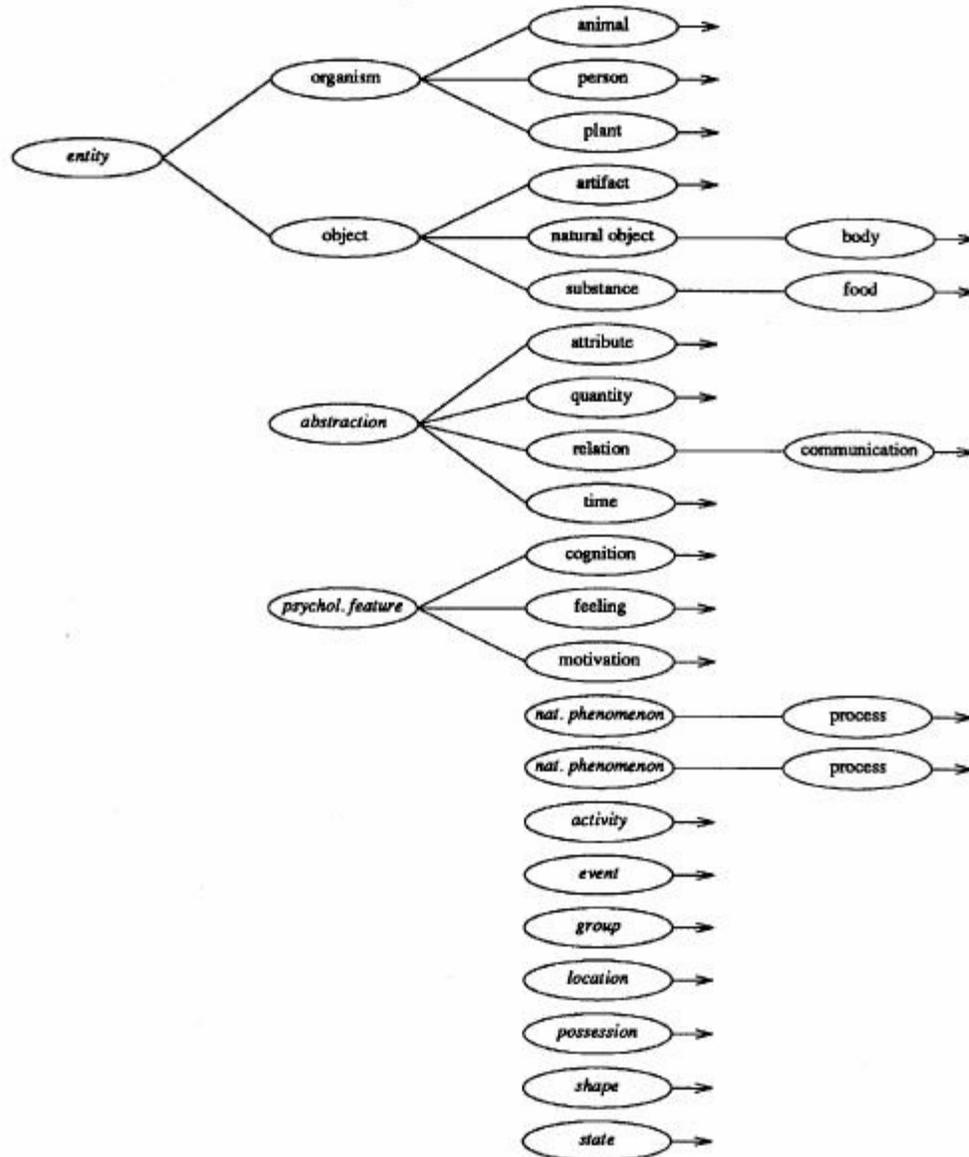


Figura 1: Categorias nominais da Wordnet: em itálico estão os *unique beginners*.

Estas categorias foram organizadas de modo a conterem o menor número possível de

³⁰ Estas 32 categorias, incluídos os *unique beginners*, foram deduzidas, em parte, de anotações não publicadas e, atualmente, extraviadas, de Jonson-Laird (Universidade de Princeton), sobre a referida combinação entre nomes e adjetivos (comunicação pessoal por correio eletrônico de Jonhson-Laird, 15 de setembro de 2005).

"unique begginer", que são as posições mais altas na árvore hierárquica, e sua forma atual é a que mostra a Fig. 1.

Também foi implementada uma árvore hierárquica para os verbos, numa relação chamada de troponímia, (cf. FELLBAUM, 1998b).

Assim, correr é hipo-tropônimo de movimentar-se, i.e., um modo de movimentar-se. Escrever, em um dos seus sentidos, é um modo de comunicar, etc. No entanto, o presente trabalho se atém, sobretudo, à categoria gramatical dos nomes.

Os adjetivos não foram hierarquizados, por isso não estão relacionados pela hiperonímia, mas se organizam principalmente por pares de antônimos e pelos adjetivos ligados a cada um destes antônimos.

Palavras de categorias fechadas não estão incluídas nas Wordnets.

Convém mencionar que o plano inicial era de que, simplesmente, o estabelecimento das relações pudesse revelar o sentido de uma determinada palavra. No entanto, com o aumento do número de conceitos abrangidos, verificou-se que isto não acontecia. E, para que o banco de dados obtivesse maior funcionalidade, acrescentaram-se glosas para os *synsets*. Elas foram feitas da maneira tradicional, como um trabalho manual de lexicografia. Veremos mais à frente que se podem sugerir meios de geração automática de glosas a partir de relações lexicais. Miller (1998) afirma que a glosa é importante para satisfazer a lacuna informativa que se estabelece na descrição do conceito *grandson*, por exemplo, pois a principal relação lexical deste conceito é a que o determina como hipônimo de *grandchild* (neto ou neta).

No entanto, as informações que são postas na glosa, escrita em linguagem natural, poderiam ser decompostas em outras relações lexicais, tais como o constitutivo, agentivo e télico. O constitutivo de *grandson* é "ser macho", que é um constitutivo diferencial de neta. Esta

diferenciação extra soluciona o problema de falta de critério de diferença quando se utiliza somente a *quale* formal. E reduz a quantidade de co-hipônimos não diferenciados (cf. MENDES; et. all., 2001).

Deve-se notar que o projeto da Wordnet não abre mão de que ela contenha todas as informações que o dicionário tradicional disponibiliza, sobre as palavras de que se compõe.

Outra característica da Wordnet (quando nos referirmos simplesmente a Wordnet estaremos nos referindo à Wordnet de Princeton, qualquer versão), é que ela contém outros objetos passíveis de busca em seu banco de dados. Também é possível encontrar pequenas expressões: sintagmas nominais ou verbais, com um sentido determinado. São exemplos, “good person” e “bad person”, que não são encontradas como lexicalizações tradicionais, ou expressões idiomáticas. No entanto, a palavra (*word form*) é a unidade básica (MILLER, 1998, p. 13).

A Wordnet, portanto, é construída ao redor de quatro classes gramaticais: substantivos, verbos, adjetivos e advérbios. Entretanto, há a estrutura de herança lexical do tipo hiperonímica somente para os substantivos e para os verbos. Os advérbios são contemplados somente por relações de sinonímia e familiaridade. Na versão 2.1, a mais recente, estão disponíveis, também, relações de “é um valor de” para os adjetivos. Por exemplo, “alto” é um valor de “grandeza”, em um determinado sentido de alto. Já os advérbios não estão relacionados por qualquer estrutura hierárquica, mesmo na versão mais recente. Tendo em vista estas características, um projeto a ser possivelmente desenvolvido é a união das redes semânticas das quatro categorias, por meio da estrutura *qualia*, o que é possível, nos dizeres de Pustejovsky (1998), uma vez que a estrutura *qualia* se aplica a todas as classes gramaticais.

Quanto à divisão do conhecimento para fins de análise lingüística, entre enciclopédico e conhecimento de mundo, a Wordnet pretende de certa forma, dar conta de ambos. Principalmente

porque se considera que a divisão entre ambos não é bem delimitada, e que os falantes do dia-a-dia necessitam de ambos os tipos de conhecimento. Uma característica de conhecimento enciclopédico contido na Wordnet é, sobretudo, o fato de contar com nomes próprios de pessoas, que sejam representativos de um determinado campo de conhecimento. Constam listas de profissionais, pesquisadores, artistas, estadistas, etc, com pequenas explicações, a exemplo das enciclopédias. Também estão citados nomes de lugares, segundo o critério de importância e representatividade. Cada um, em seu devido lugar, dentro da hierarquia conceitual. Um exemplo é a seqüência de ligações hiperonímicas a partir do nome Bloomfield. Da instância superior às inferiores, temos:

{Entity} -> {physical entity} -> {causal agent, cause, causal agency} -> {organism, being} -> {person, individual, someone, somebody, mortal, soul} -> {scientist, man of science} -> {linguist, linguistic scientist} -> {Bloomfield, Leonard Bloomfield}

Já, a lista de enlaces hiperonímicos a partir de *Brazil*, para os conceitos mais genéricos, é:

{Brazil, Federative Republic of Brazil, Brasil} -> {country, state, land} -> {administrative district, administrative division, territorial division} -> {district, territory, territorial dominion, dominion} -> {region} -> {location} -> {bject, physical object} -> {physical entity} -> {entity}³¹

Pode-se falar em dois tipos de relações entre as palavras na Wordnet: (1) as relações lexicais, ou estritamente lexicais, ou, ainda, morfo-lexicais, segundo o termo utilizado por Cruse (1986), que são relações como, por exemplo, a antonímia e a sinonímia. E (2) as chamadas relações semantico-conceituais, cujo representante principal é a hiperonímia. Outras relações deste mesmo tipo são o acarretamento (*entailment*), a meronímia (*quale* constitutivo), a agentividade e a telicidade. Veja abaixo a tabela de relações para os substantivos da Wordnet em

³¹ Exemplos extraídos da Wordnet 2.1

sua versão 1.5³²:

Relações	Termos conectados	Exemplos	Propriedades das Relações
HIPERONÍMIA	<i>n_synset-n_synset</i>	{oak}--> {tree}	É-UM(A), assimétrico, transitivo
	<i>v_synset-v_synset</i>	{hit}--> {propel, impel}	
HIPONÍMIA	<i>n_synset-n_synset</i>	{ tree } -->{ oak } { propel, impel } --> { hit }	REVERSO_É-UM(A)
	<i>v_synset-v_synset</i>	{ tree } -->{ oak } { propel, impel } -->{ hit }	
POSSUI-MEMBRO	<i>n_synset-n_synset</i>	{family, family unit} --> {child, kid}	
É-MEMBRO-DE	<i>n_synset-n_synset</i>	{child, kid} -->{ family, family unit}	Assimétrico, transitivo
POSSUI-MATÉRIA	<i>n_synset-n_synset</i>	{tank, army tank}-->{ steel}	REVERSO_POSSUI-MEMBRO
É-MATÉRIA-DE	<i>n_synset-n_synset</i>	{steel} --> {tank, army tank}	Assimétrico, transitivo
POSSUI-PARTE	<i>n_synset-n_synset</i>	{torso, body} -->{ shoulder}	REVERSO_POSSUI-MATÉRIA
É-PARTE-DE	<i>n_synset-n_synset</i>	{ shoulder } -->{torso, body}	REVERSO_POSSUI-PARTE
IMPLICA	<i>v_synset-v_synset</i>	{snore, saw wood} --> {sleep, slumber}	Assimétrico, transitivo
CAUSA	<i>v_synset-v_synset</i>	{develop} --> {grow, become larger}	Assimétrico, transitivo
PAST-PARTICIPLE	adj-verb	Developed+develop	Assimétrico
A TRIBUTO	<i>adj_synset-n_synset</i>	{hypocritical} --> {insincerity}	Assimétrico
SYNSET	<i>n_synset-n</i>	{place, property} --> place	de <i>synsets</i> para palavras
	<i>v_synset-v</i>	{travel, journey} --> travel	
	<i>adej_synset-adj</i>	{glad, happy} --> happy	
	<i>Adev_synset-sdv</i>	{well, much} --> well	
PERTENCIMENTO	adj-n	academic --> academia	Assimétrico
	adj-adj	Universalistic --> universal	
	adv-adj	Reasonably --> reasonable k presence-absence	
ANTONÍMIA	n-n	presence --> absence	Simétrico
	v-v	rise--> tfall	
	adj-adj	active --> passive	
	Adv-adv	always --> never	
SIMILARIDADE	Adj-adj	abridge -->shorten	Simétrico
VEJA-TAMBÉM	v-v	touch+touch down	Simétrico
	Adj-adj	Inadequate --> unsatisfactory	

Tabela 2: Relações da Wordnet1.5.

Uma característica importante na seleção de itens que integram o banco de dados lexicais idealizado por George Miller, e sua equipe, é a presença de conceitos que não fazem parte dos léxicos ou dicionários convencionais. Estes conceitos são realizados por pequenas expressões

³² Esta tabela foi extraída, com algumas modificações, de Harabagiu e Moldovan (1998).

assim como “pessoa má”, ou “pessoa boa”. Por isto, pode-se dizer que as palavras dependem dos conceitos, mas os conceitos não dependem das palavras. Os conceitos transitam entre as línguas, podendo ser expressos por combinação de itens lexicais. Pode haver um conceito lexicalizado em uma língua e não lexicalizado em outra, mas expressável em ambas. Além disso, não haverá uma única palavra, em qualquer linguagem natural, à qual não correspondam conceitos, o contrário não sendo verdadeiro.

Os verbos na Wordnet foram alvo de um complexo e inovador trabalho de organização, pois eram raros os estudos de redes semânticas de verbos. A primeira atitude foi identificar campos semânticos que pudessem dividir os verbos em grandes grupos, a partir dos quais pudessem ser feitas novas delimitações. Esta divisão seguiu, prioritariamente, uma “purely intuitive basis” (FELLBAUM, 1998, p. 70). A primeira destas divisões é a que diz respeito à clássica oposição entre verbos que denotam ação e eventos, e verbos que denotam estados. Dentro destes dois ramos foram estabelecidas quatorze categorias: emoção, percepção, contato, comunicação, competição, mudança, cognição, consumo, criação, emoção, percepção, posse, cuidados com o corpo e suas funções, e verbos que se referem a comportamento e interações sociais. Esta classificação não é a única possível: outro grupo de pesquisadores poderia chegar a outros tipos de categorias, assim como acontece na abordagem dos primitivos semânticos. Um grupo heterogêneo foi formado para abranger verbos que não se enquadravam nas categorias anteriores, como o verbo *to be* e auxiliares do inglês, resultando, portanto, em 15 grupos de verbos.

Abaixo, temos exemplos de relações hiperonímicas entre verbos, retirados da Wordnet 2.1. Esta relação deve ser lida como X é um modo de Y, onde X é o verbo hipônimo ou hipotrópico e Y é o verbo hiperônimo ou hiper-tropônimo:

cut --> separate, disunite, divide, part --> move, displace

write --> create verbally --> make, create

run --> travel rapidly, speed, hurry, zip --> ravel, go, move, locomote

Além dos *synsets*, das relações entre estes, de definições no formato de glosas, a Wordnet também contém pequenas sentenças como exemplos (*sample sentences*) para a melhor descrição do sentido, para certos conceitos. Por exemplo, para o sentido de “dog”, expresso pelo *synset* {dog, heel, hound, blackguard, bounder, cad}, no qual a glosa é “*someone who is morally reprehensible*”, consta a *sample sentence* “*you dirty dog*”.

Um dos objetivos que guiam a postura metodológica da Wordnet é o de modelizar o conhecimento lexical dos falantes (FELLBAUM, op. cit., p. 74). Por este motivo as tabelas de categorias são endossadas por estudos psicolinguísticos que atestam a plausibilidade psicológica do modelo de redes semânticas. No entanto, este tipo de teste empiricamente verificado serve, somente, para confirmar em linhas gerais o procedimento das redes semânticas, pois, dificilmente se poderiam atestar as minúcias das relações da Wordnet por meio de tais testes. Outra característica que faz com que a WorNet se afaste da possibilidade de ser inteiramente confirmada por procedimentos de psicolinguística, é o fato de ela conter, em diversos pontos da rede nominal, muitos degraus, os quais dificilmente teriam contraparte no léxico mental dos sujeitos em situação de teste³³.

A Wordnet se parece, em certo sentido, com a representação de uma espécie de léxico do conhecimento enciclopédico. Do ponto de vista metodológico, podemos entender que a construção de um léxico com interface computacional, como a Wordnet, não precisa estar presa às confirmações do experimento psicolinguístico, tanto quanto queiramos o estabelecimento de uma ferramenta para processamento automático de linguagens naturais (PLN) que se aproxime da

³³ Sobre este ponto ver Fellbaum et. all. (2005).

competência lingüística e não do desempenho. Afinal, basta considerar que a capacidade de memória e recuperação da máquina ultrapassa, em muito, atualmente, as capacidades cognitivas concretas da mente/cérebro humano.

Diante disto, nos perguntamos se a Wordnet, ou outra concepção de rede semântica, deve pressupor a existência, no léxico mental, das relações que propõe. Uma resposta a esta pergunta é a reformulação da mesma. Ao invés de nos perguntamos sobre o que há no léxico mental, poderíamos nos ater à questão instrumental da utilidade para fins de PLN do estabelecimento destas ou daquelas relações semânticas. Além disso, podemos encarar o léxico mental não como um repositório fechado e estático, mas como uma estrutura dinâmica, cujas mudanças se dão por combinação de alguns princípios essenciais.

Outra fonte de legitimação das relações semânticas da Wordnet tem sido a análise de padrões na construção de definições dos dicionários convencionais³⁴. Também as definições dos dicionários têm sido usadas para se extrair relações semânticas (cf. HODOROW et. all., 1985).

As redes semânticas das quatro classes gramaticais, abrangidas pela Wordnet, estão separadas, como já se disse. No entanto, a unificação destas poderia trazer vantagens para a formação de composições lexicais. Inclusive no sentido das restrições seletivas de Pustejovski:

“A importância da contribuição dos nomes ao significado dos verbos tem sido destacada em muitos trabalhos recentes de semântica lexical. Pustejovsky (1991, 1995) argumenta contra a idéia de sentidos estáticos, fixos, e propõe, ao invés, a noção de um ‘Léxico Gerativo’ mais flexível. Ele enfatiza a polissemia de alguns verbos aspectuais como ‘ser’, verbos experienciativos como ‘terminar’ em ‘terminar a bebida’ e muitos causativos, cujos significados dependem em grande medida do contexto particular no qual ocorrem. Por exemplo, terminar em (1), abaixo, significa ‘terminar de fumar’; enquanto que em (2), abaixo, significa ‘terminar de beber’:

(1) Paulo terminou o cigarro.

³⁴ “Uma fonte rica de inspiração para a representação das palavras e dos conceitos é encontrada nos dicionários tradicionais, nos quais as palavras são definidas em termos de outras palavras” (FELLBAUM, 1998, p. 76, tradução nossa).

(2) Paulo terminou a cerveja.

Aqui, o cigarro e a cerveja ‘imprimem’ o significado dos verbos. Pustejovsky propõe um mecanismo pelo qual a contribuição dos argumentos dos verbos para o significado dos mesmos é apreendida de forma sistemática. Esta proposta estimulante poderia ser capturada pela Wordnet se nomes e verbos estivessem ligados. Nomes poderiam ser relacionados a verbos expressando funções características [...]. Entretanto, não é claro a que nível as ligações precisariam ser feitas se quisermos esforçar-nos para obter economia máxima no léxico e, ao mesmo tempo, evitar grandes números de excessões.” (FELLBAUM, 1998, p. 85, tradução nossa).

Neste momento é importante frisar que a concepção de sinonímia utilizada pela Wordnet é a de sinonímia fraca. Uma concepção forte de sinonímia, baseada na famosa Lei de Leibniz, (cf. KRAUSE, 2000), é aquela segundo a qual duas palavras são sinônimas se e somente se puderem ser substituídas uma pela outra, em qualquer contexto, sem que isto altere o valor de verdade da sentença (ver Cap. 1 seção 2). No entanto, esta concepção não parece se aplicar à linguagem natural assim como se aplicaria à aritmética, pois dificilmente se atestaria este grau absoluto de sinonímia entre duas palavras diferentes. Portanto, a sinonímia fraca é aquela segundo a qual são sinônimas duas palavras desde que possam ser substituídas uma pela outra *salva veritate* em algum contexto. O problema desta definição é que, sendo fraca, torna-se demasiadamente permissiva. Pois decorre dela que quaisquer duas palavras são sinônimas, uma vez que é possível imaginar os mais variados contextos nos quais, por um acordo tácito entre os falante, duas palavras de formas diferentes resultem no mesmo referente (ou valor de verdade), no contexto discursivo. Portanto, admitiremos que os sinônimos na Wordnet são um expediente prioritariamente baseado na tradição lexicográfica, que não se assentam em qualquer princípio apriorístico da teoria da semântica relacional.

Uma pergunta que se pode fazer quanto à Wordnet, diante das questões do Léxico Gerativo, é se a Wordnet é *apenas* mais uma estratégia de enumeração de sentidos lexicais (*sense enumeration lexicon*, SEL) ou não. Pustejovski (1995, p. 243) diz que sim. No entanto, se

considerarmos como unidades mínimas de significado, não as formas de palavras, mas os *synsets*, teremos que, para cada *synset*, há uma e apenas uma glosa. O que corresponde dizer que para cada item semântico há apenas um sentido. Por outro lado, se considerarmos as *word forms*, veremos que muitas delas aparecem em diversos *synsets*, e que para cada um destes *synsets* há um e apenas um sentido estabelecido, o que configuraria um SEL.

As relações nominais visam determinar as características capazes de diferenciar os conceitos. Uma forma de ver estas características é dividindo-as entre: atributos (adjetivos), partes (substantivos) e funções (verbos). No entanto, pela possibilidade de substantivação das diversas categorias gramaticais, podemos supor que o conceito a que se referem estas características diferenciadoras não é, *per si*, característica exclusiva destas classes gramaticais. Suponhamos a função expressa pelo sintagma verbal "guardar objetos domésticos" ligada ao nome "gaveta". Podemos transcrever esta função pela paráfrase nominal "a guarda de objetos domésticos". Ora, "a guarda de objetos domésticos" passaria a ser, então, uma parte de "gaveta", enquanto "guardar objetos domésticos" seria sua função? Também dispomos do mesmo recurso no que diz respeito aos adjetivos. Podemos transformar o atributo branco da parede na parte constituinte "a brancura" da parede, etc. Além disso, um componente de diferenciação dos nomes também é a origem, ou papel agentivo, que se antepõe logicamente à finalidade, função ou papel télico.

No entanto, ainda que haja a herança lexical, esta somente acontece ao longo dos eixos de uma mesma relação semântica. “Na Wordnet informação sobre as propriedades extrínsecas não definem hiponímia ou contribuem para a hierarquia nominal de nenhum modo explícito” (MILLER, 1998, p. 37, tradução nossa). Nossa proposta será a de integrar as relações semânticas de modo a dar maior capacidade inferencial à rede como um todo. Esta integração das relações semânticas seria capaz de criar glosas automaticamente, mesmo que "Demais a mais, é muito

mais fácil ler uma definição por meio de uma glosa do que tentar inferí-la a partir de uma lista de ponteiros associados com cada conjunto de sinônimos" (...) "Um teste imaginável do sistema poderia ser, então, escrever um program de computador capaz de sintetizar glossas a partir das informações fornecidas pelos ponteiros." (Idem).

Ainda que seja um objetivo da Wordnet manter um pequeno número de relações semânticas, são assumidos 3 tipos de meronímia: componente, membro, e matéria de que é feito. A relação chamada de atributo (ver Tabela 2) também se aproxima da relação fixada pela *quale* constitutiva, nos termos de Pustejovsky (1995). Sobre isto, Miller (op. cit), diz: "a dog can be friendly or unfriendly because friendliness is an attribute of dogs, but stingy dog or shallow dog must be interpreted figuratively, since generosity and deph are not normal canine attributes."

Outra característica da Wordnet é a possibilidade de acrescentar *synsets* ad hoc, para a conveniência da classificação. Esta decisão, facilitada pelo meio eletrônico, onde o lexicógrafo não precisa se importar com a economia de papel, é também incentivada por Aristóteles. Este, de fato, menciona no Organon, que, uma vez que haja um conceito para o qual não se estabeleceram termos, temos a obrigação de criá-los. Diante disso avaliemos o seguinte problema³⁵: um constitutivo de "pássaro" é "bico". Este, por sua vez, é hipônimo de "maxilar", o qual é uma "estrutura esquelética". Aceitando estas relações, obtemos a figura abaixo:

³⁵ Este problema, bem como a solução que proporemos é abordado por Miller (1998).

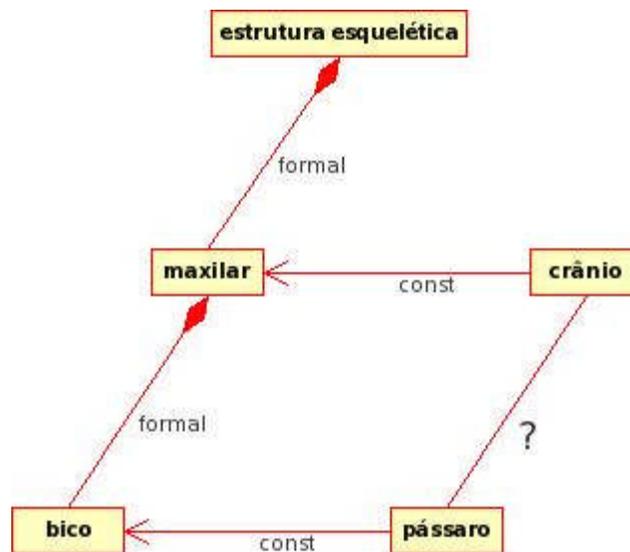


Figura 2: Hiponímia e meronímia de pássaro.

Esta figura aponta o problema de se estabelecer uma relação coerente entre pássaro e crânio. Considerando que aceitamos a lexicalização de conceitos *ad hoc*, conforme ditam a necessidade de organização e preenchimento de lacunas do léxico, podemos obter a seguinte

Figura:

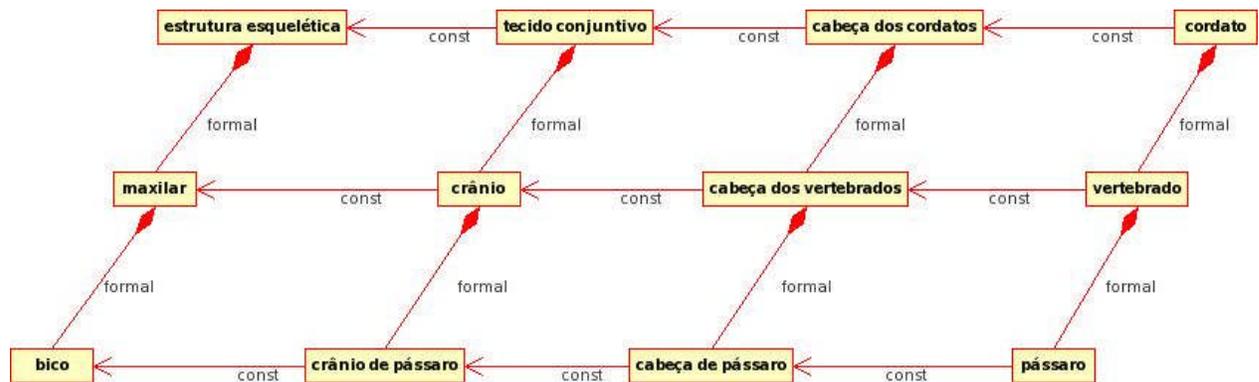


Figura 3: Acréscimo de *synsets* ad hoc.

Note-se que não se espera encontrar vários dos nós da Figura 3 em dicionários convencionais, ainda que se trate de um dicionário de grande abrangência como os que se pretendem inventários completos do léxico de uma língua. No entanto, “cabeça de pássaro” é um conceito perfeitamente inteligível que, eventualmente, é lexicalizado por meio de uma única palavra em alguma outra língua humana (ou não-humana). Esta solução estabelece que a relação

entre “pássaro” e “crânio” é, seguindo o termo de Cruse (1986), tal que “crânio” é um mero-mero-hiperônimo de pássaro. O que soluciona o ponto de interrogação da Figura 2.

1.5 A ontologia na computação: a Semantic Web e o Sistema UNL

Wordnets têm sido chamadas de ontologias, embora este termo fosse, de início, utilizado pelos cientistas da computação para designar bancos de dados relacionais, com ênfase nos conhecimentos especializados, e que não diziam respeito à totalidade de uma língua (cf. BERNES-LEE, 2001 e BERTOLI et. all, 2001).

Uma definição de ontologia comum nos manuais de ciências da computação a limita a um arquivo de computador no qual se estabelecem, formalmente, relações entre termos. Obviamente esta definição, ao mesmo tempo em que se isenta de conteúdos psicológicos, permite uma abrangência muito grande para o que se deseja, no ramo das ciências da computação, entender por ontologia. No entanto, Noy e McGuiness ([s.d.]) diminui o escopo desta definição, por meio de determinações da finalidade de uma ontologia. Neste quesito a principal característica é a possibilidade de se reutilizar os conhecimentos de uma ontologia, acrescentando-a a outra. Também se pode fazer uma relação tal que a ontologia nas ciências da computação está para a engenharia do conhecimento, assim como as redes semânticas, lingüisticamente consideradas, estão para a engenharia da linguagem ou o PLN. Uma questão interessante: não serão sempre incompletas as abordagens que separem conhecimento de linguagem?

Uma ontologia, segundo esta visão, pode ser um banco de dados que armazene informações sobre vinhos: tipos de vinhos e suas classificações, propriedades, fabricantes, alimentos com os quais se combinam, etc. Estes conceitos são organizados de modo a possibilitar a extração automática de respostas à perguntas especializadas. Por exemplo: que vinho melhor acomapnha peixe? Ou, quais são os vinhos brancos franceses?

As principais relações, assim como na Wordnet, são as que permitem hierarquias, que

transitam entre os níveis dos mais específicos aos mais gerais. Aliás, a constante presença de substâncias primeiras, segundo o termo aristotélico, ou entidades concretas, específicas, é um imperativo das ontologias segundo a perspectiva das ciências da computação. Como uma ontologia assim concebida visa uma aplicação prática bem determinada, como a de ajudar os garçons de um restaurante a responderem as perguntas sobre vinhos, é possível que na posição mais inferior da hierarquia de vinhos estejam as representações das próprias garrafas, segundo os critérios de localização na adega, e preço atualizado pela cotação do Euro, por exemplo. Outra possibilidade de aplicação para este conceito de ontologia é a de organizar os elementos de uma empresa. O software de edição de ontologias Protégé vs. 3.1.1,³⁶ traz consigo um exemplo de uma ontologia voltada para a organização dos elementos de um jornal. As categorias são: autores, conteúdo, informação sobre layout, biblioteca, jornais, organização e pessoal. A categoria de conteúdo se divide em anúncios e artigos. A categoria de pessoal (*person*) tem uma única subcategoria direta, empregados. Esta se divide em colunista, editor, jornalista, vendedor e gerente. Na categoria gerente está submetida a categoria de diretor.

Para esclarecer o significado da submissão de categorias, basta observar quais são as propriedades estabelecidas para as “entidades” que se liguem somente à categoria pessoal, e não à categoria empregados. São propriedades de pessoal: nome, telefone e outras informações. Já, a categoria de empregados contém, além destas três que recebe por herança, função, data de contratação e salário. Outras categorias que não empregados poderiam ser submetidas à categoria pessoal: colaboradores, voluntários, fornecedores, vendedores autônomos, etc.

Uma abordagem diferente, mas próxima à ontologia computacional é o sistema UNL (*Universal Network Language*), o qual se pretende uma linguagem de computadores que seja capaz de representar os conteúdos das expressões das linguagens humanas (cf. UNDL FOUNDATION, 2004). O grande objetivo deste sistema, que é capitaneado pela Universidade

³⁶ Disponível em <<http://protege.stanford.edu>>. Visitado em: 29/08/2005.

das Nações Unidas, é a tradução automática. Para isto são estabelecidas relações sintáticas e semânticas, bem como se deseja a utilização de léxicos estruturados a exemplo do sugerido por Pustejovski (1995). A UNL trata-se, portanto, de uma linguagem de computador para a codificação universal das línguas humanas.

Finalmente, a Semantic Web (BERNES-LEE, op. cit.), é um projeto pelo qual se pretende uma revolução no tipo de informação que está disponível, atualmente, pela Internet. A idéia é que as páginas da Web contivessem não apenas as convencionais informações de formatação de texto, próprias da *hyper text markup language* (html), pela qual se estabelecem os *links* entre URLs e a formatação gráfica dos textos e imagens, mas que as páginas da Internet fossem, também, preenchidas com informação semântica possível de ser compreendida por máquina. Isto significa, resumidamente, adicionar etiquetas de sentido, segundo uma ontologia, de tal forma que, por métodos de inferência e inteligência artificial, agentes eletrônicos pudessem resolver questões. Por exemplo, se alguém deseja encontrar uma clínica hospitalar com atendimento em determinada especialidade, para determinados horários, poderia pedir ao seu agente para procurar nas páginas da Semantic Web, aquela clínica que melhor se adapta à necessidade. Isto seria possível porque os construtores desta página da Semantic Web adicionariam ao texto para ser lido por humanos, um segundo texto, para ser lido por agentes eletrônicos. Este texto estaria no formato de etiquetas semânticas estruturadamente relacionadas. Outra característica relevante da Semantic Web é que ela consistiria em um repositório dinâmico de conhecimento semântico, diferente de uma Wordnet que segue o modelo das enciclopédias tradicionais. Ou seja, a Wordnet é fruto de uma empreitada lingüística sem que haja a necessidade, já estabelecida em sua arquitetura, de incremento ou modificação. A Wordnet, ao contrário das ontologias sob a perspectiva das ciências da computação, tem um caráter estático. Seu desenho não privilegia a atualização.

Estes são, em linhas gerais, os principais caminhos que a ontologia na computação vem assumindo atualmente. Apenas poderíamos acrescentar o desenvolvimento de softwares gráficos, como o Visual *Thesaurus* da ThinkMap³⁷, que traduz redes semânticas simbólicas em imagens de árvores conceituais. Este tipo de programa pretende permitir uma navegação visual pelos conceitos. E igualmente às próprias ontologias computacionais, a idéia de um *thesaurus* visual possibilita diversas aplicações, como a facilitação ao acesso de grandes bancos de dados corporativos (cf. THINK MAP, 2004³⁸).

³⁷ Cf. <www.visualthesaurus.com>, acesso em 10/10/2005.

³⁸ Também disponível no endereço acima. Acessado na mesma data.

2 Redes Semânticas

Nesse capítulo apresentaremos nossa proposta da *QualiaNet*, uma rede semântica pautada nas relações *qualia*. Abordaremos o modo como essas relações têm sido utilizadas para o enriquecimento das Wordnets e a possibilidade de integrar tais relações por meio das propriedades de herança lexical. Para este objetivo apresentaremos as Regras de Inferência *Qualia* (QFs) às quais aplicaremos testes lingüísticos. Os exemplos serão contrapostos às relações estabelecidas na Wordnet 2.0. Demonstraremos que as QFs incrementam o conteúdo semântico da Wordnet, e que isto se faz por meio da geração automática de relações hiperonímicas, entre outras.

2.1 Relações semânticas

Relações semânticas são definidas como fenômenos lingüísticos recorrentes, que podem ser subjetivamente atestados. Por exemplo, a antonímia, que se verifica pela possibilidade de elencarmos diversos pares de termos antônimos. Em outras palavras, existe alguma propriedade no léxico mental que faz com que reconheçamos uma e apenas uma relação estabelecida entre os pares alto/baixo e quente/frio. Entretanto, ainda que haja relativo consenso sobre este princípio básico para o reconhecimento de relações semânticas, diversos autores têm tratado do tema de formas diferentes. Faremos abaixo uma rápida catalogação de tipos de relações semânticas, para o que nos basearemos principalmente em Rector e Yunes (1980).

Constelação de Saussure. Esta metáfora vai ao encontro da teoria do valor na lingüística saussuriana: "um termo dado é como o centro de uma constelação, o ponto para onde convergem os outros termos coordenados, cuja soma é indefinida". Saussure apud Rector e Yunes (1980, p. 72).

Campos Semânticos de Bally. São campos semânticos associativos, considerados como "um halo que circunda o signo e cujas franjas exteriores se confundem com seu ambiente" , Bally, Charles, apud Rector e Yunes (op. cit., p. 73). São exemplos de itens de um mesmo

campo, a partir da palavra boi: vaca, touro; arado, jugo; trabalho lento e pesado, etc.

Campos Semânticos de Trier. Para este autor, um campo semântico é: "um conjunto de palavras não necessariamente aparentadas etimologicamente, nem ligadas por associações psicológicas individuais, arbitrárias e contingentes que, colocadas lado a lado como pedras de um mosaico, recobrem exatamente um domínio bem delimitado de significações, constituído tradicional ou cientificamente pela experiência humana". Op. cit., p. 75.

Campos Semânticos de Mounin. Para este autor, os campos semânticos são de "origem e de natureza conceitual empírica e as percepções da experiência não são perfeitamente lexicalizadas". Op. cit., p. 76. "As palavras não constituem um sistema; quando muito formam pequenos grupos", o que vem a contrariar o princípio de Trier. Mas a língua não necessita necessariamente de um termo para cada conceituação. "Há outras formas de manifestação, por exemplo, com auxílio de adjetivos circunstanciais". Op. cit., p. 77-8.

O que se disse acima reforça a idéia da lexicalização de itens tradicionalmente não-lexicalizados. Afinal, estas construções são importantes uma vez que um *thesaurus* admite a possibilidade de consultarmos palavras por meio de conceitos e não apenas conceitos por meio de palavras. Logo, é natural que "As palavras não constituam um sistema (...)", etc.

Análise Componencial. Os itens lexicais são definidos a partir de seus componentes. Os componentes são semelhantes aos hiperônimos. Termos como escolher, comer, dormir, e sentir teriam em comum o componente [atividade]. Outros critérios podem ser estabelecidos. Unindo os termos trocar, dar, receber, emprestar, herdar, alugar, etc., está a "mesma estrutura formal na qual o objeto é posse de diferentes pessoas em épocas diversas". Op. cit. p. 91.

Relações de Parentesco. O fenômeno das relações semânticas foi estudado por Goodenough (1956), apud Rector e Yunes (op. cit., p. 91) a partir das relações de parentesco. Com base na combinação de conceitos-chave das relações de parentesco, como pai, mãe, irmão, irmã, filho, filha, etc., o autor deduz três dimensões para definir os termos: sexo, geração e

linearidade. Vistos estes caracteres gerais, pretende-se uma definição dos termos por meio de suas respectivas posições dentro das dimensões e graduações das dimensões estabelecidas. Por exemplo: irmã: + feminino, a própria geração, parentesco por consangüinidade.

A Taxonomia para Bendix. As afirmações deste autor antecedem o conceito de encapsulamento de sentidos de Cruse (1986), o qual abordaremos mais a frente. Pois, para Bendix, “a taxonomia é um tipo de paradigma de um subconjunto natural de termos, onde um elemento superior é igualmente um componente que entra na definição de uma forma que exprime um elemento inferior, e no qual a relação entre um elemento superior e um elemento inferior é uma oposição hierarquizada, caracterizada por uma marca e uma não-marca referente a um ou mais componentes.” (RECTOR; YUNES, 1980, p. 99).

Análise Distribucional. A análise distribucional (Apresjan) pode ser vista como uma antecessora da lingüística de *corpus*. Pois leva em conta a relação entre as palavras em contextos. Segundo Harris, chama-se distribuição de um elemento lingüístico "a soma de todos os contextos nos quais ele se encontra, quer dizer, a soma de todas as posições diferentes de um elemento em relação aos outros." (op. cit. p. 111).

Análise Sêmica. A terminologia desta abordagem estabelece que traços semânticos, ou semas, são as unidades mínimas da significação. Para Bernard Potier, "os elementos lexicais (palavras e locuções) constituem um número ilimitado e infinito, ou seja, formam um inventário aberto." op. cit. p. 116. Relações como as estabelecidas pelos papéis *qualia*, para este mesmo autor, são as “substâncias do significado”. Os papéis formais são os semas genéricos ou classemas. A substância do significado pode ser entendida como a definição do termo que se utiliza das características dos papéis *qualia*. Por exemplo, a substância do significado de boi é herbívoro (papel formal) com quatro patas, rabo (papel constitutivo), etc. (papéis agentivo e télico). Curiosamente, Potier estabelece quatro relações semânticas: de oposição (co-hiponímia); de inclusão (hiperonímia); de participação (meronímica ou constitutiva); de associação, por

exemplo, limão – calor. A que podemos conferir uma conseqüência da relação télica, pois é possível a paráfrase: o limão serve para ser consumido em forma de suco, no calor. Ou, mais propriamente, calor estaria na relação de constitutivo do télico de limão.

2.2 Os papéis *qualia* como relações semânticas e as séries proporcionais

Como já deixamos entrevisto, os papéis *qualia* serão, aqui, tratados como relações semânticas. A relação formal é idêntica à hiperonímica; a constitutiva iguala-se à meronímia; o agentivo, que nos termos de Aristóteles costuma ser designado como causa eficiente, é a origem, o meio pelo qual algo vem a ser o que é; e, por fim, o télico é o objetivo, aquilo a que se destina o conceito em questão.

Estas relações podem ser consideradas relações semânticas pela possibilidade do reconhecimento subjetivo da regularidade das relações entre pares de termos. As relações de hiperonímia e meronímia não apresentam problemas, uma vez que são habituais nas abordagens de léxicos estruturados³⁹. Já as relações télica e agentiva podem gerar alguma incerteza. São exemplos de pares de termos relacionados pela agentividade: sol/calor terrestre, padeiro/pão, escritor/texto, autor/obra, construtor/construção, etc. A relação télica pode ser exemplificada pelos seguintes pares: flor/reprodução, folha/respiração, pernas/deslocamento, passarela/travessia, etc.

Estas relações podem ser dispostas no formato das séries proporcionais, as quais foram intensamente estudadas por Cruse (1985, p. 118-135). Uma série proporcional é um conjunto de elementos coerentemente relacionados.

Veja-se a tabela abaixo:

³⁹ Ipsis literis, hiper/hiponímia seria uma relação estabelecida entre quaisquer grupos de palavras que mantivessem entre si a relação de super/subordinação. Isto porque *hiper* e *hipo* são prefixos de origem grega para acima e abaixo. E *ónoma* representa nome. Então, ácido seria hipônimo de limão, no eixo constitutivo. Entretanto, estabeleceu-se a tradição de chamar de hiper/hipônimo principalmente a relação formal.

2	3
4	6

Tabela 3: Uma série proporcional matemática.

As proporções são indicadas da seguinte maneira:

2 está para 4 assim como 3 está para 6
4 está para 2 assim como 6 está para 3
2 esta para 3 assim como 4 está para 6
3 esta para 2 assim como 6 esta para 4

Um exemplo de série proporcional léxico-semântica é a que segue:

Norte	Sul
Leste	Oeste

Tabela 4: Uma série proporcional léxico-semântica.

Com a seguinte explicitação das proporções:

Norte esta para Sul assim como Leste esta para Oeste
Sul esta para Norte assim como Oeste esta para Leste
Norte esta para Leste assim como Sul esta para Oeste
Leste está para Norte assim como Oeste está para Sul

Estes dois exemplos são séries proporcionais do menor tamanho possível. Uma série proporcional de quatro elementos segue a seguinte regra geral:

A	B
C	D

Tabela 5: Esquema geral das séries proporcionais de quatro elementos.

A está para B assim como C está para D
B está para A assim como D está para C

A está para C assim como B está para D
C está para A assim como D está para B

No entanto, as séries proporcionais podem se estender para além de quatro elementos. E, teoricamente, não há limite final para o número de elementos de uma série proporcional.

Veja-se no exemplo abaixo a série proporcional matemática da qual a série representada na Tabela 3 é um subconjunto:

	1	2	3	m
1	1	2	3	$2^{(1-1)}m$
2	2	4	6	$2^{(2-1)}m$
3	4	8	12	$2^{(3-1)}m$
n	$2^{(n-1)}1$	$2^{(n-1)}2$	$2^{(n-1)}3$	$2^{(n-1)}m$

Tabela 6: Uma série proporcional matemática infinita⁴⁰.

Agora, vejamos séries semânticas compostas de mais do que quatro elementos:

Cadeira	Madeira	Verniz
Mobília	Objetos Sólidos	Revestimento

Tabela 7: Série proporcional semântica com mais de quatro elementos.

Note-se que no eixo horizontal a relação é constitutiva e no eixo vertical a relação é formal. Compare-se com a série abaixo:

Leitura	Texto	Escritor
Texto	Escritor	Alfabetizador

Tabela 8: Série proporcional formada pela agentividade e pela telicidade.

Nesta série a relação compreendida no eixo horizontal é a de agentividade e a compreendida no eixo vertical é de telicidade.

⁴⁰ A regra geral desta série proporcional, $2^{(n-1)}m$, foi determinada pelo físico Msc. Carlos Eduardo Rodrigues, em comunicação pessoal. A parte mais escura é integrante do cap. 5, seção .3 de Cruse (1986). O fundamental desta tabela para os nossos propósitos, porém, é simplesmente o fato da formação de séries proporcionais entre quaisquer que sejam quatro células, tais que as quatro formem um quadrado ou um retângulo. Assim como o termos $\{<1, 2>$, $<4, 8>$ formam uma série proporcional, os termos $\{<1, 4>$, $<4, 12>$ também formam.

A composição, ou integração, entre as séries proporcionais é aplicada à estrutura *qualia* e ao esquema de hierarquia ontológica próprio das Wordnets. Este é o coração do que chamamos de *QualiaNet*.

A referida integração pode ser representada pela seguinte figura tridimensional:

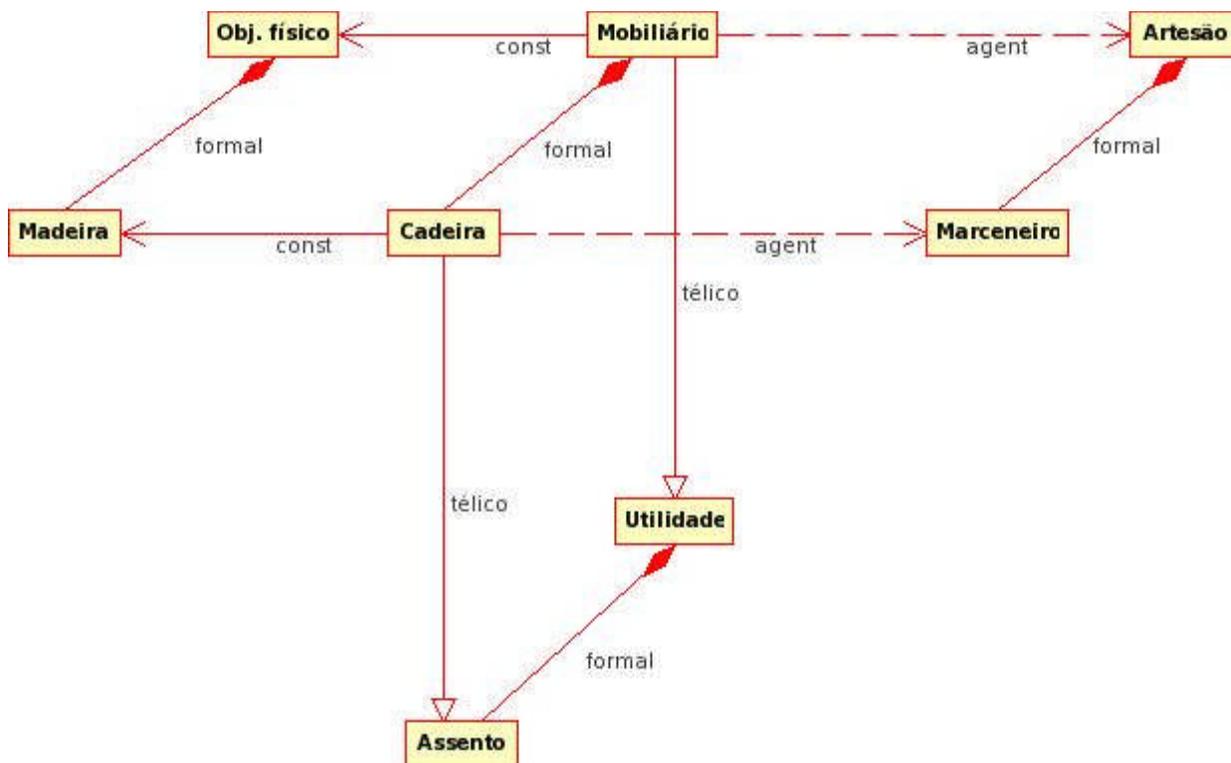


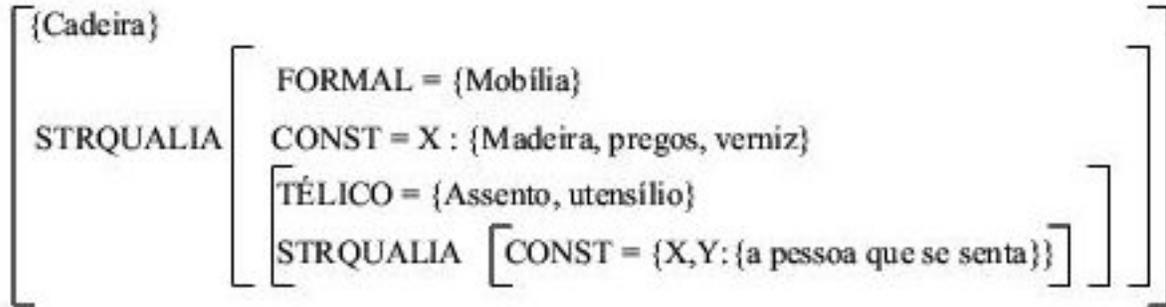
Figura 4: Integração das relações *qualia*.

Acima estão representadas as quatro relações correspondentes aos quatro papéis *qualia*: formal, téllico, agentivo e constitutivo, ao longo, respectivamente, dos eixos z, y, x (à esquerda) e x (à direita). As setas indicam o sentido das relações. Assim, *Utilidade* é o formal de *Assento*, *Assento* é o téllico de *Cadeira*, *Madeira* é o constitutivo de *Cadeira*, etc. Note-se que os elementos das relações não satisfazem a necessidade e a suficiência de descrição uns dos outros. Quando se diz que madeira é constitutivo de cadeira, deve-se entender que o é para um determinado contexto ou em um tempo t.

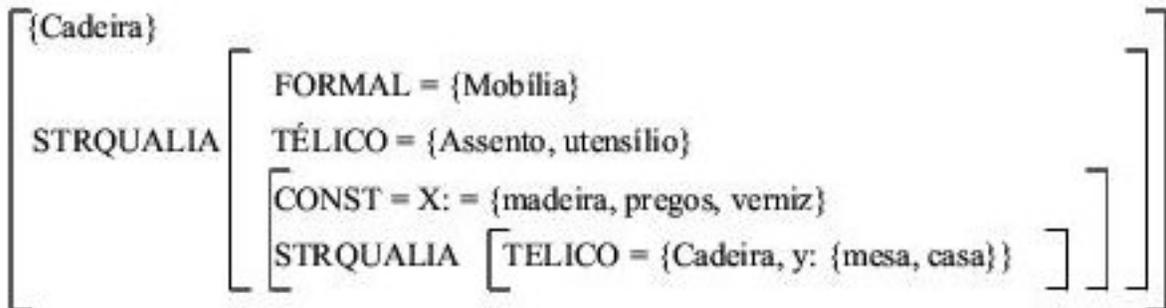
Também é fundamental perceber que os três polígonos da figura são séries proporcionais interligadas pelo eixo z. E que poderiam ser estendidas ao longo dos três eixos pelos quais se desenvolvem, tanto no sentido positivo (da esquerda para a direita, ou da frente para trás) quanto no sentido negativo (o inverso, por convenção).

As relações descritas na Figura 4 são, tradicionalmente, representadas segundo a notação de matrizes, conforme os exemplos abaixo⁴¹:

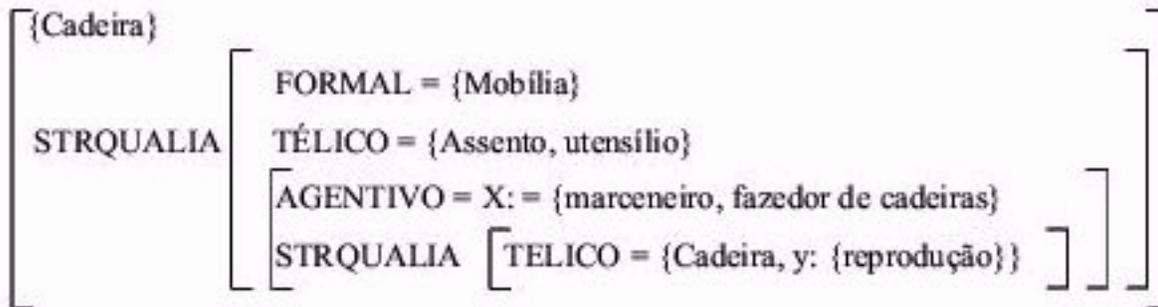
(E1)



(E2)

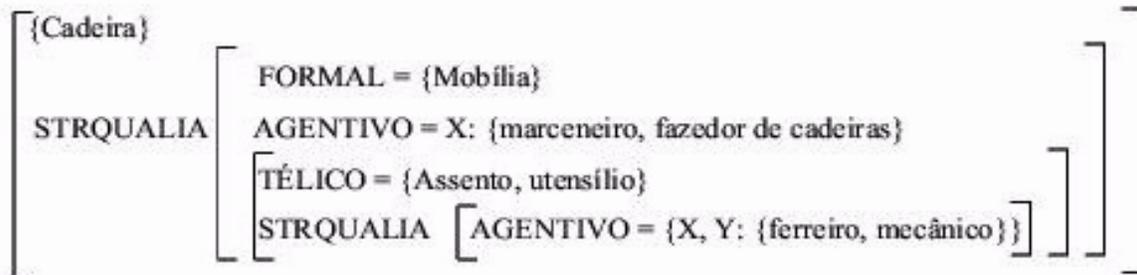


(E3)



⁴¹ Pustejovsky (1998) não proíbe a representação qualia encaixada ou recursiva de que nos utilizamos. Apenas disto se poderia supor a aceitabilidade formal da mesma. No entato, em comunicação pessoal por correio eletrônico (05/10/2005) o autor do GL nos confirmou haver trabalhos de Busa, Boulion e dele mesmo que se utilizam de tal conceito. E que “the notion of having an embedded or recursive qualia structure is perfectly within standard assumptions in GL.”

(E4)



2.3 Integração das relações semântico-conceituais

Considerados os quadros das séries proporcionais descritos acima, podemos deduzir, a partir da propriedade de herança conceitual, as interações entre os eixos das relações *qualia*. A herança de características, de um nó superior para com um nó inferior, em uma árvore hierárquica, não significa que o nó inferior terá todas, e somente todas, as características do nó superior. Ao contrário, por uma questão de diferenciação, um hipônimo compartilhará características mas terá outras que não estavam presentes nos ramos acima. Ou seja, as características do hipônimo são, necessariamente, hipônimas das características correspondentes do hiperônimo. Além disso, pela propriedade do encapsulamento de sentidos⁴², sabemos que o hiperônimo está contido no hipônimo. São estes dois aspectos que norteiam as três primeiras

⁴² Para esclarecer este conceito, optamos por apresentar diretamente sua fonte: “Há duas relações lexicais especificamente associadas com as séries proporcionais. Mas, antes que nós possamos discutí-las, outra relação precisa ser introduzida, a qual, embora tenha um papel proeminente nas séries proporcionais lexicais, não está, em princípio, restrita e estas. Esta relação será chamada de endonímia. É baseada na noção **de encapsulamento semântico**, e envolve a **incorporação de um significado de um item lexical pelo significado de outro**. A característica essencial para definir esta relação é a capacidade de gerar pleonasmos. O teste-mas provê um modo conveniente de ilustrar este aspecto. Todas as sentenças em (2) são pleonasmos:

- 2a É um cavalo, mas é um animal.
- b É um dedo, mas é uma parte de uma mão.
- c. É uma luva, mas serve para cobrir a mão.
- d. É um estábulo, mas serve para cavalos.
- e. Ele me chutou, mas ele me chutou com seu pé.

Pode haver, algumas vezes, problemas em decidir qual membro do par é endônimo e qual é exônimo, porém, na maior parte dos casos, podemos confiar em nossas intuições. **Se os termos são hiperonimicamente relacionados, então o superordenado é a escolha automática para endônimo**; sendo menos específico em

regras abaixo (as letras C, F, A e T correspondem, respectivamente, a constitutivo, formal, agentivo e télico):

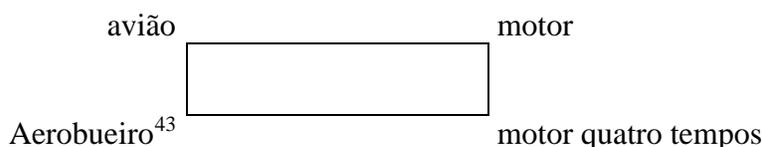
$$(QF1) C(F(x)) \subseteq F(C(x))$$

$$(QF2) A(F(x)) \subseteq F(A(x))$$

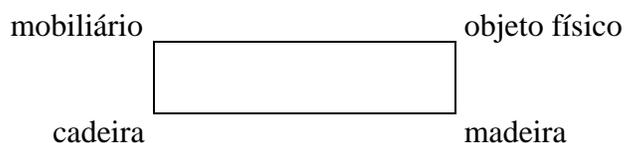
$$(QF3) T(F(x)) \subseteq F(T(x))$$

(QF1) pode ser lida como: o constitutivo do formal de x é um subconjunto do formal do constitutivo de x. Em outras palavras, o mero-hiperônimo de x é hiperônimo do merônimo de x. No caso do Exemplo 1, abaixo, se x = aerobueiro, então o mero-hiperônimo de x é y, tal que y = motor, e este, por sua vez, é hiperônimo do merônimo de x. Vejamos exemplos na forma de séries proporcionais quadradas:

Exemplo 1:



Exemplo 2:

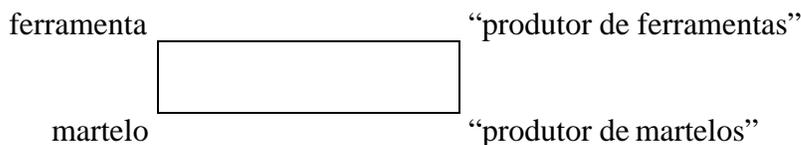


(QF2) e (QF3) são aplicações deste mesmo princípio aos papéis agentivo e télico. A primeira regra é lida como: o agentivo do formal de x é um subconjunto do formal do agentivo de x. E (QF3) pode ser lida como: o télico do formal de x é um subconjunto do formal do télico de x.

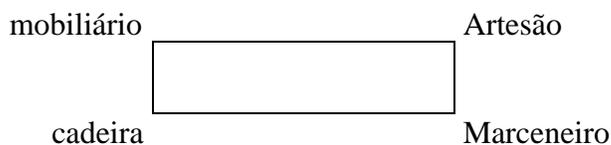
sentido, ele é, por isso, menos complexo semanticamente.” (CRUSE, 1986, p. 123-4, tradução e grifo nossos).
⁴³ Avião monomotor de origem argentina adotado pelo Departamento de Aviação Civil Brasileiro para voos de instrução.

Para a relação de agentividade podemos, também, elaborar os seguintes quadrados proporcionais, nos quais, para os fins das QFs, x deve ser sempre considerado como o termo inferior esquerdo:

Exemplo 3:



Exemplo 4:



Outras interações entre os papéis *qualia* podem ser estabelecidas. Se aceitarmos o preceito kantiano de que nada deve ser considerado outra coisa que não, pelo menos, um fim em si mesmo⁴⁴, obtemos a regra abaixo:

$$(QF4) x \subseteq T(x)$$

A qual pode ser desenvolvida nos seguintes corolários:

$$(QF5) C(x) \subseteq C(T(x))$$

$$(QF6) A(x) \subseteq A(T(x))$$

Veja-se que as duas regras acima podem ser deduzidas mecanicamente por causa de algumas propriedade formais simples. Ambas podem ser simplificadas para (QF4) por anularem-se o C e o A antes e depois de \subseteq .

As próximas proposições podem ser entendidas pelo seguinte raciocínio: aquilo de que é feita a água, a saber, hidrogênio e oxigênio, tem, entre outras finalidades, a própria água. Embora pareça, a primeira vista, anacrônico pensarmos na finalidade de objetos físicos, é plenamente plausível o diálogo: “- Para que servem o hidrogênio e o oxigênio? – Servem para possibilitar a

existência da água”. Mas, uma vez que aquilo de que se compõe a água não está presente exclusivamente na água, não pode ser, a água, sua finalidade exclusiva. Portanto, não precisamos prever a possibilidade de x ser igual ao $T(C(x))$, bem como x ser igual ao $T(A(x))$, por motivo análogo. Por isso nos permitimos, nestas duas últimas regras, a utilização do operador de subconjunto próprio.

$$(QF7) x \subset T(C(x))$$

$$(QF8) x \subset T(A(x))$$

2.4 Produção automática de relações na Wordnet 2.0 por meio de QF1

Finalmente, chegamos às últimas partes desta dissertação. Aqui apresentamos a implementação da regra (QF1) sobre o banco de dados da Wordnet 2.0. Obtivemos estes resultados na fase final de escrita deste trabalho, e, por limitações de *hardware* não pudemos fazer uma verificação dos resultados em grande escala. Mesmo assim, é surpreendente a qualidade das relações de hiperonímia geradas automaticamente. É claro que elas não foram inventadas do nada pela máquina, mas, embora tenham sido deduzidas das relações já existentes na Wordnet, não estavam previstas pelo Grinder, o “motor” de inferências da Wordnet.

Procedemos às relações de hiperonímia deduzidas através da regra (QF1), aplicadas à meronímia de substância e de membro, sem a transitividade da hiperonímia, no banco de dados. O cálculo foi feito com a seguinte expressão, em PROLOG:

**hiponimo(A, X, B, Z) :- hyp(Y, W), hyp(X, K),
ms(X, Y), ms(Z, W),
s(X, C, A, n, G, I),
s(Z, D, B, n, H, J),
K =\= Z, X =\= Z.**

Explicação: A fórmula acima é uma implicação material. O antecedente, ou cabeça da

⁴⁴ Esta interpretação corresponde ao conceito de valor intínseco (cf. DALLGNOLL, 2005).

fórmula está a direita do sinal de implicação: :- (cf. BLACKBURN et. all., 2001). À esquerda está o conseqüente, ou corpo. O conseqüente é um predicado quaternário cujas variáveis representam, A, uma palavra (*word form*) de um *synset* X, que está na relação de subordinada de uma palavra B, de um *synset* Z. Esta relação, nomeada “hipônimo”, se concretiza toda vez que forem satisfeitas as condições impostas na cabeça da implicação. Estas condições são, simplesmente, a formalização da serie proporcional relativa ao quadrado: {<X, - hiperônimo de X>, <- merônimo de X, - hiperônimo do merônimo de X>}. Ou seja, as condições procuram garantir a regularidade destas relações de forma a que, havendo três das relações, a quarta se estabelece. Neste caso, havendo as relações de Y ser hipônimo de W, X ser merônimo de Y, e Z ser merônimo de W, a relação de hiperonímia entre X e Z se estabelece. Ou seja, X é hipônimo de Z. Se procedemos a substituição destas variáveis pelos termos do Exemplo 1 , acima, obtemos as seguintes correspondências: X = motor de quatro tempos; Z = motor; Y = aerobueiro; W = avião. Como garantia de que as relações de hiperonímia geradas não são, coincidentemente, as relações de hiperonímia já estabelecidas no banco de dados da WorNet2.0, foi acrescentada a seguinte condição: se X é hipônimo de K, necessariamente, K deve ser outro *synset* que não Z.

Para evitar situações em que um *synset* seja hiperônimo de si mesmo, estabelecemos a condição de que X seja diferente de Z. Observação: obteríamos outro resultado se usássemos as propriedades de transitividade das relações de hiperonímia e meronímia. Para isso precisaríamos ter acrescentado as regras:

hyp(A, C) :- hyp(A, B), hyp(B, C).

mm(A, C) :- mm(A, B), hyp(B, C).

Infelizmente, isto traria um grande inconveniente para a velocidade do processamento, tornando-o praticamente irrealizável no hardware doméstico de que dispomos (Celeron 799Mh,

conceitos. Ou seja, aquelas que não são estabelecidas pela mera convenção social, e que estão, por isso, sujeitas a um determinado contexto cultural. Neste esquema, {cão policial} não é um hipônimo de {cão}, mas é uma possibilidade de propriedade dos cães. Ou seja, um constitutivo. Cães policiais, por exemplo, o Pupu ou o Gorila, não nasceram {cão policial}, por mais que tenham pendões para esta atividade desde o nascimento. O termo {cão policial}, e esta posição em uma ontologia, depende de uma série de fatores sociais, como, por exemplo, o haver um corpo profissional e institucionalizado chamado polícia. Podemos imaginar sociedades onde não há polícia para as quais o conceito de cão policial não encontraria respaldo na ontologia pela qual se expressam os pensamentos e palavras de tal sociedade. Assim como de uma sociedade para outra o valor da poligamia masculina adquire ou não relação com as ações imorais.

Portanto, cão policial não é nem um hipônimo de cão, nem um tético de cão, mas é uma das possibilidades de atribuição social de característica aos cães. Algo semelhante pode ser pensado em relação aos famosos exemplos do camelo para os árabes e a vaca para os indianos. Há um sem número de subespecificações, que fazem sentido naqueles contextos sociais, e que estão previstas nas respectivas ontologias, para as quais não perceberíamos correlatos em nossas idéias-língua.

Para retificar: cão policial não é um tipo de cão, mas um tipo de propriedade social de cão. E um tipo de propriedade social de cão é, por sua vez, um tipo de propriedade social, que é um tipo de propriedade.

Dito isto, convidamos o leitor a observar o Apêndice iii.i. A semelhança entre as relações a que estamos chamando de hiperonímia, ali estabelecidas, e a relação entre o cão policial e animal de estimação é patente.

Considerações Finais

Neste trabalho iniciamos por estudar questões acerca da ontologia filosófica para compreender a relação desta, ancestral, com o uso deste termo que se faz pela ciência da computação, contemporaneamente. Percebemos que diversas categorias filosóficas têm correlatos em categorias de organização de léxico, ainda que os lexicógrafos não tivessem interesse em especulações filosóficas, nem os filósofos considerassem sua atividade como um tipo de lexicografia.

Avaliamos algumas características da Wordnet na tentativa de incluir as relações *qualia* no esquema da estrutura hierárquica. Defendemos que os chamados papéis *qualia*, ou Causas de Aristóteles, correspondem, em parte, a relações semânticas tradicionalmente praticadas. Como a hiperonímia que corresponde ao papel formal e a meronímia que corresponde ao constitutivo. Tendo em visto este fato, consideramos a possibilidade do desenho de uma rede semântica pautada em tais relações, a que chamamos *QualiaNet*, e estabelecemos regras de inferência lexicais no intuito de estabelecer uma unidade entre os eixos formados pelos papéis *qualia*. Para a visualização deste fenômeno, recorreremos a um diagrama tridimensional. Este, por sua vez, se desmontado em suas faces, deixa ver os quadrados das séries proporcionais. Partindo das séries proporcionais, somadas ao conceito de encapsulamento de sentidos e à herança lexical, deduzimos diversas regras de inferência. Sendo que uma delas, a que estabelece a relação entre os merônimos e hiperônimos, mostrou-se intuitivamente bastante compreensível e adequada para ser testada diante da Wordnet. Uma vez que a Wordnet é provida, largamente, de relações hiperonímicas. Consideramos, então, que o problema da compatibilidade entre co-hipônimos poderia estar na determinação incorreta do hiperônimo. Vimos isto através do exemplo do cão policial o qual, por depender de uma característica socialmente vinculada, e, portanto, contingente, parecia ser mais adequadamente relacionada à *propriedade de cão*, e não ao *cão*

propriamente.

Feito isto, restou apenas demonstrar que as relações obtidas automaticamente por meio de um condicional escrito em PROLOG, para o banco de dados da Wordnet, poderia gerar relações do mesmo tipo que a hiperonímia de cão policial. O que possibilita um grande incremento de informação semântica às Wordnets. Cabe ainda frisar que, assim como foi obtida a hiperonímia, poderiam ser deduzidas quaisquer outras relações que fechassem um dos lados dos quadrados das séries proporcionais, ao longo do eixo tridimensional da *QualiaNet*.

Apêndices

- 1.1 Apêndice i: Tabela de Categorias.
- 1.2 Apêndice ii.i: Categorias de Roget.
- 1.3 Apêndice ii.ii: Categorias de Roget (cont.).
- 1.4 Apêndice iii.i: Pares de Hipônimos/Hiperônimos.
- 1.5 Apêndice iii.ii: Análise e contra-exemplos.

Referências

ABBAGNANO, Nicola. **Dicionário de filosofia**. 2.ed. São Paulo: Martins Fontes, 1998.

AGOSTINHO, Santo. **De Magistro**. São Paulo: Vitor Civita, 1973 [354 - 430].

ARISTÓTELES. **A Metafísica**. Porto Alegre: Globo, 1969. [sec. III a. C.].

-----, **Organon**: Categorias; Da Interpretação; Analíticos Anteriores; Analíticos Posteriores; Tópicos; Refutações Sofísticas. Bauru: Edipro, 2005. [sec. III a. C.].

BARBOSA, Jeronymo Soares. **Gramática philosophica da língua portuguesa**: ou princípios da gramática geral aplicados à nossa linguagem. Lisboa: Academia Real das Ciências, 1871.

BERNES-LEE, Tim; HENDLER, James; ORA, Lassila. The semantic web. **Scientific American**, [S. l.], n. 17, 2001.

BERTOLI, Valmor; ZAGO, Lumar Junior; FLORES, Ricardo; LUZ, Pinto Rodolfo da. Sistema UNL: uma solução para a tradução automática e sistemas de conhecimento. **Revista de Automação e Tecnologia da Informação**, Florianópolis, v. 2, n. 2, jul./dez. 2003. Disponível em: <www.ctai.senai.br/revista4artigo.htm>. Acesso em: 27 de setembro de 2005.

BÉZIEAU, Jean-Yves. **Tendências Atuais da Filosofia**: edição bilíngue. Florianópolis: Edições Nefelibata, 2003.

BIDERMAN, Maria Tereza Camargo. **Teoria Lingüística**: teoria lexical e lingüística computacional. 2 ed. São Paulo: Martins Fontes, 2001

BITTENCOURT, G. **Inteligência Artificial**: ferramentas e teorias. Florianópolis: EdUFSC, 1998

BLAKBURN, Patrick; BOS, Johan; STREIGNITZ, Kristina. **Learn Prolog Now!**. [S. l.]: 2001. Disponível em: <<http://www.coli.uni-sb.de/kris/learn-prolog-now>>. Acesso em: 10/09/2005.

BRAIDA, Celso. Reni. **Conceitos ontológicos básicos**. Material ministrado na disciplina de Filosofia da Linguagem, Pós-graduação em Filosofia, Área de Concentração Ontologia, Florianópolis: UFSC, 2002/02.

BRITO, Glaucia da Silva. **Dicionário informatizado e dicionário eletrônico**: novas estruturas para novas ferramentas. Florianópolis, 2002. Tese (Pós-graduação em Letras) — Programa de Pós-graduação em Linguística, Universidade Federal de Santa Catarina.

CAMPOS, Haroldo de. **Ideograma**: logica, poesia, linguagem. São Paulo: EDUSP, 1977.

CARNAP, R. Empiricism, Semantics and Ontology. **Revue Internationale de Philosophie**, [s.l.], v. 4, 1950.

CHOMSKY, Noam. **New horizons in the study of language and mind**. Cambridge, Reino Unido: Cambridge University Press, 2000.

CLEMENTS, George Nick. The geometry of phonological features. **Phonology Yearbook**, v. 2, p. 225-252, 1985.

COLLINS, A. M.; QUILLIAN, M. R. Retrieval time from semantic memory. **Journal of Verbal Learning and Verbal Behavior**, v. 8, p. 240-247, 1969.

CRUSE, Allan. **Lexical Semantics**. Cambridge, Reino Unido: Cambridge University Press, 1986.

DALL'AGNOL, Darley. **Valor Intrínseco**. Florianópolis: UFSC, 2005.

DAVIDSON, Donald. **O método da verdade em metafísica**. In: *Inquires into Truth and Interpretation*. 2. ed. Oxford: Clarendon, 2001.

FELLBAUM, Christiane. English Verbs as a Semantic Net. **Journal of Lexicography**, v. 3, p. 278-301, 1990.

----- . A Semantic Network of English Verbs, In: FELLBAUM, Christiane (Org.). **Wordnet**: an electronic lexical database, Massachusetts: The MIT Press, 1998.

FELLBAUM, Christiane; BOYD-GRABE, Jordan; OSHERSON, Daniel; SCHAPIRE, Robert. **Adding Dense, Weighted Connections to WORDNET**. 2005. (No prelo).

FERNANDES, Rodrigo Panchiniak. Uma tetravalência de verdade para paradoxos semânticos. **Working Papers em Lingüística**, Florianópolis, v. 7, p. 95-105, 2003.

FERREIRA, Aurélio Buarque de Holanda. **Minidicionário Aurélio**. Rio de Janeiro: Nova Fronteira, 1988.

FREGE, Gottlob. **Logica e filosofia da linguagem**. São Paulo: Cultrix: Ed. da USP, 1978 [1884].

----. Sobre a justificação científica de uma conceitografia. In **Os Pensadores**: vol. XXXVI. São Paulo: Victor Civita, 1974 [1884].

HARABAGIU, Sanda M; MOLDOVAN, Dan I. Knowledge Processing on na Extended Wordnet. In: FELLMAUM, Christiane (Org.). **Wordnet**: an eletronical lexical database. Massachusets: The MIT Press, 1998.

HEIDEGGER, Martin. **Introdução à metafísica**. Rio de Janeiro: Tempo Brasileiro, 1969.

HOLMES, Ernest. **The Science of the Mind**. [s.l.]: Tarcher, 1997.

HUSSERL, Edmund; MERLEAU-PONTY, Maurice; **Investigações lógicas**: sexta investigação (elementos de uma elucidação fenomenologica do conhecimento). São Paulo: Abril Cultural, 1975.

KANT, Emmanuel. **Fundamentação da Metafísica dos Costumes**. [S.l.]: Ediouro, 1998 [1785].

KNEALE, W.; KNEALE, M. **O Desenvolvimento da Lógica**. 2 ed. Lisboa: Fundação Calouste Gulbekian, 1980.

KRAUSE, Décio. **Introdução aos Fundamentos Axiomáticos da Ciência**. São Paulo: EPU, 2002

KRAUSE, Décio; COELHO, Antônio M. N. **Identity, indicernability, and philosophical claims**. 2000. Disponível em: <www.cce.ufsc.br/fil/dkrause.htm>. Acesso em: 10/10/2004.

LAERTIOS, D. **Vidas e doutrinas dos filósofos ilustres**. 2 ed. Brasília: Editora da Universidade

de Brasília, 1977. [sec. IV].

MANNING, Christopher D.; SCHÜTZE, Hinrich. **Foundations of statistical natural language processing**. Cambridge: MIT Press, 1999.

MARRAFA, Palmira. **Wordnet do Português**: uma base de dados de conhecimento lingüístico. Lisboa: Instituto Camões, 2002a.

----- Wordnet: General Architecture and Internal Semantics Relations. **DELTA**, v. 18, 2002b. Edição especial.

----- Extending Wordnets to Implicit Information. In: **Proceedings of LREC 2004**: International Conference on Language Resources and Evaluation. Lisboa, maio 2004.

MARRAFA, Palmira; MOURA, Heronides Maurílio Melo de. Portuguese Telic Causative Verbs. In: **Proceedings of the third international workshop on generative approaches to the lexicon**. Genève, p.144-148, 2005.

MENDES, Sara, CHAVES, Rui Pedro. Enriching Wordnet with *Qualia* Information. In: NAACL 2001 CONFERENCE. **Workshop on Wordnet and Other Lexical Resources**. Pittsburgh, 2001.

MILLER, G.; et. all. Five Papers on Wordnet. **CSL Report 43**. Princeton: Princeton University, 1990.

MILLER, G. Nouns in Wordnet. In: FELLBAUM, Christiane (Org.). **Wordnet**: An Eletronic Lexical Database. Massachusets: The MIT Press, 1998a.

-----Foreword. In: FELLBAUM, Christiane (Org.), **Wordnet**: An Eletronic Lexical Database. Massachusets: The MIT Press, 1998b.

MIOTO, C.; FIGUEIREDO SILVA, M.C.; LOPES, R.E.V. **Novo Manual de Sintaxe**. Florianópolis: Insular, 2004.

MORAVCSIK, J. **Thought and language**. Londres: Routledge, 1990.

MORTARI, Cesar. **Introdução à Lógica**. São Paulo: UNESP, 2001.

MOURA, Heronides Maurílio de Melo. **Significação e contexto**: uma introdução a questões de semântica e pragmática. Florianópolis: Insular, 1999.

-----. Dénotation et argumentation dans le discours. **Langages**, v. 142, n. 1, p. 77-91, 2001.

-----. **Notas sobre a teoria de Frege**. Material ministrado nas aulas de Semântica na Pós-graduação em Linguística. Florianópolis: UFSC, 2003.

MOURA, Heronides Maurílio de Melo; PEREIRA, J. S. V. A interface léxico-enciclopédia no léxico gerativo: um estudo do verbo preparar. **Revista da ANPOLL**. São Paulo, v. 16, n. 1, p. 57-73, 2004.

NOY, Natalya F.; McGUINNESS, Deborah L. **Ontology Development 101**: A guide to Creating your first ontology. [sl], [sd]. Disponível em: <<http://protege.stanford.edu>>. Acessado em: 29/06/2005.

PETERS, F. E. **Termos filosóficos gregos**. 2.ed. Lisboa: Calouste Gulbenkian, [s.d]

PLATÃO. **Platonis Opera**: Evthyphro, Apologia, Crito Phaedo, Cratylvs, Theaetets, Sphista, Politicvs. Reino Unido, Oxford: Oxford University Press, 1995. [sec. III a. C].

POSNER, Michael I. **Cognição**. Rio de Janeiro: Interamericana, 1980.

PUSTEJOVSKY. The Generative Lexicon. **Computational Linguistics**, v. 17, p. 409-44 1, 1991.

-----. **The Generative Lexicon**. Cambridge: MIT Press, 1995.

PUSTEJOVSKY, J.; BOULLION, P. Aspectual Coercion and Logical Polysemy. In: PUSTEJOVSKY, J; BOULLION, P. (Org.). **Lexical semantics**: the problem of polysemy. Oxford: Clarendon Press, 1996.

QUINE, Willard Van Orman. **From a logical point of view**. 2 ed. Cambridge: Harvard University Press, 1980.

RECTOR, Mônica; YUNES, Eliana. **Manual de Semântica**. Rio de Janeiro: Ao livro técnico, 1980.

RIGAU, German; et. all. Framework and Results for the Spanish SENSEVAL. In: SENSEVAL-2, 2001, Toulouse. **Proceedings of SENSEVAL-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems**. Toulouse: ACL, 2001. p. 41-45.

ROGET, S. **Roget's International Thesaurus**. 4ed. New York: Harper and Row, 1977.

RUSSELL, Bertrand; MOORE, G. E. **Logica e conhecimento**. São Paulo: Abril Cultural, 1974.

SANTOS, Leonor. **Significados lexicais como esquemas explanatórios**. Material apresentado ao Grupo de Estudos Lexicais, Pós-graduação em Lingüística, Florianópolis: UFSC, 2003.

SAUSSURE, Ferdinand de. **Curso de Lingüística Geral**. Buenos Aires: Losada, 1967.

SCHANK, R. Conceptual Dependency: A Theory of Natural Language Understanding. **Cognitive Psychology**, v. 3 (4), 1972.

SILVA FILHO, Olavo da; CABRERA, Julio. **Inferências Lexicais e Interpretação-rede de Predicados**. Brasília, 2004. No prelo.

-----. Inferências Lexicais e Interpretação-rede de Predicados. In: III COLÓQUIO NACIONAL DE FILOSOFIA DA LINGUAGEM. **Anais**. São Leopoldo: Unisinos, 2005.

SUPPES, Patrick. **Teoria Axiomática de Conjuntos**. Cali: Norma, 1968

SVENSEN, Bo. **Practical lexicography** : principles and methods of dictionary-making. Oxford: Oxford University, 1993.

THINK MAP. **Thinkmap Sdk v. 2.5**: the technical whitepaper. New York, 2004

UNDL FOUNDATION. **The Universal Networking Language (UNL)**: specifications. [S.l.]: UNL Center, 2004.

VOSSEN, Pick. **The EuroWordnet Base Concepts and Top Ontology**. University of Amsterdam, 1998. Disponível em: <www.hum.uva.nl/ewn>. Acesso em: 15/09/2005.

WIERZBICKA. Apples are not a “kind of fruit”. **American Ethnologist**, v. II, p. 313-328, 1984.

Apêndice i: Tabela de Categorias.

		FILOSOFOS														
		1	2	3	4	5	6		7		8	9	10		11	
		Platão e Plotino	Aristóteles	Estóicos	Wolf	Kant		Renouvier		Cohen		Hegel	Heidegger	Hartmann		Husserl
Categorias:	1	Ser	Substância	Substância	Deus	Quantidade:	Unidade	Especificações Da Categoria Fundamental:	Relação	Categoria Fundamental:	Sistema	Eu	Existenciais	Modais	Propriedade	
	2	Movimento	Quantidade	Qualidade	Alma Humana		Multiplicidade		Número				Ser-aí	Bipolares		Qualidade
	3	Repouso	Qualidade	Modo de Ser	Coisas Naturais		Totalidade		Extensão				Ser-das-coisas	Do real:		Princípio do valor
	4	Identidade	Relação	Relação		Realidade	Duração					Princípio da crença	Relação			
	5	Alteridade	Lugar			Negação	Qualidade					Princípio da planificação	Conjunto			
	6		Tempo			Limitação	Devir					Princípio da dependência	Número			
	7		Posição			Inerência e subsistência	Força						Etc. (?)			
	8		Ter			Causalidade e dependência	Finalidade									
	9		Agir			Comunhão	Personalidade									
	10		Sofrer			Possibilidade-impossibilidade										
	11					Existência-inexistência										
	12					Necessidade-contingência										
	13															

Esta Tabela foi confeccionada a partir, principalmente, do verbete Categorias, do Dicionário de Filosofia de Nicola Abbagnano (1998).

Apêndice ii.i: Categorias de Roget.

Plano de classificação	
Classe	Seção
Relações Abstratas	Existência
	Relação
	Quantidade
	Ordem
	Número
	Tempo
	Mudança
	Causação
Espaço	Espaço em geral
	Dimensões
	Forma
	Movimento (Motion)
Matéria	Matéria am geral
	Matéria Inorgânica
	Matéria orgânica
Intelecto: o exercício da mente	
Divisão um:	Geral
Formação de idéias	Condições e operações precursoras
	Material para o raciocínio
	Processos de raciocínio
	Resultados do raciocínio
	Extensão do pensamento
Divisão dois:	Pensamento criativo
	Natureza das idéias comunicadas
Comunicação de idéias	Modos de comunicação
	Significados das idéias comunicadas
Volição: o exercício da vontade	
Divisão um:	Volição em geral
Volições individuais	Volição prospectiva
	Ação voluntária
	Antagonismo
	Resultados da ação
Divisão dois:	Volição social geral
Volições sociais	Volição social especial
	Volição social condicional
	Prelações possessivas
Emoção, religião e moralidade	Geral
	Emoções pessoais
	Emoções Interpessoais
	Moralidade
	Religião

Apêndice ii.ii: Categorias de Roget (continuação).

Sinopse de Categorias			
Classe um: Relações Abstratas	Cabeças/ Índices (heads)		
Existência			
Abstrato:	Existência		Inexistência
Concreto:	Substancialidade		Insubstancialidade
Formal: (interno/externo)	Intrinsicalidade		Extrinsicalidade
Modal: (absoluto/ relativo)	Estado		Circunstância
Relação			
Absoluto:	Relação		Não-relação
	Cosanguinidade		
	Correlação		
	Identidade		Contrariedade
		Diferença	
Continua:	Uniformidade		Não-uniformidade
Parcial:	Similaridade		Dissimilaridade
	Imitação		Originalidade
	Coópia		Protótipo
Geral:	Concordância		Discordância
Quantidade			
Simples: (absoluto/ relativo)	Quantidade		Grau
	Equalidade		Inequalidade
Comparativo:		Média	
		Compensação	
(por comparação com um padrão)	Maior		Menor
(por comparaçã com um objeto)	Superior		Inferior
(mudanças na quantidade)	Incremento		Decréscimo
Conjuntivo:	Adição		Subtração
	Adjunção		Remainder (?)
			Decrement
	Mistura		Pureza
	União		Desunião
	Bond (?)		
	Coerência		Incoerência
	Combinação		Decomposição
	Todo		Parte
Concreto:	Completude		Incompletude
	Composição		Exclusão
	Componente		Estranho
Ordem			
Geral:	Ordem		Desordem
	Arranjo		Desarranjo
Consecutivo:	Precedente		Conseqüente
	Precursor		Sequel (?)
	Início		Fim
		Meio	
	Continuidade		Descontinuidade
	Termo		
Coletivo:	Assemblage (?)		Nonassembly (?)
	Foco		
Distributivo:	Classe		

	Inclusão		
	Generalidade		Especialidade
Categorial:	Regra		Multiformidade
	Conformidade		Inconformidade
Número			
Abstrato:	Número		
	Nymeração		
	Lista		
Determinado	Unidade		Accompaniment (?)
	Dualidade		
	Duplicação		Biseção
	Triplidade		
	Triplicação		Triseção
	Quaternidade		
	Quadruplicação		Quadriseção
	Cinco ou mais		Multiseção
Indeterminado	Pluralidade		Fração
			Zero
	Muitos		Poucos
	Repetição		
	Infinidade		
Tempo			
Absoluto: (definido/ indefinido)	Tempo		Nunca
	Período		Curso
	Duração Contingente		
	Longa duração		Transição
	Perpetuidade		Instantaneidade
	Cronométrico		Anacrônico
Relativo: (a sucessão)	Prioridade		Posteridade
	Tempo presente		Tempo diferente
	Sincronismo		
(a um período)	Futuro		Passado
	Novo		Velho
	Manhã		Tarde

Apêndice iii.i: Pares de Hipônimos/Hiperônimos (*synsets* retirados da Wordnet 2.0)

Pares de hipônimos/hiperônimos localizados automaticamente por meio da regra QF1	
Busca sobre banco de dados de mero-substância	Busca sobre banco de dados de mero-membro
Notação: X no sentido Y é hipônimo de Z no sentido W	
X = hemp Y = 114058371 Z = warp W = 104374825 ;	X = animal, animate_being, beast, brute, creature, fauna Y = 100012748 Z = phylum W = 107610999 ;
X = hemp Y = 114058371 Z = woof, weft, filling, pick	X = animal_order Y = 101268597 Z = phylum W = 107610999 ;
X = chenille Y = 102905916 Z = warp W = 104374825 ;	X = arthropod_family Y = 101679576 Z = phylum W = 107610999 ;
X = chenille Y = 102905916 Z = woof, weft, filling, pick W = 104420795 ;	X = arthropod_genus Y = 101682919 Z = phylum W = 107610999 ;
X = cotton Y = 114023275 Z = warp W = 104374825	X = 'Parazoa', subkingdom_Parazoa Y = 101824523 Z = phylum W = 107610999 ;
X = cotton_wool Y = 114023275 Z = warp W = 104374825 ;	X = coelenterate_family Y = 101826453 Z = phylum W = 107610999 ;
X = cotton Y = 114023275 Z = woof, weft, filling, pick W = 104420795 ;	X = coelenterate_genus Y = 101826581 Z = phylum W = 107610999 ;
X = cotton_wool Y = 114023275 Z = woof, weft, filling, pick W = 104420795 ;	X = 'Metazoa', subkingdom_Metazoa, ctenophore_family Y = 101826869 Z = phylum W = 107610999 ;
X = norethynodrel Y = 113904887 Z = progesterone, 'Lipo-Lutin', W = 113904141 ;	X = ctenophore_genus Y = 101836111 Z = phylum
X = mestranol Y = 113909267	

Z = progesterone, 'Lipo-Lutin',
W = 113904141 ;

X = flax
Y = 114023026
Z = warp
W = 104374825 ;

X = flax
Y = 114023026
Z = woof, weft, filling, pick
W = 104420795 ;

X = norethindrone
Y = 113904511
Z = progesterone, 'Lipo-Lutin',
W = 113904141 ;

X = norethindrone_acetate
Y = 113904511
Z = progesterone, 'Lipo-Lutin'
W = 113904141 ;

X = norethandrolone
Y = 113904511
Z = progesterone, 'Lipo-Lutin'
W = 113904141 ;

X = 'Norlutin'
Y = 113904511
Z = progesterone
W = 113904141 ;

X = estradiol
Y = 113908367
Z = progesterone
W = 113904141 ;

X = oestradiol
Y = 113908367
Z = progesterone
W = 113904141 ;

X = norgestrel
Y = 113905101
Z = progesterone
W = 113904141 ;

X = estradiol
Y = 113908367
Z = progesterone
W = 113904141 ;

X = norethindrone

W = 107610999 ;

X = worm_family
Y = 101839517
Z = phylum
W = 107610999 ;

X = worm_genus
Y = 101839845
Z = phylum
W = 107610999

X = 'Ebola_virus'
Y = 101258120
Z = 'Bunyaviridae'
W = 101257456

X = 'Ebola_virus'
Y = 101258120
Z = 'Togaviridae'
W = 101258458

X = 'Ebola_virus'
Y = 101258120
Z = 'Flaviviridae'
W = 101258766

X = 'Ebola_virus'
Y = 101258120
Z = 'Arenaviridae'
W = 101259195 ;

X = 'Ebola_virus'
Y = 101258120
Z = 'Reoviridae'
W = 101259723 ;

X = 'Ebola_virus'
Y = 101258120
Z = subfamily
W = 107617017 ;

X = chordate
Y = 101389442
Z = subphylum
W = 107611743 ;

X = craniate
Y = 101394664
Z = subphylum
W = 107611743 ;

X = vertebrate

Y = 113904511

Z = progesterone

W = 113904141

X = strychnine

Y = 113874214

Z = physostigmine

W = 111738735

Y = 101394664

Z = class

W = 107612221

X = bird_family

Y = 101427078

Z = class

W = 107612221

Apêndice iii.ii: Análise e contra-exemplos (casos constantes do Apêndice iii.i)⁴⁷

Exemplo 1: hemp/ warp

EIXO FORMAL		
EIXO CONSTITUTIVO	A = hemp X = 114058371	B = warp Z = 104374825
	C = canvas Y = 102848775	D = fabric W = 103189674
<p>Sense 1</p> <p>{14058371} <noun.substance> hemp#1 -- (a plant fiber)</p> <p>=> {14108918} <noun.substance> plant fiber#1, plant fibre#1 -- (fiber derived from plants)</p> <p>=> {14108532} <noun.substance> natural fiber#1, natural fibre#1 -- (derived from plants or animals)</p> <p>=> {14020709} <noun.substance> fiber#1, fibre#1 -- (a slender and greatly elongated solid substance)</p> <p>=> {13746072} <noun.substance> material#1, stuff#1 -- (the tangible substance that goes into the makeup of a physical object; "coal is a hard black material"; "wheat is the stuff they use to make bread")</p> <p>=> {00017572} <noun.Tops> substance#1, matter#1 -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")</p> <p>=> {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))</p> <p>=> {14140602} <noun.substance> plant product#1 -- (a product made from plant material)</p> <p>=> {14140423} <noun.substance> plant material#1 -- (material derived from plants)</p> <p>=> {13746072} <noun.substance> material#1, stuff#1 -- (the tangible substance that goes into the makeup of a physical object; "coal is a hard black material"; "wheat is the stuff they use to make bread")</p> <p>=> {00017572} <noun.Tops> substance#1, matter#1 -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")</p> <p>=> {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))</p>		

⁴⁷ Neste Apêndice apresentamos quatro séries proporcionais inteiras, que completam exemplos constantes do Apêndice iii.i. Por meio destas série podemos observar não apenas a relação de hiperonímia automaticamente produzida, mas as demais relações presentes na Wordnet 2.0 e que satisfizeram o condicional aplicado sobre as bases de dados em PROLOG. Também apresentamos, abaixo das séries, os enlances hiperonímicos dos membros da primeira coluna das séries, os quais servem como contra-exemplos. Os números dos synsets em **negrito** ressaltam que a relação hiperonímica inferida (representada na primeira linha dos quadrados proporcionais) não estava presente, originalmente, no banco de dados lexicográfico da Wordnet.

Sense 1

{02848775} <noun.artifact> canvas#1, canvass#7 -- (heavy closely woven fabric (used for clothing or chairs or sails or tents))

=> {03189674} <noun.artifact> fabric#1, cloth#1, material#4, textile#1 -- (artifact made by weaving or felting or knitting or crocheting natural or synthetic fibers; "the fabric in the curtains was light and semitransparent"; "woven cloth originated in Mesopotamia around 5000 BC"; "she measured off enough material for a dress")

=> {00019244} <noun.Tops> artifact#1, artefact#1 -- (a man-made object taken as a whole)

=> {00016236} <noun.Tops> object#1, physical object#1 -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00002645} <noun.Tops> whole#2, whole thing#1, unit#6 -- (an assemblage of parts that is regarded as a single entity; "how big is that part compared to the whole?"; "the team is a unit")

=> {00016236} <noun.Tops> object#1, physical object#1 -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Exemplo 2: norethynodrel/ progesterone

		EIXO FORMAL	
EIXO CONSTITUTIVO	A = norethynodrel X = 113904887	B = progesterone Z = 113904141	
	C = 'Enovid' Y = 103170313	D = pill W = 103788550	
Sense 1			
<p>{13904887} <noun.substance> norethynodrel#1 -- (a progesterone derivative used in oral contraceptives and in the control of menstruation and the treatment of abnormal uterine bleeding)</p> <p>=> {13905432} <noun.substance> progestin#1, progestogen#1 -- (any of a group of steroid hormones that have the effect of progesterone)</p> <p>=> {13903746} <noun.substance> steroid hormone#1, steroid2#2, sex hormone#1 -- (any hormone affecting the development and growth of sex organs)</p> <p>=> {05097833} <noun.body> hormone#1, endocrine#1, internal secretion#1 -- (the secretion of an endocrine gland that is transmitted by the blood to the tissue on which it has a specific effect)</p> <p>=> {05095511} <noun.body> secretion#2 -- (a functionally specialized substance (especially one that is not a waste) released from a gland or cell)</p> <p>=> {05089633} <noun.body> liquid body substance#1, bodily fluid#1, body fluid#1, humor#6, humour#4 -- (the liquid parts of the body)</p> <p>=> {04960498} <noun.body> body substance#1 -- (the substance of the body)</p> <p>=> {00017572} <noun.Tops> substance#1, matter#1 -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")</p> <p>=> {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))</p>			
Sense 1			
<p>{03170313} <noun.artifact> Enovid#1 -- (trade name for an oral contraceptive containing mestranol and norethynodrel)</p> <p>=> {03788550} <noun.artifact> pill2#5, birth control pill#1, contraceptive pill#1, oral contraceptive pill#1, oral contraceptive#1, anovulatory drug#1, anovulant#1 -- (a contraceptive in the form of a pill containing estrogen and progestin to inhibit ovulation and so prevent conception)</p> <p>=> {02984546} <noun.artifact> contraceptive#1, preventive2#3, preventative2#3, contraceptive device#1, prophylactic device#1, birth control device#1 -- (an agent or device intended to prevent conception)</p> <p>=> {03068033} <noun.artifact> device#1 -- (an instrumentality invented for a particular purpose; "the device is small enough to wear on your wrist"; "a device intended to conserve water")</p> <p>=> {03443493} <noun.artifact> instrumentality#3, instrumentation#1 -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)</p> <p>=> {00019244} <noun.Tops> artifact#1, artefact#1 -- (a man-made object taken as a whole)</p> <p>=> {00016236} <noun.Tops> object#1, physical object#1 -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")</p> <p>=> {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))</p> <p>=> {00002645} <noun.Tops> whole#2, whole thing#1, unit#6 -- (an assemblage of parts that is regarded as a single entity; "how big is that part compared to the whole?"; "the team is a unit")</p> <p>=> {00016236} <noun.Tops> object#1, physical object#1 -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")</p> <p>=> {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))</p>			

Exemplo 3: strychnine/ physostigmine

EIXO FORMAL		
EIXO CONSTITUTIVO	A = strychnine X = 113874214	B = physostigmine Z = 111738735
	C = nux_vomica Y = 103692387	D = medicine W = 103600430

Sense 1
 {13874214} <noun.substance> strychnine#1 -- (an alkaloid plant toxin extracted chiefly from nux vomica; formerly used as a stimulant)
 => {14183743} <noun.substance> plant toxin#1, phytotoxin#1 -- (any substance produced by plants that is similar in its properties to extracellular bacterial toxin)
 => {14181323} <noun.substance> toxin#1 -- (a poisonous substance produced during the metabolism and growth of certain microorganisms and some higher plant and animal species)
 => {14179654} <noun.substance> poison#1, poisonous substance#1 -- (any substance that causes injury or illness or death of a living organism)
 => {00017572} <noun.Tops> substance#1, matter#1 -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")
 => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
 => {13872080} <noun.substance> alkaloid#1 -- (natural nitrogen-containing bases found in plants)
 => {13886449} <noun.substance> organic compound#1 -- (any compound of carbon and another element or a radical)
 => {13974154} <noun.substance> compound#1, chemical compound#1 -- ((chemistry) a substance formed by chemical union of two or more elements or ingredients in definite proportion by weight)
 => {00017572} <noun.Tops> substance#1, matter#1 -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")
 => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 1
 {03692387} <noun.artifact> nux vomica#1 -- (a medicine made from the seeds of an Asiatic tree; contains strychnine and brucine; formerly used as a stimulant)
 => {03600430} <noun.artifact> medicine#2, medication#1, medicament#1, medicinal drug#1 -- ((medicine) something that treats or prevents or alleviates the symptoms of disease)
 => {03129572} <noun.artifact> drug#1 -- (a substance that is used as a medicine or narcotic)
 => {13935705} <noun.substance> agent#2 -- (a substance that exerts some force or effect)
 => {00005598} <noun.Tops> causal agent#1, cause#4, causal agency#1 -- (any entity that causes events to happen)
 => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Exemplo 4: cotton/ warp

EIXO FORMAL		
EIXO CONSTITUTIVO	A = cotton X = 114023275	B = warp Z = 104374825
	C = cotton_cloth Y = 103002171	D = fabric W = 103189674

Sense 1
 {13874214} <noun.substance> strychnine#1 -- (an alkaloid plant toxin extracted chiefly from nux vomica; formerly used as a stimulant)
 => {14183743} <noun.substance> plant toxin#1, phytotoxin#1 -- (any substance produced by plants that is similar in its properties to extracellular bacterial toxin)
 => {14181323} <noun.substance> toxin#1 -- (a poisonous substance produced during the metabolism and growth of certain microorganisms and some higher plant and animal species)
 => {14179654} <noun.substance> poison#1, poisonous substance#1 -- (any substance that causes injury or illness or death of a living organism)
 => {00017572} <noun.Tops> substance#1, matter#1 -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")
 => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
 => {13872080} <noun.substance> alkaloid#1 -- (natural nitrogen-containing bases found in plants)
 => {13886449} <noun.substance> organic compound#1 -- (any compound of carbon and another element or a radical)
 => {13974154} <noun.substance> compound#1, chemical compound#1 -- ((chemistry) a substance formed by chemical union of two or more elements or ingredients in definite proportion by weight)
 => {00017572} <noun.Tops> substance#1, matter#1 -- (that which has mass and occupies space; "an atom is the smallest indivisible unit of matter")
 => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Sense 1
 {03692387} <noun.artifact> nux vomica#1 -- (a medicine made from the seeds of an Asiatic tree; contains strychnine and brucine; formerly used as a stimulant)
 => {03600430} <noun.artifact> medicine#2, medication#1, medicament#1, medicinal drug#1 -- ((medicine) something that treats or prevents or alleviates the symptoms of disease)
 => {03129572} <noun.artifact> drug#1 -- (a substance that is used as a medicine or narcotic)
 => {13935705} <noun.substance> agent#2 -- (a substance that exerts some force or effect)
 => {00005598} <noun.Tops> causal agent#1, cause#4, causal agency#1 -- (any entity that causes events to happen)
 => {00001740} <noun.Tops> entity#1 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))