

ERASMO ROBERTO MARCELLINO



**CONSTRUÇÃO DE UM ONTOLÉXICO PARA
O UNIVERSO LÉXICO-CONCEITUAL DA
INDÚSTRIA DO BORDADO DE IBITINGA**

Araraquara – SP
2011

ERASMO ROBERTO MARCELLINO

CONSTRUÇÃO DE UM ONTOLÉXICO PARA O UNIVERSO LÉXICO-CONCEITUAL DA INDÚSTRIA DO BORDADO DE IBITINGA

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras, Universidade Estadual Paulista – Campus de Araraquara, como parte dos requisitos para a obtenção do título de Mestre em Letras (Área de Concentração: Linguística e Língua Portuguesa).

Linhas de Pesquisa: Estudos do Léxico; Análise fonológica, morfossintática, semântica e pragmática.

Orientador: Prof. Dr. Bento Carlos Dias da Silva

Bolsa: FAPESP

Araraquara – SP
2011

Marcellino, Erasmo Roberto

Construção de um ontoléxico para o universo léxico-conceitual da
Indústria do Bordado de Ibatinga / Erasmo Roberto Marcellino. – 2011
110 f. ; 30 cm

Dissertação (Mestrado em Linguística e Língua Portuguesa) –
Universidade Estadual Paulista, Faculdade de Ciências e Letras, Campus
de Araraquara

Orientador: Bento Carlos Dias-da-Silva

1. Lexicologia. 2. Semântica. I. Título

ERASMO ROBERTO MARCELLINO

CONSTRUÇÃO DE UM ONTOLÉXICO PARA O UNIVERSO LÉXICO-CONCEITUAL DA INDÚSTRIA DO BORDADO DE IBITINGA

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras, Universidade Estadual Paulista – Campus de Araraquara, como parte dos requisitos para a obtenção do título de Mestre em Letras (Área de Concentração: Linguística e Língua Portuguesa).

Linhas de Pesquisa: Estudos do Léxico; Análise fonológica, morfossintática, semântica e pragmática.

Orientador: Prof. Dr. Bento Carlos Dias da Silva

Bolsa: FAPESP

Data da defesa: 01/06/2011

MEMBROS COMPONENTES DA BANCA EXAMINADORA:

Presidente e Orientador: Prof. Dr. Bento Carlos Dias da Silva
Universidade Estadual Paulista/Araraquara.

Membro Titular: Prof^a. Dr^a. Clotilde de Almeida Azevedo Murakawa
Universidade Estadual Paulista/Araraquara.

Membro Titular: Prof^a. Dr^a. Patrícia Tosqui Lucks
Instituto de Controle do Espaço Aéreo da Aeronáutica/São José dos Campos.

Local: Universidade Estadual Paulista
Faculdade de Ciências e Letras
UNESP – Campus de Araraquara

A todos os ibitinguenses que já bordaram, bordam e/ou continuarão bordando...

Em especial à **Arlete**:

talentosa bordadeira (que pena que deixou a máquina de bordar!)

e excepcional mãe (que bom que nunca vai deixar de nos amar!).

AGRADECIMENTOS

Ao criador de tudo que tentamos compreender e, às vezes, até representar em ontologias.

A todos os meus professores, desde os que me alfabetizaram até os que me acompanharam na pós-graduação, e muito especialmente ao Bento, que, em qualquer momento ou local, se doa pacientemente aos orientandos e à ética pessoal e profissional.

Às professoras Clotilde, Gladis e Patrícia, que aceitaram ajudar, com muita boa vontade e propriedade, na concretização desta dissertação. Também à professora Cristina, pela sugestão bem acertada para a epígrafe.

Aos meus pais, os apoiadores incondicionais Laércio e Arlete, à minha avó Isabel, que já se foi mas muito deixou, e aos demais familiares, mesmo os que vivem perguntando “Quando você vai começar a trabalhar?”

Aos ibitinguenses que atuam nos diversos setores da indústria do bordado, particularmente às minhas informantes Aurora, Dona Ana, Dona Glória e Dona Nereide, pela consultoria tão solícita.

Aos amigos

- da convivência diária: Brá e Márcio
- de longa data e de sempre: Bruno, Diogo, Izidoro, Janaia e Leonardo (e suas respectivas famílias)
- do meio acadêmico formal e informal: Alessandra, Alexandre, Aline, Ana, Ariani, Débora, Emerson, Felipe, Filipe, Gisele, Hélio, Janaina, Larissa, Maira, Maria Paula, Patrícia, Priscila, Valdirene

A UNESP e àqueles que a fazem existir enquanto uma instituição de ensino gratuito e de qualidade.

À FAPESP pelo apoio financeiro que facilitou o desenvolvimento deste trabalho.

FÁBULA

Num fabulário ainda por encontrar será um dia lida esta fábula:

A uma bordadora dum país longínquo foi encomendado pela sua rainha que bordasse, sobre seda ou cetim, entre folhas, uma rosa branca. A bordadora, como era muito jovem, foi procurar por toda a parte aquela rosa branca perfeitíssima, em cuja semelhança bordasse a sua. Mas sucedia que umas rosas eram menos belas do que lhe convinha, e que outras não eram brancas como deviam ser. Gastou dias sobre dias, chorosas horas, buscando a rosa que imitasse com seda, e, como nos países longínquos nunca deixa de haver pena de morte, ela sabia bem que, pelas leis dos contos como este, não podiam deixar de a matar se ela não bordasse a rosa branca.

Por fim, não tendo melhor remédio, bordou de memória a rosa que lhe haviam exigido. Depois de a bordar foi compará-la com as rosas brancas que existem realmente nas roseiras. Sucedeu que todas as rosas brancas se pareciam exactamente com a rosa que ela bordara, que cada uma delas era exactamente aquela.

Ela levou o trabalho ao palácio e é de supor que casasse com o príncipe.

No fabulário, onde vem, esta fábula não traz moralidade. Mesmo porque, na idade de ouro, as fábulas não tinham moralidade nenhuma.

Fernando Pessoa (1986, p.69)
Primeira publicação em 'O Jornal' (nº1, Lisboa, 1915)

RESUMO

A cidade de Ibitinga destaca-se nacionalmente no ramo dos bordados, com sua economia voltada quase que exclusivamente à produção dessas peças. Tendo esse contexto como pano de fundo, esta dissertação discute todo o processo linguístico e linguístico-computacional de construção de *ontoléxicos* – constructos formais cuja natureza léxico-conceitual possibilita o desenvolvimento de investigações teóricas (estudos lexicogramaticais) e aplicadas (construção de recursos lexicográficos e para o processamento computacional de informação textual disponível na Web). Em particular, constrói um ontoléxico exploratório que entrelaça conceitos e itens lexicais do *domínio léxico-conceitual da Indústria do Bordado de Ibitinga*. O embasamento teórico-metodológico assenta-se nos estudos de semântica lexical (*wordnets*) e de frames (*framenets*), pura e computacional, de processamento automático de língua natural e de ontologias.

Palavras-chave: semântica lexical; semântica de frames; léxico; ontoléxico; processamento automático de língua natural.

ABSTRACT

Ibitinga, which is a Brazilian town nationally known for its outstanding position in the embroidery business, has its economy almost totally based on the production of a broad selection of embroideries. With this embroidery industry in the backdrop, this master thesis describes the whole linguistic and computational-linguistic process of construction of *ontolexicons* – formal constructs whose lexical-conceptual nature allows for both theoretical (lexical-grammar construction) and applied (lexicographical and ontological resource construction) research. In particular, the study focuses on the design and implementation of a toy ontollexicon for the *Ibitinga Embroidery Industry lexical-conceptual domain*. Its theoretical foundations have drawn heavily on lexical semantics (*wordnets*), frame semantics (*framenets*), natural language processing, and ontologies.

Keywords: lexical semantics; frame semantics; lexicon; ontollexicon; natural language processing.

SUMÁRIO

1. INTRODUÇÃO.....	9
1.1 Motivação e justificativa	10
1.2 Contextualização	10
1.3 Objetivos.....	11
2 FUNDAMENTOS: <i>CORPORA</i> , LÉXICOS, ONTOLOGIAS E EDITOR.....	13
2.1 As fontes de informação do domínio léxico-conceitual IB(I).....	14
2.1.1 Dicionários e enciclopédias	15
2.1.2 Teses, artigos, relatórios, livros e glossários	17
2.1.3 Entrevistas	18
2.1.4 Outros <i>corpora</i> estruturados e a <i>Web como corpus</i>	19
2.1.5 As redes <i>wordnets</i> e <i>framenets</i>	21
2.1.5.1 A rede <i>WordNet</i>	21
2.1.5.2 A rede <i>FrameNet</i>	25
2.2 A estratégia de coleta das informações léxico-conceituais	26
2.3 Os conceitos-chave: léxico, ontologia, e ontoléxico	31
2.3.1 Ontologias e léxicos	35
2.3.2 A interface entre léxico-ontologia	46
2.3.3 A “mão-dupla” entre léxico e ontologia	49
2.4 Um exercício de alinhamento entre léxicos e ontologias	52
2.5 O editor de ontologias Protégé-OWL.....	55
3 CONSTRUÇÃO DO ONTOLÉXICO: SYNSETS, FRAMES E O PROTÉGÉ.....	61
3.1 Sistematização preliminar.....	61
3.2 Ampliação da sistematização	64
3.3 Consolidação da sistematização	70
3.4 Implementação.....	75
4 CONSIDERAÇÕES FINAIS	91
REFERÊNCIAS BIBLIOGRÁFICAS	93
APÊNDICE 1	100
APÊNDICE 2	104
APÊNDICE 3	110

1. INTRODUÇÃO

A riqueza lexical que se manifesta na realidade da indústria do bordado de Ibitinga incentiva o recorte que fizemos nesta pesquisa, em que tratamos do domínio léxico-conceitual que chamamos de INDÚSTRIA DO BORDADO DE IBITINGA, que pressupõe o domínio léxico-conceitual mais amplo, o domínio léxico-conceitual INDÚSTRIA DO BORDADO em geral. Daí as abreviaturas: IB e IBI ou IB(I), quando não se fizerem necessárias distinções entre ambos. Ao falarmos do aspecto lexical e do conceitual relativos a esse domínio, nos inserimos no constructo de interface denominado **ontoléxico**, que nos leva a estudar, para construí-lo, os dois elementos que o embasam: as ontologias e os léxicos.

A análise das ontologias e dos recursos lexicais, nas suas variadas características pertinentes a uma pesquisa no âmbito da linguística-computacional, bem como a análise da interação complexa entre eles, nos colocam diante de diferentes teorias (como a da semântica lexical), metodologias (como as que permitem que ontologias e léxicos sejam estruturados, formalizados, conectados, etc.) e outros constructos (como as *wordnets* e a *framenets*). Esses e outros tópicos serão considerados ao longo da dissertação, que apresenta, nesta primeira seção, a motivação e a justificativa para o estudo (1.1), uma breve contextualização (1.2) e os objetivos (1.3). Na seção 2, *Fundamentos: corpora, léxicos, ontologias e editor*, apresentamos o aparato teórico-metodológico, que consiste na (2.1) especificação das fontes de informação do domínio léxico-conceitual IB(I) (dicionários e enciclopédias, teses, artigos, relatórios, livros, glossários, entrevistas, outros *corpora* estruturados, a “Web como corpus” e as redes *wordnets* e *framenets*, que são ao mesmo tempo *corpus* e estruturas que auxiliam na construção de léxicos estruturados), na (2.2) descrição da estratégia de coleta das informações léxico-conceituais adotada, na (2.3 e 2.4) avaliação das possibilidades de associação entre léxicos e ontologias e (2.5) na introdução do editor de ontologias Protégé para o leitor. Na seção 3, *A construção do ontoléxico: synsets, frames e o Protégé*, partimos para a sistematização (planejamento e descrição) e implementação (edição e apresentação gráfica) do ontoléxico, passando pelas principais etapas do processo. A seção 4 é uma breve seção de considerações finais, seguida das referências bibliográficas e de três apêndices (uma amostra do levantamento dos itens lexicais do domínio, os principais recursos usados para a estruturação do ontoléxico e um esboço preliminar da ontologia).

1.1 Motivação e justificativa

Uma parte da motivação desta pesquisa advém de estudos e atividades teóricas e práticas de Iniciação Científica¹, principalmente no que diz respeito à representação do léxico em redes *wordnets*, discutidas mais adiante na dissertação. A outra advém do fato de a atividade a partir da qual se recorta o domínio léxico-conceitual que escolhemos para trabalhar, aqui denominado INDÚSTRIA DO BORDADO DE IBITINGA (IBI), ter importância para o artesanato, indústria e cultura da região, e oferecer material lexical rico e, sobretudo, registro linguístico dessa importante atividade do município.

A cidade de Ibitinga e a sua principal atividade, o bordado, já foram objetos de vários estudos, sendo que a maior parte deles aborda, por exemplo, as condições sócio-culturais relativas à indústria (CINTRÃO, 1990; LEITE, 2009) ou o seu sistema de produção, ou seja, analisando-a do ponto de vista econômico (BACIC; SOUZA, 2008; FLORIAN; LORENZO, 2008; MATUSHIMA, 2005); no entanto, não detectamos estudos do ponto de vista linguístico, como o que aqui propomos desenvolver, trazendo para esta pesquisa um caráter de originalidade.

1.2 Contextualização

Contextualizar o domínio IBI faz-se, então, importante para explicitarmos tanto a motivação quanto a justificativa desta investigação.

Lendo um pouco sobre a história de Ibitinga, cidade do interior paulista, com aproximadamente 53 mil habitantes, observamos que ela adquiriu importância graças à atividade do bordado, introduzida, em meados de 1950, pela imigrante portuguesa Dioguina Sampaio. Desde a década de 60, com a formação da “Escola de Bordados Singer”, essa atividade vem se desenvolvendo. A cidade, que desde 1974 abriga a tradicional “Feira do Bordado de Ibitinga”, fica reconhecida como a Capital do Bordado entre as décadas de 80 e 90, período de mudanças para a indústria, que se reestrutura, para acompanhar o cenário nacional e mundial, ampliando e diversificando sua produção.

¹ Projeto PLN-BR, processo CNPq n° 550.388/2005-2 (02/2006 a 11/2007), e PIBIC-CNPq, processos n° 100018/2008-4 (01/2008 a 07/2008) e n° 104037/2008-3 (08/2008 a 02/2009).

De acordo com dados recolhidos por Turco (2007), o setor dos bordados, que engloba mais de 350 empresas formais e 1200 informais, vem representando aproximadamente 80% da economia do município, 12,6% do mercado de São Paulo e 5% do mercado nacional de produtos de cama, mesa e banho.

Em Ibitinga, o bordado – desde aquele confeccionado artesanalmente até o mais tecnológico, com produção em grande escala e fruto de tecnologias desenvolvidas especificamente para o setor – alimenta uma indústria que agrega inúmeros profissionais e utiliza os mais diversos materiais e instrumentos. Por isso, produzi-lo exige conhecimentos técnicos e artísticos que, em termos linguísticos, traduz-se em um universo lexical rico e específico que possibilita a comunicação eficiente entre os profissionais do setor, proporcionando, não só entre esses profissionais, como também entre eles e o público geral, a discursivização de, por exemplo, agentes, técnicas, instrumentos, materiais, suportes, processos e produtos que constituem o universo discursivo dessa indústria.

Esse universo lexical tem uma contrapartida conceitual que, conforme dissemos, pode ser sistematizada em uma ontologia de conceitos. Estudar essa forma de representação de conceitos e dos itens lexicais a eles ancorados, para esta pesquisa, é estudar a constituição e formalização de léxicos e ontologias de um domínio da atividade humana e da combinação de ambos em termos de ontoléxicos, objetos de estudo da subseção 2.3.

1.3 Objetivos

O objetivo deste estudo é sistematizar, do ponto de vista do processamento automático de línguas naturais² (DIAS-DA-SILVA, 1996, 2006), o domínio léxico-conceitual da IBI em termos de uma ontologia do domínio e nela “ancorar” as parcelas correspondentes dos léxicos do português e do inglês, construindo, assim, um “ontoléxico”. Podemos especificar dois grandes objetivos complementares desta pesquisa: (I) a construção de uma ontologia do domínio léxico-conceitual da IBI e (II) a ancoragem, nessa ontologia, dos fragmentos dos léxicos correspondentes do português e do inglês. As atividades em (I) consistem em: (Ia) especificação dos conceitos do domínio e (Ib) descrição desses conceitos em termos de

² A literatura da linguística computacional em português, conforme sinaliza Dias-da-Silva (1996, 2006), traduz “natural language processing” (NLP) de maneiras diferentes, como “processamento computacional de línguas naturais” ou “processamento de linguagem natural”, entre outras. Neste trabalho, como propõe Dias-da-Silva, adotamos o termo “Processamento automático de Línguas Naturais” e a sigla PLN.

glosas³ e dos diferentes tipos de relações (paradigmáticas e sintagmáticas) que se estabelecem entre eles, para, então, propormos a ontologia, nos moldes indicados pela literatura. Já em (II), procedemos à (IIa')⁴ seleção, em *corpus*, dos itens lexicais que compõem os fragmentos dos léxicos da IBI, e também, quando pertinente, da indústria do bordado em geral (IB), de cada uma das duas línguas e que serão estruturados em termos de sua ancoragem à ontologia, isto é, em termos da (IIc) especificação da relação de significação entre o item lexical e o conceito da ontologia por ele lexicalizado, e do (IIId) alinhamento semântico entre os itens lexicais das duas línguas decorrente dessa ancoragem, que resulta, conforme adiantamos, na proposição dos objetivos, em um ontoléxico do domínio, um dos alvos aplicados da pesquisa e com o qual faremos o exercício exploratório de implementação no editor Protégé para demonstrar o seu potencial de gerar automaticamente uma obra de referência léxico-conceitual bilíngue.

³ Glosa é uma definição informal que explicita um conceito; no contexto da WordNet, por exemplo, o synset {*embroidery, fancywork*} tem como glosa “decorative needlework” (trabalho de agulha decorativo).

⁴ Como a atividade Ia' (coleta de itens lexicais) é concomitante à atividade Ia (identificação de conceitos), usamos **a'** para indicar essa simultaneidade.

2 FUNDAMENTOS: *CORPORA*, LÉXICOS, ONTOLOGIAS E EDITOR

Nesta seção, apresentaremos, em 2.1, os variados recursos dos quais nos valem para cumprir parte dos objetivos propostos, sendo que uma atividade para um desses objetivos, a coleta de informações, será especificamente tratada na subseção 2.2. Já em 2.3, estudaremos os léxicos e as ontologias, bem como a interface entre ambos, que nos levará à investigação de um terceiro conceito fundamental: o ontoléxico. Um exercício que exemplifica o alinhamento entre léxicos e ontologias poderá ser conferido na subseção 2.4, e em 2.5 será apresentado o editor de ontologias Protégé-OWL, importante ferramenta para se trabalhar na interface estudada.

Em se tratando de um estudo do ponto de vista do processamento automático de línguas naturais, estruturamos os níveis de investigação conforme metodologia proposta por Dias-da-Silva (1996, 2006): o linguístico, o linguístico-computacional e o computacional, que correspondem, respectivamente, a:

[...] a “extração do solo” (isto é, a explicitação dos conhecimentos e habilidades linguísticas), a “lapidação” (isto é, a representação formal desses conhecimentos e habilidades) e a “incrustação” (isto é, a construção do programa de computador que codifica essa representação). (DIAS-DA-SILVA, 2006, p.122)

Em cada nível do estudo PLN se entrecruzam variadas disciplinas, como a Inteligência Artificial, as Ciências da Computação, a Filosofia da Linguagem, a Linguística, entre outras, cada uma oferecendo os recursos teóricos e metodológicos de sua especialidade. No que diz respeito à construção de ontologias, a interação com os recursos linguísticos se tornou um dos tópicos em evidência no PLN e vem acrescentando novas possibilidades aos produtos desenvolvidos, pois ontologias linguisticamente motivadas, ou ontoléxicos, são o futuro da Web Semântica (HUANG et al., 2010; HIRST, 2004).

Especificamente, fundamentamos nosso estudo na teoria da semântica lexical, pura (CRUSE, 1986; FILLMORE, 1976; JACKENDOFF, 2002; SINGLETON, 2000) e computacional (FELLBAUM, 1998; HANDKE, 1995; IDE; GREENSTEIN; VOSSEN, 1998; VOSSEN, 1998), com auxílio de pesquisas em *corpus* (ALUÍSIO; ALMEIDA, 2006; TEUBERT; ČERMÁKOVÁ, 2007), assim como na metodologia de construção de ontologias

(DING; FOO, 2002; FRANÇA, 2009; GELLER; PERL; LEE, 2004; GRUBER, 1993; HUANG et al., 2010; PINTO; MARTINS, 2004). A pesquisa culmina com a proposição de um ontoléxico do domínio conceitual da IBI (FRANÇA, 2009; HIRST, 2004; HUANG et al., 2010; PRÉVOT; BORGO; OLTRAMARI, 2010), cuja implementação (HORRIDGE, 2004) tem o potencial de gerar produtos como dicionários bilíngues (KRIEGER; FINATTO, 2004).

2.1 As fontes de informação do domínio léxico-conceitual IB(I)

O trabalho empírico de construção do ontoléxico, detalhado na seção 3, concentra-se na coleta de itens lexicais⁵ do português e do inglês que denotam conceitos do domínio léxico-conceitual da IBI, sem descuidarmos, quando relevante para a descrição do ontoléxico desse domínio, da coleta de itens lexicais que denotam os conceitos do domínio léxico-conceitual mais geral da indústria do bordado (IB). No domínio léxico-conceitual da IB temos inseridos, então, todos os conceitos relativos à indústria do bordado, que, por sua vez, não lhe são necessariamente exclusivos e podem ser compartilhados com outros domínios (por exemplo, o conceito TESOURA), conforme o esquema da Figura 1, que, com cores e intersecções diversas, mostra diferentes domínios e sobreposições de domínios conceituais.

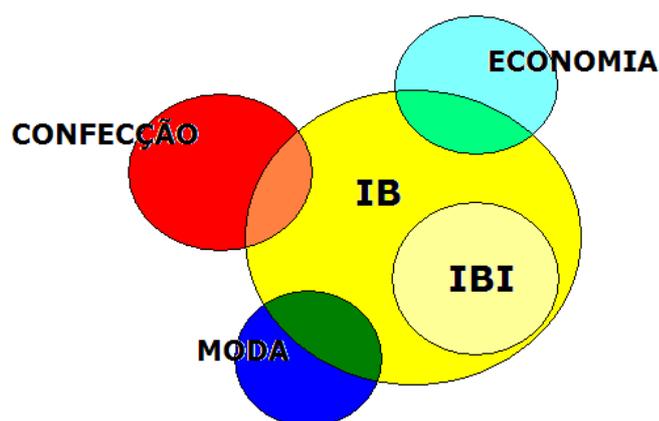


Figura 1. Esquemática do relacionamento de diferentes domínios conceituais.

⁵ Dada a problemática, há muito tempo discutida, sobre a noção de “palavra” (ALLAN, 2001; SINGLETON, 2000), empregamos a noção de “item lexical” que advém de Jackendoff (2002, p.152-153), para quem desde afixos até expressões idiomáticas ou frases inteiras convencionalizadas são considerados itens lexicais; isso porque tais estruturas da língua, armazenadas no “léxico mental” (LM), ou “memória lexical de longo-prazo”, podem ser articuladas (em correspondências com outros constituintes fonológicos, sintáticos e conceituais) por um falante numa construção *online* capaz de expressar partes de seu discurso que não estavam armazenadas no LM. Reservamos, como o fazem os estudiosos da rede *FrameNet*, o termo “unidade lexical” para designar cada sentido de um item lexical.

Essa coleta de itens lexicais e da apreensão dos conceitos de domínio por eles lexicalizados é feita a partir do *corpus* exposto nesta seção. A identificação, análise e sistematização das informações dessas duas naturezas, lexical e conceitual, são essenciais para a proposição do ontológico domínio léxico-conceitual da IBI. Por essa razão, ao longo desta dissertação, descrevemos a base teórico-metodológica e as ferramentas que nos amparam no desenvolvimento dessas três atividades.

O *corpus* da pesquisa, por sua vez, é constituído tanto por recursos não-estruturados (livros, teses, artigos, relatórios e entrevistas) quanto por recursos estruturados (dicionários, enciclopédias, glossários e duas redes estruturadas lexicogramaticalmente, as redes WordNet e FrameNet, que apresentamos na subseção 2.1.5) e embasa-se fundamentalmente no conceito de *corpus* adotado pela Linguística:

Conjunto do material recolhido e bem delimitado no tempo e no espaço apto a servir para a descrição linguística. O *corpus* é uma amostra representativa da língua a descrever, pois de sua análise resultará a descrição, que dele não pode ou não deve fugir. Pode-se obter um *corpus* por diversas maneiras: selecionando informantes e anotando as informações; gravando conversações casuais ou preparadas de um número determinado de falantes; recolhendo obras escritas que representem um determinado registro de língua e baseando a descrição num determinado número delas. (BORBA, 1976, p.25)

Parte do material que utilizamos não está em formato eletrônico⁶ e é, então, trabalhada manualmente, já outra parte coletada originalmente nesse formato ou transcrita para ele pode ser analisada com mais agilidade por meio de ferramentas e procedimentos que descrevemos ao longo desta seção.

2.1.1 Dicionários e enciclopédias

Dicionários e enciclopédias são fontes para a coleta dos itens lexicais e para a consulta de informações de natureza lexical e enciclopédica relacionadas a esses itens.

⁶ Em seu artigo, Aluísio e Almeida (2006, p.156-158) destacam que um dos pontos nos quais a Linguística e a Linguística de *Corpus* se diferenciam é exatamente no conceito de *corpus* que cada uma adota: para a última há a necessidade de os dados estarem em formato eletrônico para que haja processamento por máquina.

Os léxicos das línguas naturais, parcialmente registrados em produtos como dicionários, relacionam-se com o conhecimento enciclopédico registrado também parcialmente em enciclopédias. Ao adotarmos a concepção de léxico e enciclopédia de Allan (2001), consideramos o léxico um acervo de informações sobre os itens lexicais e a enciclopédia um acervo com informações sobre o que os itens lexicais denotam. Nessa concepção, o léxico faz parte da enciclopédia: “A enciclopédia é uma base de conhecimentos gerais da qual o conhecimento lexical é uma parte própria – a informação lexical é um tipo particular de informação enciclopédica.” (ALLAN, 2001, p.101, tradução nossa⁷)⁸.

Assim, ao mesmo tempo em que o léxico é parte da enciclopédia, o conhecimento enciclopédico pode também fazer parte dos dicionários, conforme atestam Prévot et al. (2010): [...] quando a língua é tomada como um artefato social no qual o contexto cultural é um elemento ativo de sua definição, muitas regularidades do conhecimento enciclopédico merecem ser integradas em dicionários. (p.20)⁹.

As obras selecionadas para a extração desses tipos de conhecimentos são, pois, as enciclopédias Mirador (1976), Britannica (1974) e Larousse (1970) e os dicionários Aurélio (FERREIRA, 2004), Dicionário de usos do português (BORBA, 2002), Houaiss (HOUAISS, 2001) e Michaelis (WEISZFLOG, 1998), para o português brasileiro, e Macmillan (2007) e Random House (1996), para o inglês. Como quase todas elas são obras legíveis com máquina¹⁰, o acesso a informações, via itens lexicais, é dinâmico e ainda conta com a pesquisa reversa, que permite a busca de itens lexicais nas definições das entradas dos dicionários.

⁷ Dada a informação de que traduzimos todas as citações diretas de textos originalmente escritos em inglês, dispensamos a notação “tradução nossa” em cada citação.

⁸ “The encyclopedia is a general knowledge base of which lexical knowledge is a proper part – lexical information is just one kind of encyclopedic information.”

⁹ “[...] when language is taken as a social artifact in which cultural context is an active element of its definition, many regularities from encyclopedic knowledge deserve to be integrated in dictionaries.”

¹⁰ WILKS et al. (1993, p.341) esclarecem que esses tipos de dicionários, os legíveis com máquina, contêm conhecimento sobre língua e mundo essenciais para tarefas em PLN, mas, porque foram desenvolvidos por lexicógrafos para leitores humanos, não apresentam esse conhecimento em um formato que possa ser diretamente aproveitado no PLN. Eles enaltecem a necessidade de dicionários tratáveis por máquina, que são dicionários legíveis com máquina transformados, por algum método computacional, em um formato manipulável pelo PLN.

2.1.2 Teses, artigos, relatórios, livros e glossários

Teses, artigos e relatórios que descrevem a realidade do domínio léxico-conceitual da IBI do ponto de vista econômico e/ou do sócio-cultural, conforme citamos na contextualização da subseção 1.2, compõem os recursos para o *corpus* elencados no Quadro 1, juntamente com *folders*, catálogos e outros materiais de divulgação dos produtos comercializados na cidade de Ibitinga. Teses, artigos, relatórios, livros e glossários, dentre outros, compõem recursos, como os destacados no Quadro 2, que abordam a arte e/ou a indústria do bordado em geral, ou seja, no domínio léxico-conceitual da IB.

- *Introdução da prática de medição de desempenho para produção sustentável no cluster industrial de Ibitinga* (ARAUJO et al., 2008)
- *Um estudo sobre os mecanismos de coordenação na determinação de custos e formação de preços no aglomerado de pequenas empresas de Ibitinga* (BASIC; SOUZA, 2008)
- *O trabalho das bordadeiras de Ibitinga: fragmentação e subordinação* (CINTRÃO, 1990)
- *A experiência da FIESP junto ao APL de enxovais de cama, mesa e banho e enxovais para o bebê de Ibitinga* (FIESP, 2006)
- *Território e ambiente institucional: o arranjo produtivo local (APL) “Bordados de Ibitinga-SP”* (FLORIAN; LORENZO, 2008)
- *Aglomerado de pequenas e médias empresas como ambiente propício à melhoria de desempenho – caracterização do arranjo produtivo local do setor de confecção de bordados de Ibitinga* (GEROLAMO et al., 2005)
- *As bordadeiras de Ibitinga: trabalho a domicílio e prática sindical* (LEITE, 2009)
- *Especialização produtiva e aglomeração industrial: uma análise da indústria de confecções de Ibitinga-SP* (MATUSHIMA, 2005)
- *A Experiência no APL de enxovais bordados de cama, mesa e banho de Ibitinga* (TURCO, 2007)
- *Porque acreditar em Ibitinga* (ROSA, 1991)
- *39 artigos do “Jornal da Cidade”¹¹ publicados entre 1999 e 2010*

Quadro 1. Recursos não-estruturados para investigação do domínio léxico-conceitual da IBI.

- *De bonecas, flores e bordados: investigações antropológicas no campo do artesanato em Brasília* (CANANI, 2008)
- *Do arado ao bordado: mudança no trabalho do homem do Sertão* (CHAGAS, 2007)

¹¹ O “Jornal da Cidade”, da cidade de Bauru, vizinha de Ibitinga, é o periódico que mais noticia a cidade de Ibitinga em sua região.

- *Glossário de termos têxteis e afins* (COSTA, 2004)
- *Entrevista com bordadeiras do Morro São Bento de Santos: uma reminiscência dos bordados da Ilha da Madeira* (CRUZ, 2007)
- *Art in Needlework – a book about embroidery* (DAY, 1900)
- *Handbook of Embroidery* (HIGGINS, 1880)
- *Bordado como assinatura: tradição e inovação do artesanato na comunidade de Barateiro – Itapajé/CE* (OLIVEIRA, 2006)
- *Bordados e rendas – cama, mesa e banho: estudos de mercado SEBRAE/ESPM* (SEBRAE, 2008)
- *Bordados Tradicionais Portugueses* (SILVA, 2006)
- *Bordados Tradicionais Portugueses* (VIEIRA, 2002)

Quadro 2. Recursos não-estruturados para investigação do domínio léxico-conceitual da IB.

2.1.3 Entrevistas

As entrevistas com informantes especialistas, além de auxiliarem na contextualização do domínio, são recursos que agilizam a coleta dos itens lexicais, o recorte e a descrição dos conceitos que compõem o léxico e a ontologia do domínio, pois explicitam conhecimentos lexicais e enciclopédicos. Por exemplo, ao perguntarmos a um informante o que é uma *semaninha*, ele, trilhando inconscientemente o percurso semasiológico, nos fornece a glosa, “jogo de sete panos de prato bordados ou pintados, cada um deles representando um dia da semana”, que nos leva ao conceito SEMANINHA¹² lexicalizado pelo item lexical *semaninha*. Se perguntarmos o que é e como se chama a peça bordada na qual ele trabalha, o informante, agora pelo percurso onomasiológico, nos retorna tanto o conceito quanto o item lexical com o qual ele se refere àquela peça em português¹³. Com as entrevistas fica evidenciado o “vai e vem” pelos percursos onomasiológico e semasiológico, sendo que tal movimento é recorrente também na análise de outros recursos do nosso *corpus*, por exemplo, em buscas de conceitos/itens lexicais em dicionários.

¹² A representação dos conceitos é o seu rótulo em caixa alta.

¹³ Babini (2006, p.39-40) retoma Bernard Pottier na explicação dos dois percursos gerativos da enunciação: no percurso onomasiológico o emissor vai da intenção de dizer ao enunciado, já no semasiológico o receptor vai do enunciado a sua interpretação. Por exemplo, se o conceito SEMANINHA lexicalizado por *semaninha* é desconhecido para o leitor desta dissertação, será necessário recorrer à definição proposta no texto, para então, semasiologicamente, apreender o conceito SEMANINHA.

2.1.4 Outros *corpora* estruturados e a *Web como corpus*

O universo digital da Web disponibiliza, para o estudioso da linguagem preparado e atento, um volume considerável e rico de material linguístico, ao conter grande número de textos, das mais diversas naturezas, disponíveis para a análise. Não é sem tempo que se encontram trabalhos que demonstram a relevância de utilizarmos a “*Web como corpus*” (LENCI, 2010)¹⁴. Na Web, nos valem também de *corpora* estruturados acessíveis gratuitamente e com ferramentas de busca inteligentes, como o ‘*Corpus do NILC*’¹⁵ (Núcleo Interinstitucional de Linguística Computacional, da USP de São Carlos), que contém textos brasileiros do registro jornalístico, didático, epistolar e redações de alunos; ou o maior *corpus* de acesso gratuito para o inglês, ‘*The Corpus of Contemporary American English*’¹⁶, ou ainda seu correspondente para o português, o ‘*Corpus do Português*’¹⁷.

São também parte de nosso *corpus* textos coletados de sites desenvolvidos para buscas específicas, dentre os quais citamos o ‘*One Look dictionary search*’¹⁸, que localiza definições em várias referências *on-line* como dicionários e enciclopédias, e o ‘*WebCorp Live*’¹⁹, que permite fazer buscas com filtros, concordância de palavras e/ou em domínios específicos.

Não desconsideramos os sites em geral para a localização (com URL direto ou por motores de busca como o *Google*) de anúncios comerciais, notícias, *e-books* de cursos, etc., tanto os relativos ao domínio léxico-conceitual relativo à cidade de Ibitinga quanto os de outras localidades (no domínio léxico-conceitual da IB).

Lenci (2010) caracteriza bem a importância da utilização desses tipos de recursos em língua natural no desenvolvimento de pesquisas em domínios de conhecimento:

[...] documentos – desde a Wikipedia a artigos científicos e relatórios técnicos – são o repositório primário de conhecimento de uma certa comunidade. Portanto, eles podem ser minerados para identificar os itens de

¹⁴ Embora a Web facilite o acesso e a distribuição de diversos *corpora*, não ignoramos a advertência de Aluísio e Almeida (2007), para quem a web como *corpus* deve ser usada com cautela. As autoras, por outro lado, apontam para a alta relevância dos *corpora* manipuláveis pelo computador, em oposição aos *corpora* compostos por textos impressos.

¹⁵ Disponível em: <<http://lusiadas.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>>. Acesso em: 04 jun. 2011.

¹⁶ Disponível em: <<http://www.americancorpus.org/>>. Acesso em: 04 jun. 2011.

¹⁷ Disponível em: <<http://www.corpusdoportugues.org/>>. Acesso em: 04 jun. 2011.

¹⁸ Disponível em: <<http://www.onelook.com/>>. Acesso em: 04 jun. 2011.

¹⁹ Disponível em: <<http://www.webcorp.org.uk/>>. Acesso em: 04 jun. 2011.

conhecimento mais relevantes para caracterizar um domínio particular, e usá-los para alimentar o processo de criação da ontologia. (p.249)²⁰.

Dentre os recursos do *corpus*, demos relevo ao uso de teses, artigos, relatórios e livros para a extração de conceitos e de itens lexicais do domínio léxico-conceitual da IBI, já que os demais recursos são mais apropriados para extração de conceitos e itens lexicais domínio léxico-conceitual IB. Conforme argumenta Sardinha (2000, p. 348):

[...] a quantidade de textos de uma variedade ou domínio específicos nos *corpora* gerais é pequena. Assim, normalmente, *corpora* compilados em pequena escala por pesquisadores individuais acabam sendo mais representativos do que os respectivos sub-*corpora* dos *corpora* gerais.

Teubert e Čermáková (2007) expõem os limites enfrentados pela Lexicografia para representar o vocabulário de uma comunidade de discurso:

Não podemos esperar tornar o discurso em um todo acessível para nossas investigações lexicográficas, mas podemos compilar corpora maiores e maiores e podemos também usar a crescente Internet como um *corpus* virtual. No entanto, como novas palavras e construções são cunhadas dia a dia, é conceitualmente impossível propor um *corpus* que compreenda todo o vocabulário de uma comunidade de discurso. (p. 46)²¹.

Essa limitação, talvez impossível de ser superada, dada a própria natureza das línguas naturais, se impõe para as pesquisas, desde aquelas que focam um domínio pontual, como no nosso caso, até aquelas que visam a descrever toda uma língua ou a estudar aspectos específicos dela (sozinha ou em comparação com outras línguas). Por isso, a constituição do *corpus* com o qual trabalhamos é diversificada.

²⁰ “[...] documents – from Wikipedia to scientific papers and technical reports – are the primary repository of the knowledge of a certain community. Therefore, they can be mined to identify the knowledge items most relevant to characterize a particular domain, and use them to feed the ontology creation process.”

²¹ “We cannot hope to make the discourse as a whole accessible to our lexicographic enquiries, but we can compile larger and larger corpora, and we can also use the ever-growing Internet as a virtual corpus. Nevertheless, as new words and phrases are coined day by day, it is conceptually impossible to come up with a corpus that comprises the whole vocabulary of a discourse community.”

2.1.5 As redes *wordnets* e *framenets*

Conforme adiantamos, fazem parte dos recursos lexicogramaticais estruturados do nosso *corpus* duas redes semântico-conceituais decisivas para a montagem do ontológico em construção: as redes WordNet de Princeton e FrameNet de Berkely.

2.1.5.1 A rede *WordNet*

O constructo linguístico-computacional WordNet (FELLBAUM, 1998), doravante rede WN.Pr, desenvolvido por George Miller e sua equipe, entre as décadas de 1980 e 90, na Universidade de Princeton, é uma rede que estrutura os conceitos expressos no léxico do inglês, sendo que esses conceitos lexicalizados organizam-se em termos de *synsets* (*synonym sets* = conjuntos de sinônimos). Um *synset*, fundamentando-se na sinonímia contextualmente motivada, reúne itens lexicais como *embroidery* e *fancywork*, porque podem ser usados, em um dado contexto, para expressar um mesmo conceito²².

Além da sinonímia, que agrupa os itens lexicais em *synsets*, a rede WN.Pr abriga outros três tipos de relações: a antonímia (oposição de sentidos), a hiponímia/hiperonímia (subordinação/superordenação) e a meronímia/holonímia (parte-todo), que relacionam os *synsets* (isto é, os conceitos lexicalizados)²³.

Desse modo, na constituição da rede, cada *synset* é um nó e cada relação que ele estabelece com outros *synsets* é um arco. O Quadro 3 exemplifica as relações que estruturam uma rede como a rede WN.Pr.

Além dos *synsets* formados por substantivos, a rede WN.Pr também contém *synsets* formados por verbos, para os quais prevê as relações semânticas da troponímia e acarretamento²⁴, e adjetivos, para os quais prevê a relação de antonímia, também prevista para os *synsets* de substantivos. Marginalmente, contém também *synsets* de advérbios.

²² Entendido o fato de que a sinonímia exata é rara em línguas naturais, para a rede WN.Pr, são considerados sinônimos os itens lexicais que são intercambiáveis em um dado contexto.

²³ A rede WN.Pr estrutura-se, então, em termos de relações lexicais (entre os itens lexicais sinônimos que compõem os *synsets*) e relações conceituais (entre os conceitos da rede, lexicalizados nos *synsets*).

²⁴ Troponímia é um termo cunhado pelos desenvolvedores da WN.Pr para denotar a relação de hiponímia entre *synsets* de verbos. Por exemplo: {*embroider*, *broider*} tem como tropônimo {*purl*} (bordar com linha de ouro ou prata), ou seja, este codifica um modo particular de executar a ação codificada naquele; já a relação (unilateral)

SYNSETS			
(a) { <i>tambour1, embroidery frame, embroidery hoop</i> } (b) { <i>framework, frame2, framing</i> } (c) { <i>brace, bracing</i> }			
RELAÇÕES LÉXICO-CONCEITUAIS			
hiperonímia / hiponímia		meronímia / holonímia	
(b) é hiperônimo de (a)	(a) é hipônimo de (b)	(b) tem (c) como parte	(c) é parte de (b)

Quadro 3. Estruturação léxico-conceitual na rede WN.Pr.

Fellbaum (1998) reconhece que “[...] não existem relações semânticas diferentes suficientes na WordNet.” (p.36)²⁵, já que seus desenvolvedores limitaram-se às relações que acreditavam ser de mais ampla aplicabilidade e de maior familiaridade; em decorrência disso, ocorre o que se denomina *the tennis problem* (FELLBAUM, 1998²⁶; MARCELLINO; DIAS-DA-SILVA, 2009), ou seja, “o problema do jogo de tênis”. Trata-se da falta de conexão entre itens lexicais de um mesmo domínio conceitual dentro da rede WN.Pr: não há, por exemplo, em uma wordnet, uma relação que ligue o synset {*darning needle, embroidery needle*} ao synset {*thread, yarn*}.

Seguindo a metodologia de montagem da rede WN.Pr para a descrição do léxico do inglês norte-americano, outros projetos foram propostos para o desenvolvimento de outras wordnets e de wordnets multilíngues, como a rede EuroWordNet (VOSSSEN, 1998), uma “multiwordnet” que alinha semanticamente as wordnets em construção para as línguas da União Europeia. A rede WordNet.Br, doravante rede WN.Br (DIAS-DA-SILVA, 2007, 2006, 2004), motivadora de estudos e produções de nossa Iniciação Científica (MARCELLINO, 2008; MARCELLINO; DIAS-DA-SILVA, 2008; RODRIGUES; MARCELLINO; DIAS DA SILVA, 2008) e desta pesquisa, é uma iniciativa, em andamento, de construção de uma rede wordnet similar para o português brasileiro.

Para facilitar a construção dessas wordnets internacionais e o mapeamento entre elas, Vossen (1998) propôs o Índice-Interlíngue (*Interlingual-Index*, ILI) que engloba todos os synsets da rede WN.Pr versão 1.5 e permite que as wordnets em construção para as outras

de acarretamento entre synsets de verbos pode ser exemplificada pelos synsets {*dream*} e {*sleep, kip, slumber, log Z's, catch some Z's*}, em que o primeiro acarreta o segundo.

²⁵ “[...] there are not enough different semantic relations in WordNet.”

²⁶ Fellbaum (1998, p.34) ilustra esse problema explicando que o vocabulário especializado do tênis está presente em toda parte e em parte alguma da rede WN.Pr: os equipamentos do tênis localizam-se no arquivo noun.artifact, os jogadores em noun.person, a quadra em noun.location, etc; os substantivos que co-ocorrem em discussões sobre tênis estão, então, espalhados pela rede sem que nada os relacione.

línguas conectem os seus synsets aos synsets correspondentes à base ILI, o que torna o ILI um agrupamento de todos os conceitos de todas as wordnets cujo mapeamento de synsets é muito útil para várias aplicações no cruzamento de línguas (PEASE; FELLBAUM, 2010, p.34).

Vossen (em VOSSSEN, 1998 e IDE; GREENSTEIN; VOSSSEN, 1998) ainda define a metodologia para o mapeamento entre as wordnets, parte importante dela, da qual nos valemos nesta pesquisa, corresponde à especificação do alinhamento semântico (ou co-indexação) entre synsets. Por exemplo, o synset $\{\textit{dirigir, guiar, pilotar}\}$ da rede WN.Br é alinhado a $\{\textit{drive}\}$ (wordnet do inglês), a $\{\textit{guidare}\}$ (wordnet do italiano) e a todos os outros synsets de redes diferentes que se alinharem ao mesmo ILI $\{\textit{drive}\}$.

A notação *EQ_SYNONYM* é utilizada para indicar a relação direta entre synsets sinônimos de wordnets diferentes, dessa maneira, no alinhamento entre a rede WN.Br e a rede WN.Pr temos, por exemplo, $\{\textit{bordador}\}$ *EQ_SYNONYM* $\{\textit{embroiderer}\}$, conforme ilustrado na Figura 2.

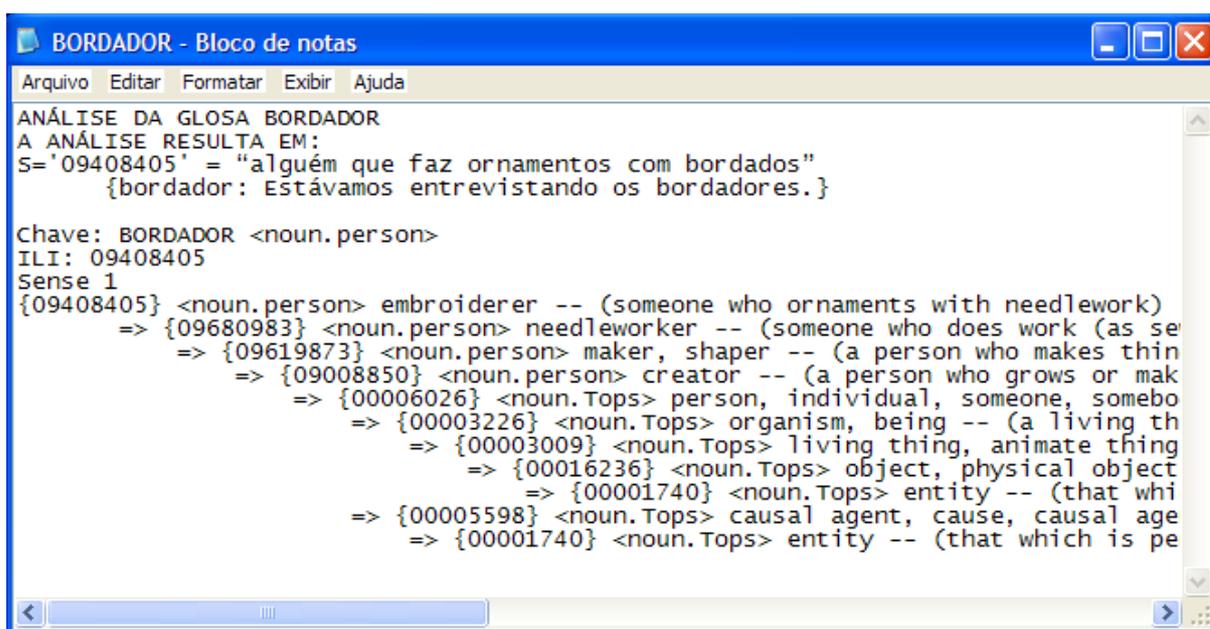


Figura 2. Synset $\{\textit{bordador}\}$ construído para o português, e o seu alinhamento com o synset $\{\textit{embroiderer}\}$.

Observamos, na Figura 2, que a construção²⁷ do synset envolve a representação de várias informações: o próprio item lexical (nesse caso apenas um, *bordador*, mas poderia haver outros se fossem identificados sinônimos para ele), a glosa “alguém que faz ornamentos

²⁷ Esse synset, assim como os outros em desenvolvimento para a base brasileira e, no caso desta pesquisa, para o ontológico do domínio léxico-conceitual da IB(I), foi construído no arquivo padrão que segue as especificações necessárias para que seja possível sua implementação no editor que viabiliza o alinhamento entre a rede WN.Br e a rede WN.Pr através do ILI.

com bordados” e a frase-exemplo de seu emprego, extraída de *corpus*, “Estávamos entrevistando os bordadores”. Além disso, o synset recebe²⁸ o mesmo ILI do synset da rede WN.Pr (indicado pela numeração 09408405). Desse modo, os synsets, que então representam um mesmo conceito, estão alinhados pela relação de sinonímia.

A co-indexação das bases, por sua vez, permite à rede WN.Br herdar o tipo semântico²⁹ do synset (uma informação importante para sua categorização) e a estrutura hierárquica da rede WN.Pr, pela Figura 2, notamos que é possível conhecer, por exemplo, o synset hiperônimo de {*embroiderer*}, o {*needleworker*}, sob o qual estão representados, por meio de synsets, outros conceitos além de o de BORDADOR; ter acesso a esses outros conceitos, bem como aos itens lexicais que os lexicalizam, já nos permite enriquecer a base da rede WN.Br e, sobretudo, construir o ontolético do domínio léxico-conceitual da IBI, conforme veremos ao longo da seção 3.

No entanto, podemos nos deparar com a impossibilidade de fazer alinhamento direto entre synsets, ou qualquer outro tipo de alinhamento, o que reflete lacunas nas estruturas das wordnets ou lacunas nas línguas, o que é comum. Diferentes notações foram definidas para os casos em que o alinhamento entre um synset X (de uma wordnet) e o synset Y (de outra wordnet) é indireto:

- X é mais específico que Y $\rightarrow X EQ_HAS_HYPERONYM Y$;
- X é mais genérico que Y $\rightarrow X EQ_HAS_HYPONYM Y$;
- X associa-se a Y e a Y' $\rightarrow X EQ_NEAR_SYNONYM Y$;
- X e X' associam-se a Y $\rightarrow X EQ_NEAR_SYNONYM Y$.

A importância de redes com organização paradigmática conceitual, como as wordnets, se reflete nos diversos trabalhos de PLN que utilizam a rede WN.Pr de várias maneiras, inclusive aproveitando a ontologia que lhe subjaz: “A ontologia implícita nas hierarquias dos substantivos têm recebido especial atenção dos linguistas computacionais” (FELLBAUM,

²⁸ Rodrigues, Marcellino e Dias-da-Silva (2008) descrevem parte dos procedimentos metodológicos para o alinhamento de synsets de substantivos, que envolve, basicamente, a análise dos synsets do português e do inglês fundamentada nas teorias de investigação léxico-semântica e nas teorias de construção e co-indexação de wordnets, contando também com o auxílio de recursos como dicionários e *corpora*.

²⁹ Os tipos semânticos propostos na base da rede WN.Pr são: noun.Tops, noun.act, noun.animal, noun.artifact, noun.attribute, noun.body, noun.cognition, noun.communication, noun.event, noun.feeling, noun.food, noun.group, noun.location, noun.motive, noun.object, noun.person, noun.phenomenon, noun.plant, noun.possession, noun.process, noun.quantity, noun.relation, noun.shape, noun.state, noun.substance e noun.time. Esses tipos semânticos acompanham os synsets, agrupando-os nessas grandes categorias semânticas, e derivam do desenvolvimento dos *unique beginners* que embasaram a WN.Pr.

1998, p.44)³⁰. Sobre as características da rede, Prévot et al. (2010) resumem que a “[...] WordNet é, em princípio, um léxico, já que todas as suas entradas são expressões linguísticas, mas a estrutura semântica definida pelos synsets e por suas relações tem sido frequentemente utilizada como uma ontologia linguística [...]” (p.11)³¹.

2.1.5.2 A rede *FrameNet*

A rede FrameNet (FILLMORE; JOHNSON; PETRUCK, 2003), por sua vez, é uma implementação computacional da semântica de frames (FILLMORE, 1976), em que os *frames*, que representam situações/cenas/domínios, são estruturados, a partir da análise de frases e textos escritos em inglês norte-americano, em termos de relações temáticas dentro do frame (*Role, Means, Place*, dentre outras) e de relações entre frames (*Inherits From, Is Used By, Has Subframes*, dentre outras). As informações de uma rede FrameNet, por estarem sintagmaticamente estruturadas, de certa forma, complementam as informações de uma rede WordNet.

Uma rede FrameNet inclui também unidades lexicais (*lexical unities*, LUs), sendo que cada LU evoca um frame com o seu conjunto de elementos, os elementos do frame (*frame elements*, FEs)³². Por exemplo, o frame *Physical_artworks* contém FEs, que são as relações temáticas, como *Artifact*, considerado o elemento nuclear do frame (*Core*), *Creator, Material, Time_of_creation, Type*, dentre outros, e está associado a LUs como *drawing.n, painting.n, photograph.n, sculpture.n., etc.*

Em linhas muito gerais, um sistema de frames pode ser concebido como uma representação conceitual linguisticamente motivada de um domínio experienciado. No nosso caso, o domínio léxico-conceitual da IBI. Os frames, nas palavras de Allan (2001), “[...] identificam os traços, atributos e funções características de um objeto do mundo, e a suas interações características com coisas típica ou necessariamente associadas a ele.” (p.251)³³.

³⁰ “The ontology implicit in the noun hierarquies has received special attention from computational linguists.”

³¹ “[...] WordNet is primarily a lexicon since all its entries are linguistic expressions, but semantic structure defined by the synsets and their relations have frequently been used as a linguistic ontology [...]”

³² É importante destacar que, na rede FrameNet, a unidade lexical, por definição, é uma unidade léxica não ambígua.

³³ “Frames identify the characteristic features, attributes, and functions of a denotatum, and its characteristic interactions with things necessarily or typically associated with it.”

Assim, neste trabalho, prevemos que a representação da valência semântica dos itens lexicais que devem integrar o ontológico domínio léxico-conceitual da IBI seja elaborada em termos de uma rede FrameNet, já que, por exemplo, o item lexical *bordado* pode ser entendido como uma LU do frame *Physical_artworks*, posto que os FEs *Artifact*, *Creator* e *Material*, dentre outros, são instanciados pelas LUs *bordado*, *bordadeira* e *linha*, respectivamente, da mesma forma, como exemplificam os autores, que as LUs, *drawing*, *Garofalo* e *marble*.

2.2 A estratégia de coleta das informações léxico-conceituais

O *corpus* da pesquisa, descrito na subseção 2.1, fornece informações de várias naturezas via língua natural, por meio da qual temos acesso a duas partes dessas informações que são objetos de busca e análise aqui: a identificação de conceitos (cf. Ia) e a coleta das parcelas dos léxicos do português e do inglês que se relacionam com os conceitos (cf. IIa' e IIc). Nesta subseção, é principalmente da coleta de itens lexicais, que então nos dá acesso a informações de natureza lexical e conceitual, que tratamos, com foco nas ferramentas e nos procedimentos que nos auxiliam nessa tarefa. O alinhamento entre wordnets também é abordado para mostrarmos como outras tarefas são executadas a fim de coletarmos informações particulares a esses tipos de redes e igualmente importantes para o desenvolvimento do ontológico do domínio léxico-conceitual da IBI.

A coleta parcial dos itens lexicais nos recursos desta pesquisa se dá majoritariamente de modo manual, conforme é verificado no Apêndice 1, mas ela pode ser facilitada às vezes:

- por meio de um ambiente específico para armazenamento de sites, o ‘*Offline Explorer Enterprise*’³⁴ (OEE), que possibilita baixar e armazenar sites, inteiros ou partes, para consultas posteriores. Perante a dinamicidade da Web é importante ter um objeto fixo para se trabalhar, e com o OEE o trabalho consiste em incrementar os recursos da pesquisa com os textos de sites escolhidos, dos quais coletamos os itens lexicais. Através das ferramentas específicas do OEE conseguimos atualizar, no programa, o site trabalhado (opção: “Explorar a URL atual Online”), carregando

³⁴ *Offline Explorer Enterprise 5.8.3158 Release*: Disponível em: <<http://www.metaproducts.com/>>. Acesso em: 04 jun. 2011.

eventuais dados, como imagens que tenham sido inicialmente dispensadas, e também fazer buscas através de palavras-chave (opção: “Buscar conteúdo”);

- pelas opções de busca em sites, conforme apresentamos os *corpora* como recursos para a pesquisa, ou nos próprios documentos e programas. Por exemplo, o visualizador de arquivos *Adobe Reader*®³⁵ permite buscas avançadas no próprio documento ou em alguma pasta específica armazenada no computador.

A Figura 3 mostra o ambiente OEE. A Figura 4 exemplifica um tipo de busca nesse ambiente, em que a chave “porta” retorna itens lexicais como *porta condimentos*, *porta azeite*, *porta galão*, *porta assadeira*, *porta papel toalha* e *porta pão*. A Figura 5 exemplifica um exercício de pesquisa, usando o item lexical *bordado*, com o *Adobe Reader*®.

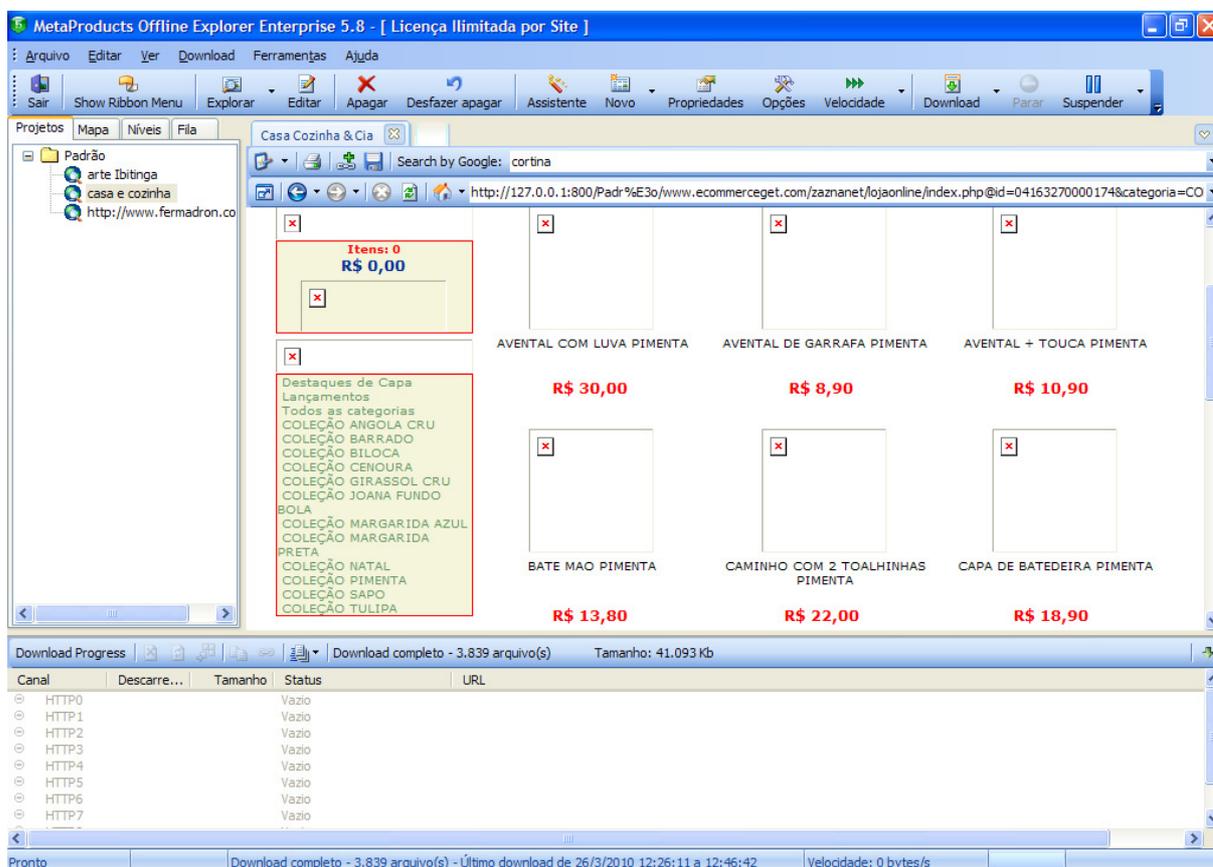


Figura 3. O ambiente OEE (*Offline Explorer Enterprise*).

³⁵ Visualizador de documentos eletrônicos disponível para download gratuito em <<http://www.adobe.com/br/products/reader/>> Acesso em: 04 jun. 2011.

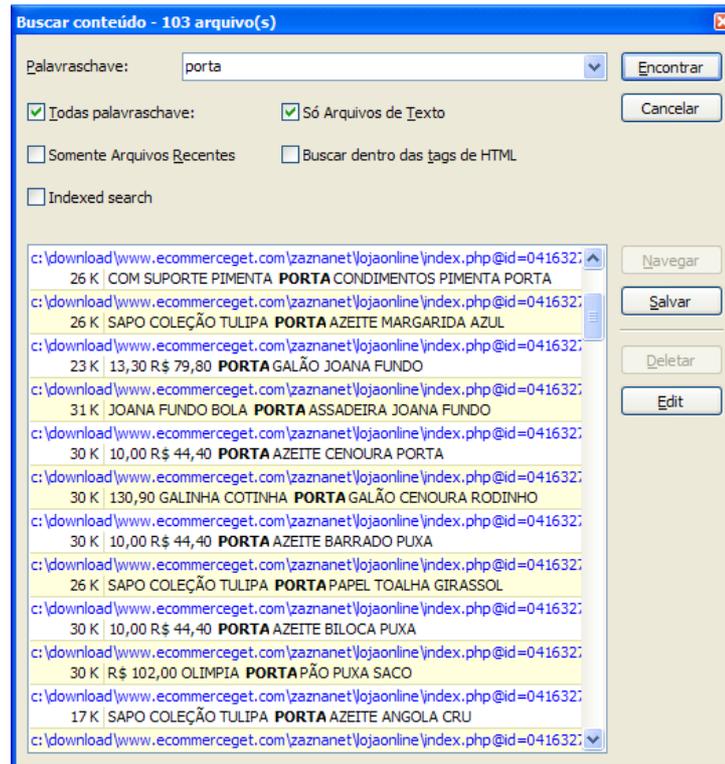


Figura 4. Resultado da busca com a palavra-chave “porta” no ambiente OEE.

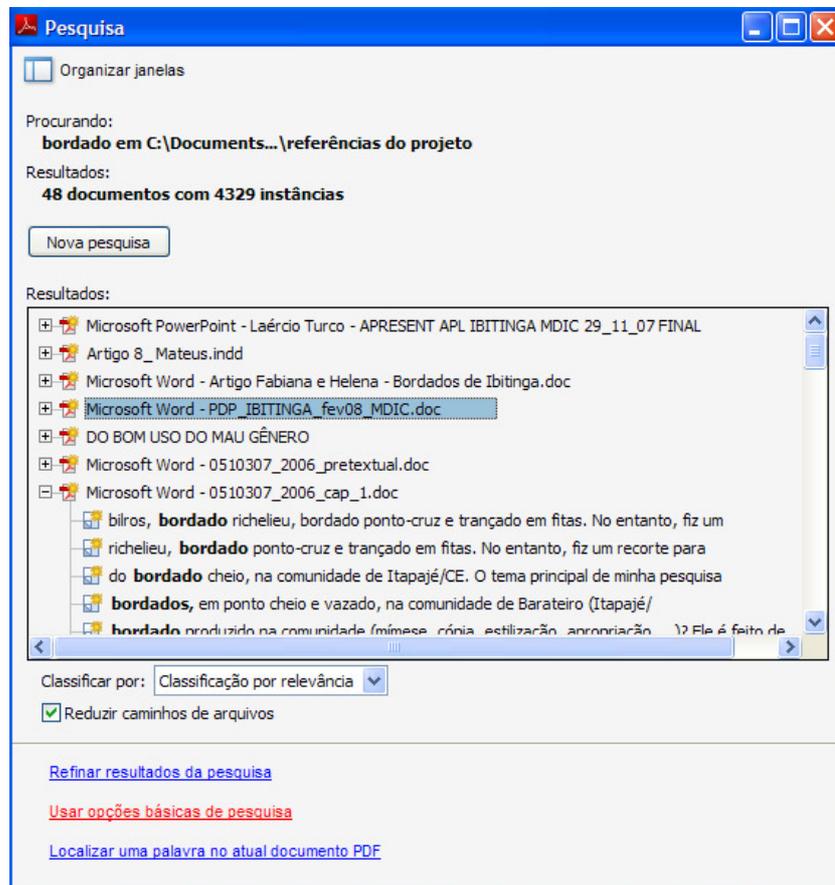


Figura 5. Exemplo de busca com o Adobe Reader©.

Conforme já adiantamos, a maioria dos dicionários que compõem nossos recursos para a pesquisa é do tipo legível com máquina, o que nos permite contar com a funcionalidade da pesquisa reversa. Por exemplo, digitando a palavra-chave “bordado”, coletamos 71 palavras-entrada do Houaiss (HOUAISS, 2001) contendo essa palavra-chave em suas definições, conforme mostra parcialmente a Figura 6.

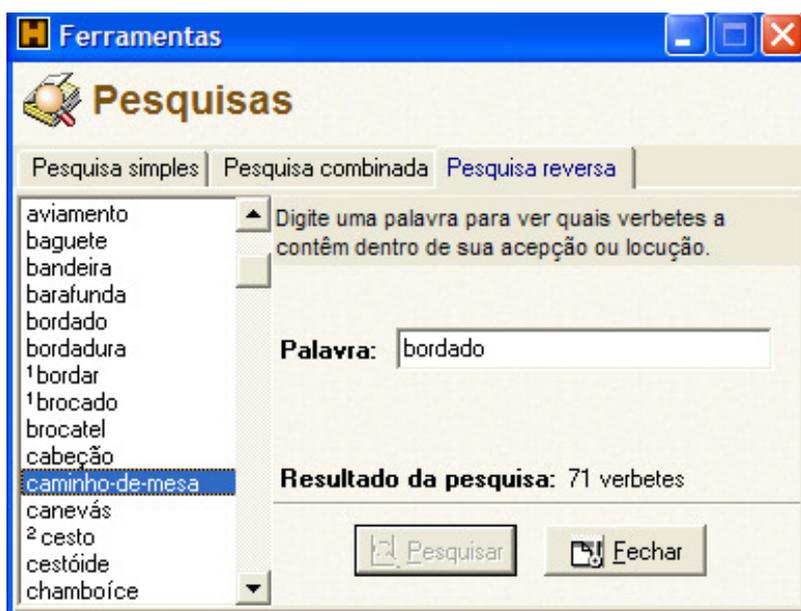


Figura 6. Resultado da pesquisa reversa com a palavra-chave “bordado” no dicionário Houaiss (2001).

Quanto ao trabalho com wordnets, se tomarmos a rede WN.Pr como uma ontologia³⁶ (possibilidade explicada há pouco na subseção 2.1.5.1), teremos uma base preliminar motivando a construção da ontologia o domínio léxico-conceitual da IBI para ancorar os fragmentos de léxico desse domínio. Partindo, então, de um conceito coletado nas referências da pesquisa ou na própria WN.Pr, damos início às atividades que levarão à ancoragem e desenvolvimento do ontológico o domínio léxico-conceitual da IBI:

1. uma vez determinado o conceito da ontologia (satisfeito o objetivo Ia);
2. procedemos à coleta do item lexical (ou itens) no português que o lexicaliza, caso o conceito já não tenha sido determinado a partir do item lexical, mas em qualquer situação, partindo do conceito ou partindo do léxico, cumprem-se IIa' e IIc;
3. com as informações conceituais e lexicais já levantadas, o synset pode ser construído, frases-exemplos podem ser extraídas dos recursos e uma glosa pode

³⁶ Podemos também tomar a coleção de conceitos mais ampla, os ILIs, usados no alinhamento entre wordnets.

ser proposta, o que corresponde a parte do objetivo Ib (descrição do conceito em termos de uma glosa);

4. o synset pode, então, ser alinhado ao ILI, que corresponde ao seu conceito e ao qual também foi alinhado um synset da rede WN.Pr, o que promove a sua conexão direta (logo, IId é satisfeito) e indireta com os synsets da rede (herança de relações semânticas)³⁷. Logo, a outra parte de Ib, a descrição do conceito em termos da definição de suas relações, se cumpre.

Já com a entrevista, com “o vai e vem” pelos percursos onomasiológico e semasiológico, que ilustramos na subseção 2.1.3, o informante nos possibilita, simultaneamente, coletar itens lexicais (cf. IIa’) e identificar conceitos (cf. Ia), o que, por sua vez, nos fornece informações essenciais para a categorização conceitual dos itens lexicais (cf. IIc), um passo fundamental para a proposição do ontolêxico que buscamos, conforme afirma Lenci (2010, p.243):

O objetivo primário dos recursos ontolêxicos é caracterizar os tipos semânticos das expressões linguísticas, ou seja, as classes às quais as expressões linguísticas pertencem em virtude de seus significados. Os tipos semânticos podem ser tomados como modos formais e simbólicos de identificar os conceitos expressos pelas expressões linguísticas. Na medida em que significados são relacionados a entidades no mundo, os tipos semânticos correspondem também a categorias de entidades referidas por itens linguísticos.³⁸

Esse outro papel do informante, na categorização de conceitos, é, então, muito relevante para a pesquisa: seu auxílio de especialista permite a identificação e o refinamento das categorias³⁹ ontológicas do domínio léxico-conceitual da IBI. Até o momento propomos algumas categorias para a ontologia, de acordo com o exposto na subseção 3.3.

³⁷ Retomando o exemplo de construção e alinhamento de synset da Figura 2 (p.23): a conexão direta é ilustrada pelo alinhamento {bordador} EQ_SYNONYM {embroiderer}, que proporciona também a conexão indireta, por meio da herança de relações, do synset {bordador} com outros synsets da hierarquia, como o seu hiperônimo {needleworker} ou o seu synset coordenado {sewer}, ou seja, o seu synset “irmão” na hierarquia.

³⁸ “The primary aim of *ontolexical* resources is to characterize the *semantic types* of linguistic expressions, i.e. the classes to which linguistic expressions belong in virtue of their meanings. Types can be regarded as formal, symbolic ways of identifying the concepts expressed by linguistic expressions. To the extent that meanings are related to entities in the world, semantic types also correspond to the categories of entities referred to by linguistic items.”

³⁹ Categoria é aqui entendida como conjunto de elementos que compartilham uma ou mais propriedades.

2.3 Os conceitos-chave: léxico, ontologia, e ontoléxico

Conforme já explicamos, constituirão o ontoléxico lexicalizações dos conceitos dos domínios léxico-conceituais tanto da IBI quanto da IB, sendo que ambos compartilham conceitos com outros domínios, como esquematiza a Figura 1 (p.14). É importante, para situar a pesquisa no cenário teórico dos estudos do léxico, expormos a natureza dessas lexicalizações, que revestem os conceitos com unidades lexicais que, dependendo do cenário comunicativo, ora são chamadas de itens lexicais ora de termos.

Dois estudos em especial precisam ser levados em consideração na abordagem desta questão: a Lexicologia e a Terminologia, que

[...] embora aproximem-se, porquanto ambas constituem ciências do léxico, distinguem-se pela especificidade de seus objetos. A diferença entre estes, cabe ressaltar, não é outra senão a propriedade que possuem as unidades lexicais chamadas de termos de estruturas linguísticas que, em sua dualidade sígnica, denominam e circunscrevem cognitivamente objetos, processos e conceituações pertinentes ao universo das ciências, das técnicas e das tecnologias; enquanto as palavras, realizando o mesmo processo denominativo e conceitual, cobrem toda a abrangência da realidade cognitiva e referencial apreendida e construída pelo homem. (KRIEGER; FINATTO, 2004, p.43).

Desse modo, as conceitualizações humanas em sua totalidade manifestam-se na língua por meio de itens lexicais (ou palavras, para as autoras), sendo que partes específicas dessa totalidade são representadas por termos.

Eugen Wüster, difusor das primeiras ideias da Terminologia moderna, que deram origem à Teoria Geral da Terminologia, considerava a Terminologia como um ramo da Linguística Aplicada, diferenciando bem, no entanto, a Linguística da Terminologia, considerando que cabe à primeira a investigação de todos os aspectos da língua geral e que a segunda deve se ocupar do léxico especializado, em que figuram termos com significados específicos da área (monorreferencialidade). Por outro lado, novos estudos terminológicos, incitados por linguistas, começaram a criticar a Teoria Geral da Terminologia e seu ideal normalizador, questionando, por exemplo, as fronteiras traçadas entre item lexical e termo, entre língua comum e língua de especialidade. Dentre as novas investigações terminológicas destaca-se a proposta de Maria Teresa Cabré (CABRÉ, 1999), a Teoria Comunicativa da Terminologia. Essa teoria:

[...] articula-se baseada na valorização dos aspectos comunicativos das linguagens especializadas, bem como na compreensão de que as unidades terminológicas formam parte da linguagem natural e da gramática das línguas. O conteúdo de um termo não é fixo, mas relativo, variando conforme o cenário comunicativo em que se inscreve. Não há termos, nem palavras, mas somente unidades lexicais, tendo em vista que estas adquirem estatuto terminológico no âmbito das comunicações especializadas. (KRIEGER; FINATTO, 2004, p.35)

Com essa visão da Terminologia de caráter mais linguístico, explica Finatto (2004), o objeto texto precede o objeto termo, em um estudo que foca mais o texto que tenha termos e menos o estudo de termos em textos, afinal, os termos “[...] são as unidades semânticas dominantes ou mais salientes nos textos técnico-científicos. Não obstante, o termo é uma unidade lexical e sua aceção estará definida no texto-fonte, integrando a totalidade da tessitura textual.” (FINATTO, 2004, p.351-352).

Priorizando, então, a realização em língua, o estudo terminológico de base comunicativa afirma que “[...] os termos são itens lexicais que não se distinguem da palavra do ponto de vista de seu funcionamento.” (KRIEGER; FINATTO, 2004, p.78), sendo assim, um termo “[...] é elemento da linguagem em funcionamento, dada a sua presença em textos e discursos especializados [...]” (KRIEGER; FINATTO, 2004, p.79).

Krieger e Finatto ainda afirmam que “[...] palavra e termo obedecem aos mesmos padrões e sofrem os mesmos efeitos da gramática dos sistemas linguísticos.” (2004, p.46) e que “[...] tal como as outras unidades lexicais dos sistemas linguísticos, as especializadas sofrem processos de sinonímia e comportam variações das mais diferentes naturezas.” (2004, p.79); tais fatores, desconsiderados pela Teoria Geral da Terminologia, que não levava em conta a linguagem em seu funcionamento, são verificáveis em constructos como os *synsets*. Por exemplo, os *synsets* da WN.Pr {*sunflower, helianthus*}, {*cameraman, camera operator, cinematographer*} ou {*cancer, malignant neoplastic disease*} abarcam itens lexicais e termos, assim como o ontoléxico para o domínio IBI pretende fazer (por exemplo, em {*máquina de costurar, maquininha*}), já que não há o comprometimento teórico e metodológico estrito com a Lexicografia e nem com a Terminografia, estudos que abordamos a seguir.

A Lexicografia, comumente entendida como a arte (técnica/ciência) de fazer dicionários, explicam Krieger e Finatto (2004, p.47-49), é uma atividade milenar que cada vez mais vem focando e aprimorando o aspecto aplicado do fazer lexicográfico, que é a produção de diversificadas obras dicionarísticas para o léxico geral, e o aspecto teórico, que visa a uma metodologia científica, desse fazer. A Terminografia, ao ser também denominada

Lexicografia Especializada, já indica ser correlata à Lexicografia, mas, ao invés de tomar a palavra como objeto de descrição e aplicação, toma o termo, que vai figurar, então, em glossários, dicionários técnicos ou terminológicos e bancos de dados (KRIEGER; FINATTO, 2004, p. 50-53).

Os variados produtos dessas duas ciências do léxico admitem certas categorizações. Por exemplo, Babini (2006, p.38-39) menciona os repertórios semasiológicos, que organizam itens lexicais em função da forma, e os repertórios onomasiológicos, que organizam itens lexicais em função dos conceitos que representam. Desse modo:

O problema que um dicionário onomasiológico deve resolver é exatamente o inverso daquele de um dicionário semasiológico: dada uma ideia (noção ou conceito), deve-se encontrar a unidade lexical ou o termo que a exprima. Em um dicionário semasiológico, o ponto de partida é o significante de um termo ou palavra; em um dicionário onomasiológico o ponto de partida é o significado. (BABINI, 2006, p. 39).

Desta forma, a Lexicografia, com seu método semasiológico, tem como produto o dicionário de língua geral⁴⁰, e a Terminografia, com seu método onomasiológico, tem o dicionário terminológico. Krieger nos lembra que “[...] para cumprir o programa de repertoriar as unidades lexicais, definir-lhes o significado e configurá-las descritivamente, a prática lexicográfica necessita fazer frente à heterogeneidade constitutiva do léxico.” (2006, p. 144). Essa heterogeneidade acaba refletida nas obras. Por exemplo, é comum verificarmos que um dicionário de língua geral contém mais termos de áreas específicas do que era de se esperar, já que, explicam Krieger e Finatto (2004, p. 131-132), em tese, é inserido em um dicionário comum apenas o vocabulário técnico de base de grandes áreas especializadas. No entanto, estimativas indicam que, em média, a metade do volume registrado nesses dicionários é constituída de léxico especializado, o que denuncia a dificuldade em se traçar fronteiras entre a língua comum e a língua de especialidade.

Uma vez que “não há termos, nem palavras, mas somente unidades lexicais”, conforme já citamos em referência de Krieger e Finatto (2004, p.35), o nosso trabalho com o domínio léxico-conceitual da IBI envolve, pois, elementos denominados “palavras” e outros

⁴⁰ Dicionário de língua geral: “[...] denominação muitas vezes aplicada em razão da obra cobrir a totalidade das realizações léxicas de um idioma. Neste caso, a totalidade não significa o registro exaustivo das palavras de uma língua, desde as mais antigas até os neologismos mais recentes, até porque a lexicografia não consegue acompanhar o dinamismo lexical; mas caracteriza a abrangência do componente léxico, sem privilegiar uma temática específica. A especificidade está nos dicionários que têm por objeto um subconjunto léxico, por exemplo, a terminologia da química, da informática ou do meio ambiente.” (KRIEGER, 2006, p.143-144)

denominados “termos”, todos, na verdade, lexicalizando as peças-chave da ontologia que sustenta o nosso ontoléxico: os conceitos. Embora desenvolvamos um trabalho na área dos estudos do léxico e possamos utilizar partes do arcabouço teórico e metodológico desse campo interdisciplinar, ressaltamos que o mapeamento dos conceitos e de seus revestimentos lexicais, neste trabalho, se dá sem a necessidade teórica da classificação dessas lexicalizações. Por isso, adotamos a expressão “itens lexicais”, sem nos esquecermos, é claro, de que tal classificação não poderia ser descartada em uma pesquisa especificamente voltada à Lexicografia e/ou à Terminografia.

Conforme viemos explicando, os itens lexicais com os quais trabalhamos são coletados, então, em fontes linguísticas da chamada língua de especialidade (o que os classificariam como “termos”), mas também da chamada língua comum (o que os classificariam como “palavras”). A opção por tais recursos heterogêneos se deu em decorrência da escassez de recursos estruturados que contemplassem com robustez o domínio alvo da pesquisa, sendo esse um empecilho comum aos trabalhos com domínios específicos, conforme relatam Almeida, Oliveira e Aluísio:

Embora já exista um número razoável de *corpora* genéricos (ou de referência, como são chamados) para várias línguas, o número de *corpora* específicos disponíveis para suporte à pesquisa terminológica ainda é deficiente. Essa deficiência dá-se pela própria especificidade de tais *corpora* que são muitas vezes construídos para serem utilizados por um período curto de tempo e somente em um projeto, daí se questionar o investimento de grandes esforços na sua compilação e anotação que visam a sua reutilização (ou reuso, como se costuma referir em linguística computacional). Ainda que seja possível construir tais *corpora* pela busca manual na web, esse processo consome muito tempo, se levarmos em conta os benefícios para pesquisas tão pontuais. (2006, p.43)

Desse modo, evidencia-se o fato, já esperado, de contarmos com mais recursos abrangendo o domínio léxico-conceitual mais geral da IB⁴¹ do que o domínio léxico-conceitual mais específico da IBI: os textos do Quadro 1 (p.14), juntamente com as transcrições das entrevistas e com os materiais de divulgação que circulam na cidade, por exemplo, estão em desvantagem quantitativa em relação aos demais recursos do nosso *corpus*, no entanto, abarcam informações altamente relevantes para o desenvolvimento desta pesquisa,

⁴¹ Advertimos que “mais geral” não se confunde com “geral”. Aquela é uma relação relativa e esta absoluta e se aplica à definição das ontologias gerais ou *Top ontologies*, que estrutura conceitos gerais como entidades, eventos, tempo, espaço, etc., conforme será explicado mais adiante.

que, ao organizar o domínio léxico-conceitual da IBI, não pode deixar de contemplar os seus níveis superiores.

Os estudos sobre a interface ontologia/léxico que visamos a desenvolver neste estudo de mestrado baseiam-se principalmente na obra “Ontology and the lexicon” (HUANG et al., 2010)⁴², uma referência do que há de mais atualizado nesse campo e que pode beneficiar diversos profissionais como pesquisadores da Engenharia do Conhecimento e da Web Semântica que levam em consideração as línguas naturais, linguistas que visam a investigar como o conhecimento lexical pode ser formalizado e tratado por máquinas e pesquisadores da Linguística Computacional que se interessam por construção de léxicos computacionais.

Como muitas são as áreas e subáreas envolvidas na investigação dessa interface, procuramos seguir a terminologia adotada em Huang et al. (2010), que ainda nos fornece grande parte das definições para os conceitos com os quais trabalhamos, não descuidando das importantes contribuições de outros pesquisadores como Hirst (2004), Vossen (2003) e Handke (1995).

2.3.1 Ontologias e léxicos

Vossen (2003) discute que, no processamento de informações, nos valem de informações de naturezas distintas, armazenadas em léxicos e ontologias; para ele, não há consenso na identificação de exatamente quais são as semelhanças e as diferenças entre ambos. Por essa razão, Huang et al. (2010) trazem considerável contribuição para o campo, ao buscarem um olhar apurado sobre as diferentes abordagens e os diferentes quadros teóricos e aplicados que tratam de ontologias e léxicos.

Para contextualizarmos e definirmos o que entendemos por ontologia, adotamos a noção vigente no âmbito da “Representação do Conhecimento”, baseada em Gruber, que é a discutida em Prévot et al. (2010) e a que se utiliza no estudo do PLN.

Como mostram Geller, Perl e Lee (2004), em levantamento histórico, quando Ross Quillian publicou o artigo *Semantic Memory*, em 1968, descrevendo um programa de computador que gerava expressões simples de língua natural, ele alcançou um feito que

⁴² Desde o ano de 2000 ocorre o “OntoLex”, um evento de âmbito internacional, focado justamente no estudo da interface entre ontologias e recursos lexicais. Huang et al. (2010) reúnem os principais artigos, revistos e atualizados, que foram apresentados nesses 10 anos de eventos.

inspirou, dentre outras coisas, o desenvolvimento do campo de estudos que seria denominado “Representação do Conhecimento”. Um dos grandes marcos nesse campo deu-se no início da década de 1990, com Thomas Gruber, que lhe oferece uma abordagem diferenciada, a da construção de ontologias⁴³: “Uma especificação de um vocabulário representacional para um domínio de discurso compartilhado – definições de classes, relações, funções e outros objetos – é chamada ontologia” (GRUBER, 1993, p.199)⁴⁴. Essa definição se aplica também ao léxico, se o concebermos como o conhecimento linguístico que se pode extrair da experiência linguística. No entanto, indicam PRÉVOT et al. (2010, p.3)⁴⁵, para “[...] entendermos diferenças mais sutis, temos que olhar mais de perto os elementos centrais da criação de ontologia: *conceitualização e especificação*. O que diferencia léxicos e ontologias encontra-se na interpretação mais clara de tais noções.”

Essas noções serão analisadas mais adiante.

Chishman (2009, p.113) complementa a definição de Gruber, explicitando os objetivos de uma ontologia:

[...] (i) compartilhar conhecimento estruturado de informações comuns entre pessoas e máquinas (sistemas computacionais); (ii) possibilitar o reuso⁴⁶ do conhecimento de determinado domínio; (iii) tornar explícito o conhecimento sobre determinado domínio; (iv) separar o conhecimento de um domínio do conhecimento operacional de construção de um sistema; (v) analisar o conhecimento de um domínio.

Hirst (2004) alerta para o fato de que, embora sejam aparentemente semelhantes a ponto de serem intersubstituíveis em alguns casos, léxicos e ontologias não são a mesma coisa:

Uma ontologia, afinal, é um conjunto de categorias de objetos ou ideias no mundo, juntamente com certas relações entre eles: não é um objeto

⁴³ Seguindo o direcionamento de Huang et al. (2010), aceitamos a diferenciação entre “Ontologia” (em maiúscula e no singular) enquanto campo filosófico e “ontologias” (no plural) enquanto artefatos de representação de conhecimento.

⁴⁴ “A specification of a representational vocabulary for a shared domain of discourse – definitions of classes, relations, functions, and other objects – is called an ontology.”

⁴⁵ “In order to understand more subtle differences one has to look closer at the central elements of ontology creation: conceptualization and specification. What distinguishes lexicons and ontologies lies in a sharper interpretation of these notions.”

⁴⁶ Pinto e Martins (2004) identificam dois processos principais de reuso de ontologias: a fusão/incorporação, na qual os conhecimentos armazenados em duas ou mais ontologias de um mesmo domínio são unificados em uma só ontologia, e composição/integração, na qual ontologias de diferentes domínios, depois de passarem por adaptações e especificações, são combinadas em uma só ontologia.

linguístico. Um léxico, por outro lado, depende, por definição, de uma língua natural e dos sentidos das palavras dela. (HIRST, 2004, p.216)⁴⁷.

Evidências dessa diferenciação são analisadas por Prévot et al. (2010). Por exemplo, enquanto que a sinonímia e a sinonímia aproximada são relações importantes para os léxicos semânticos, elas não o são para as ontologias formais, como as baseadas em linguagem de lógica de primeira ordem: “Ter dois símbolos que são logicamente equivalentes é uma redundância em qualquer teoria matemática” (PEASE; FELLBAUM, 2010, p.29)⁴⁸; já no que diz respeito ao uso de itens lexicais, os léxicos apresentam informações sobre usos, mas as ontologias tradicionais não.

A distinção entre ontologias formais e ontologias linguísticas também é destacada por Prévot et al. (2010, p.4), sendo que as primeiras são apreendidas através da lógica e são estruturas conceituais formalmente bem formadas, e as segundas, que se baseiam em língua natural, são linguisticamente convencionalizadas, não são formalmente precisas e, por isso, não são estruturas conceituais *per se*.

Quanto aos léxicos, Handke (1995) lembra-nos de que os itens lexicais podem ser armazenados na mente, em livros de referência e em dispositivos de armazenamento conectados a computadores, conforme mostra a Figura 12.

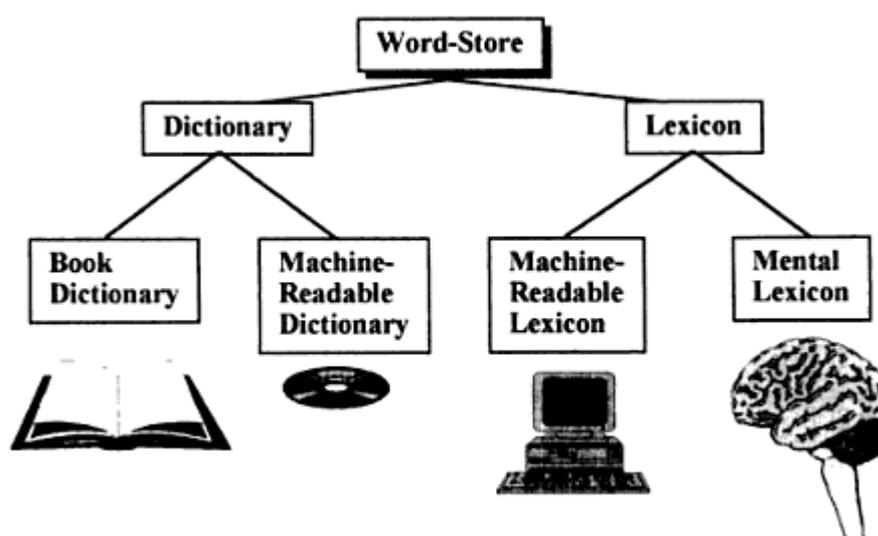


Figura 12. Uma tipologia dos acervos de itens lexicais (HANDKE, 1995, p.49).

⁴⁷ “An ontology, after all, is a set of categories of objects or ideas in the world, along with certain relationships among them; it is not a linguistic object. A lexicon, on the other hand, depends, by definition, on a natural language and the word senses in it.”

⁴⁸ “Having two symbols that are logically equivalent is a redundancy in any mathematical theory.”

Os recursos para esta pesquisa descritos em 2.1 podem ser, então, conformados com a representação do acervo de palavras descrito na Figura 12: de um lado, temos, para ‘dictionary’, tanto obras impressas, como o *Dicionário de usos do português do Brasil*, (BORBA, 2002), quanto obras em CD-ROM, como o *Miniaurélio eletrônico* (FERREIRA, 2004); de outro, temos, para ‘lexicon’, constructos como as já citadas redes WordNet e FrameNet, cada uma tratando do léxico mental para a qual foi desenvolvida.

Ressaltamos que um léxico computacional é a representação formal, com vistas a aplicações em PLN, de parte de um léxico mental, e que sua capacidade representacional depende do refinamento das relações hierárquicas que contém e da sua ancoragem conceitual. Prévot et al. (2010) explicam que muitas vezes léxico computacional, léxico relacional e recurso lexical são sinônimos no contexto do estudo em PLN, para o qual, informam os autores, dentre as definições possíveis, um léxico pode ser definido como “[...] uma coleção de conceitos linguisticamente convencionalizados [...]” (p.6)⁴⁹.

O constructo de interface, o ontoléxico, emerge das características distintas, embora relacionadas, da ontologia e do léxico. Apesar de ambos representarem conceitualizações compartilhadas, explica-nos Prévot et al. (2010), a primeira trata de conceitos e permite que máquinas processem conhecimento de modo direto (em aplicações da Web Semântica, por exemplo); já o segundo trata de itens lexicais e permite a conexão entre agentes humanos e conhecimento (em aplicações de tecnologia de língua humana, por exemplo). Os autores ainda afirmam que “É nesse contexto que a interface ontoléxico se torna um tópico de pesquisa crucial que conecta o conhecimento humano ao conhecimento da web.” (PRÉVOT et al., 2010, p.5)⁵⁰.

A conceitualização e a especificação de ontologias

De acordo com Gruber, toda base (ou sistema) de conhecimento está explicitamente ou implicitamente envolvida com alguma conceitualização⁵¹: “[...] uma visão do mundo abstrata e simplificada que desejamos representar para algum propósito.” (GRUBER, 1993,

⁴⁹ “A lexicon can be defined as a collection of linguistically conventionalized concepts [...]”

⁵⁰ “It is in this context that the ontolex interface becomes a crucial research topic connecting human knowledge to web knowledge.”

⁵¹ França (2009) explica que é com base na noção de conceitualização, como definida por Gruber, que os trabalhos no domínio da Ontolinguística são desenvolvidos. Estudar o conhecimento ontológico, como fazem os pesquisadores da interface ontoléxico, também é um dos interesses da Ontolinguística, que, ao inserir-se no âmbito da linguística cognitiva, caracteriza-se como “[...] uma área de estudo onde se procura encontrar uma ponte entre os mecanismos linguísticos que usamos no nosso dia-a-dia e o nosso conhecimento ontológico.” (FRANÇA, 2009, p.109).

p.199)⁵². Prévot et al. (2010, p.5) explicam que conceitualização, independente de situações específicas ou de línguas de representação, é o processo que leva à extração e generalização de informações relevantes a partir da experiência. Além disso,

Cada conceitualização está presa a um único agente, isto é, ela é um produto mental que representa a visão de mundo adotada pelo agente; é através das ontologias, que são especificações desses produtos mentais em alguma linguagem, que agentes heterogêneos (humanos, artificiais ou híbridos) podem julgar se uma dada conceitualização é compartilhada ou não e escolher se vale a pena negociar o significado ou não. (PRÉVOT et al., 2010, p.5)⁵³.

Vale lembrar que a linguagem referida na citação anterior não é necessariamente uma língua natural, mas sim um formalismo representacional artificial com vocabulário próprio. É apenas através dessa linguagem que se tem acesso a conceitos. Por isso, se o leigo geralmente se vale de língua natural, uma linguagem formal compreensível por máquinas se faz necessária em sociedades híbridas, compostas por diversos agentes como computadores, robôs e humanos (PRÉVOT et al., 2010, p.5).

A conceitualização tem que ser compartilhada entre agentes. No caso de agentes humanos, a conceitualização que a língua natural representa é um processo coletivo e o conteúdo informacional é definido pela coletividade de falantes (PRÉVOT et al., 2010, p.6). O fato de a língua natural servir como fonte de informação é determinante na diferenciação entre ontologia linguística e ontologia conceitual: a primeira, também chamada de ontologia descritiva, toma como recursos para o conhecimento ontológico a língua natural e o senso comum; a segunda, também chamada de ontologia revisionista, se opõe à primeira, ao capturar a natureza intrínseca de um domínio, sem levar em conta os agentes que fazem a conceitualização.

Já em ontologias lexicais, a conceitualização baseia-se em critérios linguísticos, ou seja, em informações encontradas em recursos lexicais como dicionários e *thesaurus*, que, em alguns casos, abarcam também conhecimento do mundo (ou, em outras palavras, conhecimento enciclopédico ou do senso comum). Uma breve relação entre os conhecimentos identificáveis em dicionários e enciclopédias foi apresentada na subseção 2.1.1.

⁵² “[...] an abstract, simplified view of the world that we wish to represent for some purpose.”

⁵³ “Every conceptualization is bound to a single agent, namely it is a mental product which stands for the view of the world adopted by that agent; it is by means of ontologies, which are language-specifications of those mental products, that heterogeneous agents (humans, artificial or hybrid) can assess whether a given conceptualization is shared or not and choose whether it is worthwhile to negotiate meaning or not.”

A ontologia cuja conceitualização se assenta nas experiências compartilhadas entre especialistas é também de interesse do PLN, que estuda como tais ontologias podem ser integradas a outras mais genéricas, conforme relatamos em outras partes deste estudo.

No que diz respeito à especificação, ela é a operação através da qual uma ontologia especifica a conceitualização em termos de uma linguagem de representação acima referida, ou seja, uma linguagem que formula a ontologia e que não precisa depender de uma língua natural. Essa linguagem permite que a ontologia, independente do quão complexa e explícita possa ser, funcione como “[...] a base da comunicação, a ponte através da qual o entendimento comum é estabelecido” (PRÉVOT et al., 2010, p.7)⁵⁴.

A natureza dessa linguagem determina se uma ontologia é formal, expressa em uma linguagem formal que impossibilita a existência de ambiguidades, se é informal, expressa, por exemplo, em uma língua natural, ou semi-formal, expressa em uma linguagem que mescla as duas anteriores.

Prévot et al. (2010, p.7) também indicam que uma ontologia pode ser bem rigorosa e precisa, mesmo se formulada em língua natural. Nesse caso, denomina-se “ontologia linguística”. Esse tipo de ontologia usa os sentidos de itens lexicais definidos em recursos lexicais (de maneira formal ou semi-formal, como nas wordnets) para criar os conceitos que a constituem. Observamos, entretanto, que esse procedimento precisa ser realizado com rigor, para não correremos o risco de construir recursos pobres do ponto de vista ontológico.

Handke (1995) discute uma especificação importante do léxico enquanto parte de um sistema de PLN, conferindo ao léxico o estatuto de ser:

[...] o módulo central de um sistema de processamento de língua natural, seja do homem ou da máquina. Ele interage intimamente com os outros componentes do processador da língua e fornece informações detalhadas sobre as palavras a serem produzidas ou compreendidas. (HANDKE, 1995, p.50)⁵⁵.

Desse modo, explica Handke (1995, p.68), na produção de língua natural, os itens do léxico são coletados para preencher a estrutura conceitual pré-concebida a ser manifestada em língua; já na recepção de língua natural, padrões fonológicos ou grafológicos ativam o léxico mental. Tanto para a produção quanto para a recepção, há a necessidade da especificação de

⁵⁴ “[...] the basis of communication, the bridge across which common understanding is established.”

⁵⁵ “[...] the central module of a natural language processing system, whether human or machine. It closely interacts with the other components of the language processor and provides detailed information about the words to be produced or comprehended.”

cada item lexical nos aspectos fonológicos, grafológicos, morfológicos, sintáticos e semântico-conceitual. Especificar e explicitar para a máquina tais aspectos vem se mostrando um grande desafio para o PLN, principalmente no que diz respeito ao aspecto semântico.

Handke (1995) afirma ainda que a descrição semântica para os itens do léxico “[...] tem que especificar, com bastante detalhe, as relações de sentido, usando formalismos de representação adequados.” (p.90)⁵⁶, conforme descrevem Prévot et al. (2010) nos parágrafos anteriores, tratando da (in)formalidade de, por exemplo, ontologias linguísticas.

A descrição dos fragmentos dos léxicos, frisada por Handke, pode recobrir também aspectos morfológicos, envolvendo, por exemplo, estudos sobre a formação dos itens lexicais que revestem lexicalmente a ontologia do domínio, assim como sobre o caráter de inovação deles. *Almofadrom*⁵⁷, por exemplo, é um item lexical característico de um tipo de formação de palavras que Sandmann (1991) denomina cruzamento vocabular. Esse autor descreve os principais tipos de formação de palavras, desde os mais frequentes, como composição e derivação, aos chamados de tipos especiais, que são a reduplicação (*tititi, corre-corre*), a analogia (*metroviário, bode explicatório*), a abreviação (*japa, PT, óvni*) e o cruzamento vocabular (*intelijumência, furaquinho*). Aprofundando um pouco, citamos os tipos de combinações entre classes de palavras que Borba (2003, p.25) nos traz: nome + nome (*ponto fantasia*), adjetivo + adjetivo (*luso-brasileiro*), pronome + nome (*Sua Majestade*), numeral + nome (*dois-de-paus*), advérbio + verbo, adjetivo ou nome (*bem-querer*), verbo + nome (*bate mão*), verbo + (\pm e) + verbo (*vai vem*).

Esses processos de formação de palavras destacados por Sandmann e Borba exemplificam a forte predisposição que uma língua tem no que diz respeito à inovação lexical. Sobre isso fala-nos Correia e Lemos (2005), explicando que a neologia “[...] traduz a capacidade natural de renovação do léxico de uma língua pela criação e incorporação de unidades novas, os neologismos.” (p.13). Na língua temos, como demonstração da criatividade lexical, a neologia denominativa e a criação neológica estilística, sendo que a primeira é particularmente interessante para nós, uma vez que estudamos um domínio de conhecimento que, a todo o momento, nomeia novas realidades (especialmente novos objetos) através da criação de itens lexicais como *capa de microondas* ou *bordado computadorizado*.

⁵⁶ “[...] has to specify sense relationships in great detail, using adequate representation formalisms.”

⁵⁷ O *almofadrom* é uma [almofada + edredom], que, quando desempacotada, se transforma em um edredom. Essa peça, que é suporte para vários tipos de bordados, foi criada em 2008 na cidade de Ibitinga. O conceito ALMOFADROM pode não ter contrapartida linguística em outras línguas, mas, certamente, integra a ontologia da IBI em algum nível.

Correia e Lemos (2005, p.17-18) propõem três níveis para a classificação dos tipos dos neologismos, conforme resume o Quadro 4.

- Novidade formal (a sua forma significante é nova): quando o neologismo apresenta uma forma não atestada no estágio anterior de registro de língua;⁵⁸
- Novidade semântica: quando o neologismo corresponde a uma nova associação significado-significante, isto é, uma palavra já existente adquire um novo sentido;
- Novidade pragmática: quando a neologia resulta da passagem de uma palavra previamente usada num dado registro para outro registro da mesma língua.

Quadro 4: Tipos de neologismos (CORREIA; LEMOS, 2005, p.17-18).

Os diferentes níveis de ontologias

As pesquisas com ontologias reconhecem três níveis de especificidade que esses constructos podem apresentar: a ontologia de nível superior (*top-level*), a de nível intermediário (*core level* ou *reference*) e a de domínio (*domain ontology* ou *domain level*), conforme está ilustrado na Figura 13, adaptada de Prévot et al. (2010). No nível superior, encontram-se as categorias e relações mais gerais da taxonomia de conhecimento, cujos termos, em sua grande maioria, não estão/são lexicalizados, enquanto que a ontologia de domínio abarca domínios específicos de aplicação (que podem ser bem restritos e cujos termos são praticamente todos lexicalizados); conectando os níveis opostos, está a ontologia de nível intermediário, que tem a maioria dos seus termos lexicalizados.

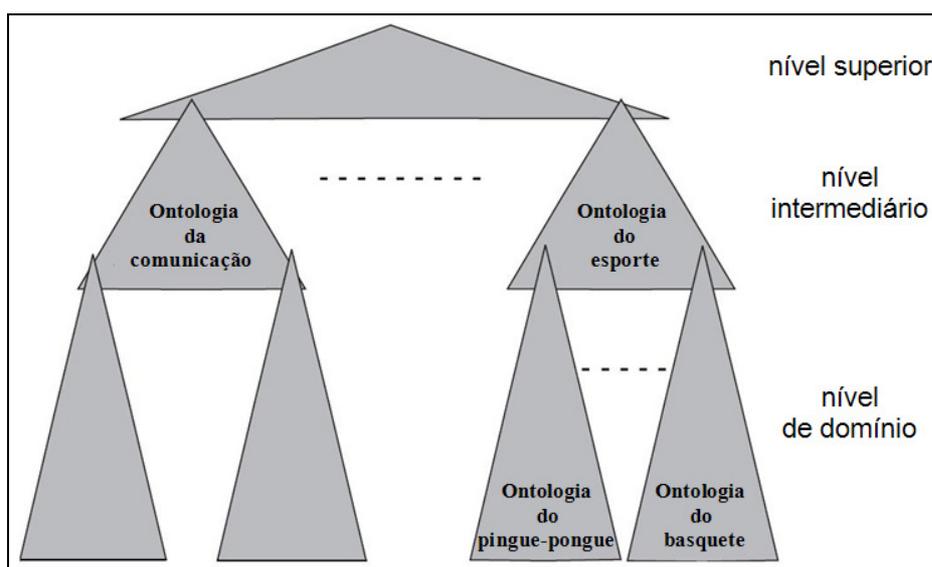


Figura 13. Níveis de especificidade das ontologias (PRÉVOT et al., 2010, p.8).

⁵⁸ Os itens lexicais que resultam de importação, os estrangeirismos, como *richelieu*, caracterizam, então, esse tipo de novidade.

O conteúdo das ontologias

Como se deduz, os elementos de uma ontologia são de natureza conceitual e, como explicam Prévot et al. (2010, p.10), são denominados conceitos, tipos, categorias ou propriedades. Frequentemente eles são caracterizados extensionalmente em termos de classes e correspondem, nesse caso, a conjuntos de instâncias ou indivíduos. Indivíduos (denotados por nomes próprios e outras entidades nomeadas) às vezes são tratados como outros conceitos em ontologias derivadas diretamente de recursos lexicais⁵⁹, no entanto, a diferença entre classes e instâncias tem que ser bem estabelecida para se construir uma ontologia robusta:

[...] a confusão entre conceito e indivíduo não é nada além do produto da “falta de expressividade”. Na verdade, se houvesse uma relação **instância-de**, poderíamos distinguir entre uma relação conceito-conceito (subsunção) e uma relação indivíduo-conceito (instanciação). (GANGEMI et al., 2010, p.37)⁶⁰.

Usando a rede WN.Pr como exemplo de um léxico baseado no sentido (*sense-based lexicon*), Prévot et al. (2010, p.11) colocam a problemática de como situar esse tipo de léxico em que os nós não são nem termos puros e nem conceitos puros, mas sim sentidos de palavras que correspondem a um uso convencionalizado de uma palavra. Desse modo, em princípio a rede WN.Pr é um léxico, já que todas as suas entradas são expressões linguísticas, mas a estrutura semântica definida pelos synsets e pelas relações entre synsets tem sido usada como uma ontologia linguística.

Na construção de ontologias são destacados dois procedimentos principais⁶¹:

- 1) *Top-down*: o principal passo do processo é a determinação de uma taxonomia guia que forneça as principais categorias e a arquitetura taxonômica para elas organizada por meio da relação ‘é-um-tipo-de’ (*is-a-kind-of*). Adaptar tais estruturas a léxicos constitui uma abordagem *top-down*, pois os sentidos das

⁵⁹ As primeiras versões da rede WN.Pr, por exemplo, não diferenciam tipos e instâncias, por isso o synset {*Rio de Janeiro, Rio*} é indicado como “hipônimo-de” (um tipo de) {*city, metropolis, urban center*} quando na verdade ele é um indivíduo (uma instância). Atenta a essa problemática, a WN.Pr instaura, a partir de sua versão 2.1, a relação “instância-de”.

⁶⁰ [...] the confusion between concept and individual is nothing but the product of a “lacking in expressiveness”. In fact, if there was an **instance-of** relation, we could distinguish between a concept-to-concept relation (subsumption) and an individual-to-concept one (instantiation).

⁶¹ Ding e Foo (2002) citam também a abordagem *middle-out*, que visa à construção da ontologia partindo dos conceitos centrais em direção aos mais gerais e aos mais específicos.

palavras serão fortemente determinados pela posição de suas ligações na taxonomia. (PRÉVOT et al., 2010, p.11).

- 2) *Bottom-up*: nesse processo a ontologia emerge dos itens lexicais, sendo que o questionamento sobre a existência ou não de sentidos de palavras⁶² pode ser levado em consideração, mas não é essencial; o foco fica sobre os cuidados em se lidar com a granularidade entre os vários recursos de onde provêm os itens lexicais. (PRÉVOT et al., 2010, p.12).

Variadas ontologias já foram construídas e estão disponíveis atualmente para toda sorte de aplicação. Sobre a profusão de ontologias desenvolvidas sob diferentes prismas falamos Prévot, Borgo e Oltramari (2010):

Ontologia, enquanto um ramo da representação do conhecimento, é uma área de pesquisa nova e com várias fragilidades, duas das quais são a falta de metodologias estabelecidas e a falta de um critério de avaliação confiável. Então, não deveria ser uma surpresa descobrir que as ontologias disponíveis hoje têm sido construídas seguindo abordagens discrepantes que resultam em sistemas bem diferentes. (p.188-189).⁶³

Prévot et al. (2010, p. 12) falam de relações de natureza conceitual e de natureza lexical: nas ontologias os conceitos são integrados em um todo coerente por meio de relações que são conceitualmente guiadas e que tomam conceitos como argumentos, nas ontologias formais, por exemplo, a relação ‘é-um’ está sobrecarregada e vem sendo muito usada, embora de maneira pouco definida e sendo correspondida a intuições ativadas pela sua expressão em língua natural (‘é-um’); já nos recursos lexicais, nos quais a preocupação recai sobre a organização de itens lexicalizados, as relações têm apenas natureza conceitual indireta, como no caso da antonímia (estabelecida entre formas de palavras e não entre conceitos) e da hiperonímia (entre sentidos de palavras). Enquanto as ontologias formais vêm se preocupando em como definir a natureza das propriedades formais das relações (se elas se relacionam a classes ou indivíduos, se são reflexivas, simétricas, transitivas, etc.), os recursos lexicais se

⁶² Prévot et al. (2010) apontam para a impossibilidade de se listar exhaustivamente a polissemia regular ou o uso criativo em uma língua natural. Hirst (2004, p.210) comenta que, às vezes, os sentidos diferentes de uma palavra são discretos o bastante a ponto de podermos tratá-los como distintos, mas nem sempre é assim, e tentar separá-los em sentidos discretos, como as definições dicionarísticas tentam fazer, não é possível.

⁶³ “Ontology, as a branch of knowledge representation, is a young research area with several weaknesses, two of which are the lack of established methodologies and the lack of reliable evaluation criteria. Thus, it should not come as a surprise to discover that ontologies available today have been built following disparate approaches resulting in quite different systems.”

debruçam sobre métodos para descobrir relações automaticamente e sobre a aplicação delas na extração de conhecimento lexical (PRÉVOT et al., 2010, p. 12).

Embora possam ter o mesmo nome nas abordagens conceituais e nas lexicais, as relações são diferentes quando analisadas sob um olhar minucioso, no entanto, a classificação geral dessas relações em paradigmáticas ou sintagmáticas é válida para as abordagens de ambas as naturezas, conforme acompanhamos, com Prévot et al. (2010, p.13-14), a seguir.

As relações paradigmáticas aplicam-se entre elementos de mesma natureza (termos intersubstituíveis em um dado contexto) que pertencem a um paradigma comum e que geralmente são da mesma categoria sintática. Cruse (1986) associa essas relações a relações de congruência entre classes (identidade, inclusão, sobreposição e disjunção); enquanto que no domínio lexical as relações paradigmáticas consagradas são sinonímia, antonímia, meronímia, hiperonímia e hiponímia, relações conceituais correspondentes são definidas formalmente em ontologias em termos de oposição conceitual, ‘parte-de’ e ‘é-um-tipo-de’⁶⁴.

Prévot et al. (2010, p.13) ainda citam a relação denominada paranímia (*paronymy*), que pode ser utilizada, explicam Huang et al. (2008), para enriquecer o conhecimento ontológico de wordnets ao identificar e classificar grupos conceituais entre termos coordenados (ou irmãos); eles citam como exemplo o fato de que os synsets {*north*}, {*south*}, {*east*} e {*west*}, apesar de serem hipônimos de {*cardinal compass point*}, logo termos irmão, não são equivalentes, pois há pelo menos dois outros pares conceituais salientes que desempenham papéis importantes em outras classificações conceituais, *north-south* e *east-west*, que poderiam ser incluídos na rede via relação de paranímia⁶⁵.

Às vezes a variação terminológica faz com que rótulos que relacionam entidades lexicais sejam usados para relacionar conceitos, por exemplo: hiperonímia (e hiponímia) funcionando como a relação ‘é-um’ (ou ‘é-um-tipo-de’) em afirmações como CAMINHÃO ‘é-hipônimo-de’ VEÍCULO (onde o correto seria *caminhão* ‘é-hipônimo-de’ *veículo* ou CAMINHÃO ‘é-um-tipo-de’ VEÍCULO); sobre isso Hirst (2004) explica que:

⁶⁴ Hirst (2004, p.216) explica que, além das relações lexicais clássicas, como as utilizadas na rede WN.Pr, há muitas outras, como as relações associativas ou de tipicidade, ele cita como exemplos a relação entre *cachorro* e *latir*, entre *prego* e *martelo* e entre *médico* e *hospital*. Podemos complementar, como será feito mais adiante, vislumbrando algumas relações importantes dentro do domínio focado pela pesquisa, por exemplo, entre *bordadeira* e *bastidor* ou entre *bordado* e *guardanapo*.

⁶⁵ Huang et al. (2008) detalham que a paranímia é, então, uma relação paradigmática que marca agrupamentos conceituais entre termos irmãos e complementa as relações de hiperonímia/hiponímia presentes na rede WN.Pr. Eles citam que a rede ganharia em representatividade semântica se a relação de paranímia fosse aplicada, por exemplo, na estruturação das estações: os synsets {*rainy season*}, {*dry season*}, {*winter, wintertime*}, {*summer, summertime*}, {*fall, autumn*} e {*spring, springtime*} são todos hipônimos de {*season, time of year*}, mas synsets intermediários entre o hiperônimo e seus hipônimos podem ser propostos com a relação paranímica, especificando-se um synset {*tropical climate*} que abarque {*rainy season*} e {*dry season*} e um outro {*general climate*} que abarque os demais.

Independentemente de como entendemos os sentidos das palavras, porque eles pertencem de algum modo a categorias no próprio mundo, relações lexicais entre sentidos de palavras espelham, talvez imperfeitamente, certas relações que se aplicam entre as próprias categorias. (p.214-215)⁶⁶.

As relações sintagmáticas aplicam-se entre entidades de naturezas diferentes, geralmente lexicalizadas por itens de categorias sintáticas diferentes que co-ocorrem mas que não são intersubstituíveis. Na semântica lexical essas relações são associadas a estudos na interface sintaxe/semântica com foco em predicação e papéis temáticos; há também o estudo de relações sintagmáticas enquanto relações de caso, conforme mostra Fillmore em sua gramática de casos, que, hoje, se consolida na semântica de frames.

Se a rede WN.Pr e outros trabalhos baseados em ontologias vêm focando as relações paradigmáticas, a FrameNet, que implementa computacionalmente a semântica de frames, vem focando as relações sintagmáticas. Tanto os construtores de ontologias quanto os linguistas computacionais reconhecem que a complementaridade entre relações paradigmáticas e sintagmáticas das redes WordNet e FrameNet, respectivamente, e a combinação eficiente delas é um “[...] elemento importante dessa área de pesquisa aplicada” (PRÉVOT et al., 2010, p.14). Na subseção 3.2, desenvolveremos um exercício exploratório que aproveita as vantagens da complementaridade entre essas duas redes.

2.3.2 A interface entre léxico-ontologia

Entendidos, então, os recursos lexicais ou as ontologias linguísticas como estruturas motivadas por língua natural, em especial pelo léxico, e as ontologias formais como ontologias motivadas por considerações filosóficas, pela eficiência na representação de conhecimento e por princípios de reutilização (*reusability*), podemos situá-los no vasto campo da representação de conhecimento, em que diferentes abordagens se distinguem e se entrecruzam:

⁶⁶ “Regardless of exactly how one conceives of word senses, because they pertain in some manner to categories in the world itself, lexical relationships between word senses mirror, perhaps imperfectly, certain relationships that hold between the categories themselves.”

O t3pico da interface entre ontologias e recursos lexicais 3 ent3o um re-exame de assuntos tradicionais da psicolinguística, linguística, intelig3ncia artificial e filosofia sob a luz de avanços recentes nessas disciplinas e em resposta a um interesse renovado nesse t3pico devido a sua relev3ncia para as principais aplicaç3es da Web Sem3ntica. (PR3VOT et al., 2010, p.14)⁶⁷.

Ontologias e recursos lexicais, embora divergentes no que diz respeito à conceitualizaç3o, especificaç3o e escopo, t3m uma natureza parcialmente similar, um paralelismo de estruturas, e tomam como base, conforme afirmam Pr3vot et al. (2010) os mesmos objetos: “Conceitos (ou palavras) e relaç3es s3o, portanto, os dois primeiros objetos a serem considerados no trabalho com ontologias e recursos lexicais. Esse paralelismo na estrutura deles define a interface ontol3xico.” (p.10)⁶⁸.

A combinaç3o da informaç3o conceitual encontrada em ontologias e da informaç3o lexical 3, ent3o, o ponto-chave para as disciplinas citadas e, conseqüentemente, para as discuss3es deste estudo. Por esse motivo, apresentaremos, oportunamente, uma s3mula das principais estruturas te3ricas que est3o por tr3s da interface “ontol3xico”.

No campo dos estudos cognitivos a sem3ntica componencial de Katz e Fodor (KATZ; FODOR, 1963) teve papel de destaque, mas sua proposta de categorizaç3o de palavras atrav3s de traços sint3ticos e sem3nticos foi se tornando limitada com os desenvolvimentos da noç3o de prototipia (ROSCH, 1973), que foi a pioneira nesse tipo de modelagem escalar da conceitualizaç3o humana: “[...] a associaç3o de palavras e at3 conceitos a categorias 3 uma noç3o escalar [...] O problema das fronteiras categoriais, de suas exist3ncias e da suas determinaç3es potenciais, 3 s3rio.” (PR3VOT et al., 2010, p.15).

Al3m da quest3o da prototipia, fala-se do fato de a l3ngua ter impacto significativo na categorizaç3o.

A predicaç3o tem sido igualmente focada, j3 que ela direciona a interpretaç3o das sentenç3as. Pr3vot et al. (2010, p.16) citam o pioneirismo de Fillmore ao propor, com a *Gram3tica de Casos*, que se analisem as palavras em relaç3o umas com as outras dentro do *frame* ou *script* em que ocorrem.

⁶⁷ “The topic of the interface between ontologies and lexical resources is therefore a re-examination of traditional issues of psycho-linguistics, linguistics, artificial intelligence, and philosophy in the light of recent advances in these disciplines and in response to a renewed interest in this topic due to its relevance for the Semantic Web major applications.”

⁶⁸ “Concepts (or words) and relations are therefore the first two objects to consider while working with ontologies and lexical resources. This parallelism in their structure defines the ontol3x interface.”

Desde Aristóteles⁶⁹ até autores atuais, a Filosofia vem tentando determinar um sistema de categorias ontológicas que possa propor um exame de questões como *Quais tipos de coisas existem?*, afirmam Prévot et al. (2010), que complementam dizendo que os desenvolvedores de ontologia modernos “[...] não procuram pela ontologia perfeita, mas consideram as várias ontologias potenciais que dizem respeito aos diferentes domínios e que captem as diferentes percepções da mesma realidade.” (p.17)⁷⁰.

No campo dos estudos filosóficos os principais trabalhos visam a postular um conjunto de categorias fundamentais que auxiliem a tomada de decisões sobre a posição dos conceitos nas ontologias, mesmo que não se disponha de uma taxonomia perfeita. Outro aspecto importante é o das abordagens revisionistas e descritivas, em que se baseiam as ontologias linguísticas e conceituais, respectivamente, conforme informamos na subseção 2.3.1.

Já que os dicionários tradicionais são os ancestrais dos nossos recursos lexicais computacionais, não é de se surpreender que os mais recentes estudos no campo da lexicografia tenham se deparado com a maioria das dificuldades que as abordagens contemporâneas enfrentam (PRÉVOT et al., 2010, p.19), e tais dificuldades incluem, dentre outras questões, a caracterização de diferentes tipos de dicionários (enciclopédicos X gerais, semasiológicos X onomasiológicos, etc.) e a deliberação sobre a economia da obra (inclusão ou exclusão e de quais entradas) e sobre as relações com o conhecimento enciclopédico (partes de algumas dessas dificuldades foram discutidas na subseção 2.2).

Juntamente com os filósofos da linguagem, os semanticistas interessam-se, no desenvolvimento de sistemas de línguas naturais e nos estudos da linguagem, por ontologias de língua natural e por saber com quais categorias ontológicas o estudo de língua natural está envolvido (PRÉVOT et al., 2010, p.20). Nesse campo da linguística e dos estudos lexicais contemporâneos, a teoria do léxico gerativo teve impacto no desenvolvimento de recursos práticos, já que contribui enormemente para o estudo do léxico e sua relação com ontologia.

A teoria de Pustejovsky, o *Léxico Gerativo*, que une um arcabouço filosófico-computacional com uma robusta teoria léxico-semântica, na qual a predicação tem papel central, aborda, com novos olhos, os recursos lexicais tradicionais, que restringem-se na

⁶⁹ Teixeira-da-Mata faz um apanhado sobre análises da obra clássica ‘Categorias’, de Aristóteles, reforçando algumas teses do estudioso Franz Brentano, que entende que, de acordo com Aristóteles, as categorias existem independentemente da mente: “[...] Aristóteles empresta muita exterioridade à sua tábua [das categorias], dando-lhe, por conseguinte, um caráter totalmente autônomo do pensamento. Desse modo, as categorias dariam os contornos mais gerais da própria realidade sensível.” (TEIXEIRA-DA-MATA, p.22). O autor ainda destaca a visão brentaniana de que “[...] a tábua das categorias atende a uma preocupação de natureza ontológica primordial.” (p.23).

⁷⁰ “Modern ontology designers are not looking for a perfect ontology but consider many potential ontologies concerning different domains and capturing different views of the same reality.”

construção de listas de sentidos para as palavras e, por isso, são denominados, pelo autor, de “léxicos enumerativos”. Em oposição a essa visão, Pustejovsky propõe o “léxico gerativo”, em que as entradas lexicais são dinâmicas e correspondem a combinações de vários aspectos do significado, entre eles, os papéis *qualia*, permitindo também que mecanismos de coerção de sentidos, entre outros, gerem novos sentidos a partir de sentidos já existentes no léxico. Sendo assim, indo além da visão limitada de uma árvore para representar os sentidos, o léxico gerativo, para a melhor organização do conhecimento, se vale de relações que levam em consideração a visão multidimensional do sentido lexical.

Exemplifiquemos, apenas como ilustração e de modo informal, como a estrutura *qualia* especifica quatro aspectos essenciais do significado de um item lexical, conforme explica Pustejovsky (1996, p.76), denominados papéis ou “*quale*”. Tomemos o item lexical *linha*:

- **Papel constitutivo:** indica a relação entre o objeto e suas partes constituintes: *linha* é constituída de FIBRAS (algodão, seda, lã, etc.)
- **Papel formal:** distingue o objeto dentro de um domínio maior (suas características físicas): *linha* é um tipo de FIO (torcido, maleável, segmentável, colorido, etc.)
- **Papel télico:** indica o objetivo e a função do objeto: *linha* serve COSTURA/BORDADO (para costurar, bordar, etc.)
- **Papel agentivo:** indica como se dá a origem do objeto: *linha* é ARTEFATO (fabricado pelo homem a partir de fibras).

2.3.3 A “mão-dupla” entre léxico e ontologia

A visão defendida em Huang et al. (2010) e Hirst (2004), que adotamos neste estudo, é a de que há uma bidirecionalidade da relação entre ontologias e recursos lexicais (léxicos). Prévot et al. (2010) se manifestam sobre essa questão da bidirecionalidade: “Consideramos como a metodologia das ontologias e o conhecimento podem aprimorar os recursos lexicais e como esses recursos podem, por sua vez, beneficiar as ontologias.” (p.21)⁷¹.

⁷¹ “We consider both how ontological methodology and knowledge can enhance lexical resources and how these resources can in turn benefit ontologies.”

Hirst (2004, p.222) também contribui, quando contrasta as ontologias baseadas em léxicos (*lexically based ontologies*) e os léxicos baseados em ontologias (*ontologically based lexicons*), ou seja, há tanto a possibilidade de uma ontologia servir de base para a construção de léxicos quanto a possibilidade de léxicos estruturados semanticamente servirem de base para a construção de uma ontologia, sobretudo em se tratando da construção da ontologia de um domínio técnico, no qual a correspondência entre os itens lexicais e os conceitos da ontologia do domínio é mais próxima que na construção de ontologias para domínios gerais (as ontologias de nível superior em que conceitos são denotados por itens lexicais da língua geral):

[...] em domínios técnicos onde existem vocabulários explícitos (incluindo glossários, léxicos, dicionários de termos técnicos, etc., apoiados ou não por uma autoridade), uma ontologia existe pelo menos implicitamente [...] E onde uma ontologia explícita existe, um vocabulário explícito certamente também existe; na verdade, frequentemente se diz que a construção de qualquer ontologia de domínio específico implica a construção paralela de um vocabulário para ela [...]. (HIRST, 2004, p.223)⁷².

Quando Hirst (2004) explica que é comum haver a construção paralela de uma ontologia e de fragmentos de léxico (vocabulário), ele está descrevendo procedimentos aproximados aos que adotamos nesta pesquisa: a determinação dos conceitos da ontologia prevista no item (Ia) dos objetivos é acompanhada pela determinação do relacionamento do significado entre itens lexicais e conceitos da ontologia (cf. IIc); ou seja, parte da ontologia é construída juntamente com parte do revestimento lexical ancorado a ela no processo de edificação do ontoléxico. O autor ainda alerta para a impossibilidade de uma ontologia, principalmente uma de nível superior, ser concebida partindo apenas do léxico, pois este, devido ao fenômeno das lacunas lexicais⁷³, “[...] omitirá qualquer referência a categorias ontológicas que não são lexicalizadas na língua – categorias que requereriam uma descrição multi-palavra (possivelmente longa) para serem referidas na língua.” (HIRST, 2004, p.218)⁷⁴.

⁷² “[...] in technical domains where explicit vocabularies exist (including glossaries, lexicons and dictionaries of technical terms, and so on, whether backed by an authority or not), an ontology exists at least implicitly [...] And where an explicit ontology exists, an explicit vocabulary certainly does; indeed, it is often said that the construction of any domain-specific ontology implies the parallel construction of a vocabulary for it [...]”.

⁷³ As lacunas lexicais refletem o fato de haver mais conceitos que sentidos de itens lexicais (HIRST, 2004, p.218). Nesse ponto podemos apreciar novamente a distinção entre os diferentes níveis de especificidade das ontologias em comparação com o fenômeno da lexicalização (subseção 2.3.2).

⁷⁴ “[...] will omit any reference to ontological categories that are not lexicalized in the language – categories that would require a (possibly long) multi-word description in order to be referred to in the language.”

Prévoit et al. (2010, p.22) continuam discorrendo sobre a bidirecionalidade da relação entre ontologias e recursos lexicais:

- na direção “das ontologias para os léxicos”, busca-se o aprimoramento ontológico dos recursos lexicais, usando-os como fontes consistentes de conhecimento para aplicações na engenharia do conhecimento.
- na direção “dos recursos lexicais para o conhecimento ontológico”, busca-se a população de recursos ontológicos com termos lexicais. Desse modo é possível criar grandes ontologias do senso comum ou de conhecimento de domínio, seja populando uma ontologia de nível superior já existente, seja criando ontologias completas para um domínio.

A interface ontoléxico, que é uma “[...] tentativa de lidar com a necessidade crescente de se modelar o inter-relacionamento complexo entre léxicos e ontologias” (LENCI, 2010, p.241)⁷⁵, pode ser trabalhada, então, conforme o direcionamento metodológico escolhido para se adequar aos objetivos pretendidos. Nesse sentido, Prévoit, Borgo e Oltramari (2010, p.186-187) complementam o parágrafo anterior detalhando três métodos disponíveis para a interface.

- 1) **Reestruturação de um léxico computacional** com base em princípios guiados por ontologias: o foco é dado ao recurso lexical, envolvendo a ontologia apenas no *metanível (metalevel)*. Nessa opção metodológica, a reestruturação ontológica é realizada seguindo-se restrições formais de um modelo ontológico.
- 2) **Enriquecimento de uma ontologia com informação lexical** de duas maneiras: (a) populando a ontologia, a fim de incrementar o volume e a cobertura dela, com entradas lexicais, e (b) anexando informação lexical aos conceitos da ontologia. Em (a), o léxico computacional e a ontologia são utilizados como taxonomias de termos vazias, o primeiro contendo apenas conceitos lexicalizados e relações linguísticas e a segunda fornecendo estrutura formal para conceitos lexicalizados e não-lexicalizados e para relações. Dessa maneira, as unidades lexicais são mapeadas nas entradas ontológicas, focando-se o *nível do objeto (object-level)*, por isso as restrições formais recaem sobre as categorias e relações ontológicas já implementadas na ontologia existente. Em (b), a estrutura do conhecimento, ao

⁷⁵ “[...] is an attempt to address the increasing need of modelling the complex interrelationship between lexicons and ontologies.”

contrário do que acontece em (a), permanece a mesma, apenas os elementos ontológicos é que são linguisticamente incrementados com informações lexicais (por exemplo, com partes do discurso).

- 3) **Alinhamento entre ontologia e recurso lexical:** considerada, pelos autores, a opção metodológica mais completa, pois cria um sistema ontologicamente robusto e linguisticamente motivado, ao combinar os dois níveis das opções anteriores, o *metanível* e o *nível do objeto*.

Com a bidirecionalidade discutida nesta subseção, o PLN tanto serve como aplicação e ferramenta para a interface dos ontoléticos – por exemplo, técnicas de PLN para a construção semi-automática de ontologias variam da análise semântica e *parsers* (analísadores gramaticais) à extração de termos e relações – quanto se beneficia de recursos lexicais ontologicamente tratados – por exemplo, em aplicações de recuperação de informação e na resolução de co-referência (PRÉVOT et al., 2010, p.22).

2.4 Um exercício de alinhamento entre léxicos e ontologias

Baseados em Pease e Fellbaum (2010), lembramos que o falante conhece milhares de conceitos e que os itens lexicais que expressam esses conceitos devem ser armazenados e recuperados de modo eficiente e econômico. A rede WN.Pr é, pois, uma proposta de modelar a relação entre conceitos e itens lexicais.

Uma vez que o léxico pode ser definido como o mapeamento de conceitos em itens lexicais (PEASE; FELLBAUM, 2010, p.27), a rede WN.Pr, enquanto um léxico estruturado, mostra que o léxico obedece a padrões de organização semântica; no entanto, não é raro identificarmos lacunas estruturais (às vezes a geometria das relações requer um item lexical que a língua não possui) ou padrões de lexicalização diferentes entre línguas (quando comparando léxicos de línguas diferentes). Então, “[...] os mapeamentos conceito-palavra de qualquer língua são, até certo ponto, acidentais; palavras existentes não refletem totalmente o inventário de conceitos que está disponível.” (PEASE; FELLBAUM, 2010, p.27)⁷⁶. Esse inventário, continuam os autores, pode ser representado numa ontologia não-lexical como a SUMO.

⁷⁶ “[...] the concept-word mappings of any given language are to some extent accidental; existing words do not fully reflect the inventory of concepts that is available.”

A ontologia formal SUMO (*Suggested Upper Merged Ontology*) (NILES; PEASE, 2001) baseia-se em uma linguagem de lógica de primeira ordem com vários termos (conceitos nomeados na ontologia) e axiomas (afirmações na lógica). Embora não seja possível objetivar quais conceitos possam ser considerados de nível superior ou de nível de domínio, nesse corte arbitrário diz-se que a ontologia SUMO é uma ontologia de nível superior, cobrindo conceitos gerais como tempo, relações espaciais, processos, etc.

Enquanto que, numa rede semântica como as wordnets, usa-se muito de definições em língua natural para expressar conceitos, numa ontologia formal apenas os axiomas como afirmações matemáticas dão aos termos seus significados. Por isso, a ontologia SUMO abre mão de relações semânticas entre itens lexicais (como a sinonímia utilizada pelas wordnets) para se valer de ligações entre conceitos axiomatizados formalmente e de vários rótulos (que podem incluir itens em diferentes línguas naturais). Esse caráter de independência em relação às línguas naturais confere à ontologia SUMO a possibilidade de se criar uma interlíngua.

Pease e Fellbaum (2010) descrevem possíveis tipos de mapeamento WordNet-SUMO e atestam a aplicação de três:

- Por equivalência aproximada (*rough equivalence*): por exemplo, {*satellite, artificial satellite, orbiter*} é mapeado ao termo ARTIFICIAL_SATELLITE;
- Por subsunção (*subsuming*)⁷⁷: por exemplo, {*elk, European elk, moose, Alces alces*} é mapeado ao termo HOOFED_MAMMAL;
- Por instância (*instance*): por exemplo, {*Washington, George Washington, President Washington*} é mapeado ao termo HUMAN.

Essa abordagem apresenta, no entanto, uma limitação: [...] um mapeamento único de uma entidade lexical a um termo formal não captura totalmente o sentido de alguns itens lexicais, mesmo se existisse a opção de ligar ou criar um termo formal bem específico. (PEASE; FELLBAUM, 2010, p.32)⁷⁸.

Os autores citam o exemplo do synset {*continue*} *exist over a prolonged period of time*, que não pode ser expresso por um único termo (já que ele pode referir-se a vários tipos de processos não relacionados). Seria necessária, então, uma estrutura de relações mais complexa para expressar a semântica desse item lexical. Porém, defendem eles que tal

⁷⁷ Esse mapeamento é motivado pela noção de subsunção advinda da lógica; Vossen (2003, p.467) explica a subsunção definindo-a como a relação entre dois conceitos C_1 e C_2 em que C_1 subsume C_2 se e somente se todas as propriedades que são verdadeiras para C_1 são também para C_2 , mas há certas propriedades verdadeiras de C_2 que não são necessariamente verdadeiras em C_1 . Na construção de ontologias a subsunção se dá em termos da relação 'é-um-tipo-de'.

⁷⁸ “[...] a single mapping from a lexical entity to a formal term does not fully capture the meaning of some lexical items, even IF there is the option of linking to or creating a very specific formal term.”

problemática não nega a utilidade e a necessidade de se expressar as relações em casos mais simples.

Por fim, explicitam a possibilidade de trabalharmos com a abordagem híbrida: a de construção de ontologias formais e a de ontologias baseadas em redes semântico-lexicais como as wordnets:

Muitas wordnets têm sido alinhadas à WordNet do inglês e assim também à ontologia SUMO. A natureza formal e de independência de uma língua da ontologia SUMO possui certa promessa de permitir aos criadores de novas wordnets a verificação desses alinhamentos entre línguas ao testá-los em comparação com uma definição lógica e formal, ao invés das definições e das relações semânticas da WordNet. (PEASE; FELLBAUM, 2010, p.35)⁷⁹.

Outro artigo presente em Huang et al. (2010), referência crucial para o desenvolvimento desta pesquisa, diz respeito à conexão entre ontologias e outro importante recurso lexicogramatical, ora em construção para o inglês, a rede FrameNet. Scheffczyk, Baker e Narayanan (2010) destacam alguns dos procedimentos dessa abordagem, que envolveu a conversão automática de parte da rede FrameNet para linguagens formais que possibilitassem a ligação desse recurso lexical a ontologias do tipo OWL ou SUMO.

A rede FrameNet definiu cerca de 40 tipos semânticos em sua rede. Por exemplo, o tipo semântico ‘month_name’ associado a *January.n*, *February.n* e etc., permite que seja possível capturar o fato de que essas unidades lexicais se comportam de maneira semelhante (FILLMORE; JOHNSON; PETRUCK, 2003, p.245). Uma vez que a rede FrameNet privilegia uma descrição de língua e não de conhecimento de mundo (mas do mundo experienciado e registrado nos textos), ela especifica tipos semânticos que são mais superficiais, e, portanto, bastante específicos, tendo menos relações entre eles do que os especificados pelas classes de ontologias. Mas, mesmo assim, seus tipos semânticos são bem similares aos conceitos semânticos (ou classes) definidos em ontologias como a ontologia SUMO (SCHEFFCZYK; BAKER; NARAYANAN, 2010, p.56).

O alinhamento dos tipos semânticos da rede FrameNet à ontologia SUMO permite o desenvolvimento de uma abordagem semi-automática que conecta os FEs (elementos dos frame) às classes da ontologia SUMO, o que, por sua vez, possibilita alinhamentos de

⁷⁹ “Many wordnets have been linked to English WordNet and thus also to SUMO. The formal and language-independent nature of SUMO holds some promise in enabling creators of new wordnets to verify these cross-language links by testing them against a formal, logical definition, rather than WordNet’s definitions and semantic relations.”

domínio específico de FEs na SUMO através de exemplos anotados de um domínio conceitual particular. Tais procedimentos, aliados a outros de caráter mais computacional que não descreveremos aqui, geram uma união entre a rede FrameNet e a ontologia SUMO que tem potencial para, por exemplo, ajudar os analisadores semânticos na desambiguação de sentidos dos itens lexicais (SCHEFFCZYK; BAKER; NARAYANAN, 2010, p.66).

Outros exemplos para a interface entre recursos lexicais e ontologias são facilmente encontrados na literatura; como uma das referências mais importantes indicamos Huang et al. (2010), no qual acompanhamos, dentre outras propostas, o alinhamento entre a rede WN.Pr e a ontologia DOLCE (GANGEMI et al., 2010).

A ontologia DOLCE é uma ontologia de nível superior que tem uma orientação cognitiva que busca capturar as categorias ontológicas subjacentes à língua natural e ao senso comum, sendo que tais categorias envolvem-se, então, na representação do mundo projetado (não o mundo real) pela percepção, pela cultura e pelas convenções sócio-culturais humanas (GANGEMI et al., 2010, p.39). Dessa forma, as orientações nas quais se baseia a ontologia DOLCE assemelham-se às que baseiam a rede WN.Pr, daí a interoperabilidade entre os dois constructos e a importância deles para a interface dos ontológicos e, por conseguinte, para a nossa pesquisa (mais desenvolvimentos com a ontologia DOLCE acompanhamos na subseção 3.4).

2.5 O editor de ontologias Protégé-OWL⁸⁰

De uso gratuito e irrestrito⁸¹, o **Protégé** consagra-se como a **melhor ferramenta** para a criação, edição e visualização de ontologias (PRÉVOT; BORGIO; OLTRAMARI, 2010, p.197), podendo modelá-las no formato de frames (**Protégé-Frames**) ou no formato da linguagem OWL (**Protégé-OWL**).

Além dessas características apontadas, esta é a principal razão para termos adotado esse editor⁸² para a implementação do constructo “ontológico”: permitir, como demonstraremos na subseção 3.4, a adaptação de suas categorias formais (indivíduos,

⁸⁰ Linguagem OWL (*Web Ontology Language*): criada, no contexto da Web Semântica, pelo *World Wide Web Consortium* (W3C); a Web Semântica, por sua vez, busca maior eficiência no compartilhamento e na acessibilidade das informações veiculadas pela Internet.

⁸¹ É um *software* livre (*free source*) e de código aberto (*open source*).

⁸² Protégé-OWL - versão 3.3.1, de março de 2009. Disponível em <<http://protege.cim3.net/download/old-releases/3.3.1/>>. Acesso em: 25 maio de 2011.

propriedades e classes⁸³) para representar as categorias essenciais de um ontoléxico (itens lexicais, relações/propriedades e conceitos) através das seguintes correlações: **indivíduos ↔ itens lexicais, propriedades ↔ relações/propriedades e classes ↔ conceitos.**

O editor de ontologias Protégé-OWL, ao qual também podemos nos referir apenas como Protégé, ganha em funcionalidade quando acompanhado de *plug-ins*⁸⁴ como:

- O *plug-in* **TGVizTab**⁸⁵ – que permite a representação gráfica do que foi inserido no editor, possibilitando a visualização dinâmica da estruturação ontológica e facilitando, quando identificados problemas, atividades como a reanálise e a reestruturação de partes da ontologia.
- O *plug-in* **OntoLing** – que permite, conforme explicam Prévot, Borgo e Oltramari (2010), a população⁸⁶ de uma ontologia usando léxicos computacionais do tipo wordnet. Com uma interface amigável (*user-friendly*), o OntoLing “[...] também pode ser empregado na simples navegação no recurso lexical [...] Acreditamos que esse *plug-in* é uma ferramenta necessária para qualquer desenvolvedor que queira criar ontologias e léxicos de forma intuitiva.” (PRÉVOT; BORGIO; OLTRAMARI, 2010, p.197)⁸⁷.

No que diz respeito à edição de ontologias, é possível inserir no Protégé qualquer ontologia já construída⁸⁸ e com a qual se deseja trabalhar, modificando-a de acordo com propósitos específicos assentados no princípio da reutilização de recursos, a saber, a reutilização de conhecimento como exemplificado por Chishman (2009) e Pinto e Martins (2004) na subseção 2.3.1 e, no que diz respeito a *corpora*, por Almeida, Oliveira e Aluísio (2006) na mesma subseção (2.3).

⁸³ *Individuals, properties, e classes*, cada componente explicado com detalhes em Horridge (2004).

⁸⁴ Um *plug-in* é um programa auxiliar que disponibiliza funções adicionais a um programa principal. Ele é acionado concomitantemente ao programa principal e aparece como uma opção em um menu apropriado. (DOWNING et al., 2009, p.370)

⁸⁵ Mais detalhes sobre esse *plug-in* disponíveis em < <http://users.ecs.soton.ac.uk/ha/TGVizTab/>>. Acesso em: 04 jun. de 2011.

⁸⁶ No âmbito do desenvolvimento de ontologias, “popular” uma ontologia significa basicamente nela inserir determinados tipos de informação. A população de ontologias pode se dar de várias maneiras. Por exemplo, com o auxílio fornecido pela área da extração de informações (EI). Nédellec e Nazarenko (2005, p.155-156) discutem os benefícios da associação entre as áreas, dizendo que a EI tanto precisa de ontologias como parte do processo de extração de informações relevantes, quanto da extração, de textos, de conhecimentos novos a serem integrados às ontologias. As ontologias que relacionam os conhecimentos conceituais a suas realizações linguísticas podem, então, ser usadas pela EI e ser populadas pela EI.

⁸⁷ “[...] can also be employed for simple navigation in the lexical resource [...] We think that this *plug-in* is a necessary tool for any developer desiring to create ontologies and lexicons in an intuitive way.”

⁸⁸ Uma biblioteca de ontologias para o Protégé, nos mais diferentes formatos, está à disposição dos usuários em <http://protegewiki.stanford.edu/wiki/Protege_Ontology_Library>. Acesso em: 04 jun. de 2011.

As figuras de 7 a 11 resumem e ilustram os principais componentes do Protégé. A título de ilustração, carregamos no programa a ontologia DOLCE⁸⁹, que é a ontologia de nível superior que utilizamos como base nos desenvolvimentos da pesquisa.

Quando selecionada, conforme mostra a Figura 7, a aba ‘OWL Classes’ exibe uma representação hierárquica dos conceitos⁹⁰ da ontologia (‘SUBCLASS EXPLORER’, à esquerda) e um editor no qual os conceitos são editados (‘CLASS EDITOR’, à direita). Com o editor da direita é possível, por exemplo, adicionar comentários sobre a classe selecionada à esquerda. Na figura 7, para a classe ‘physical-realization’ a ontologia DOLCE traz o comentário (‘Value’) “Any physical particular that realizes a non-physical enduring [...]”⁹¹.

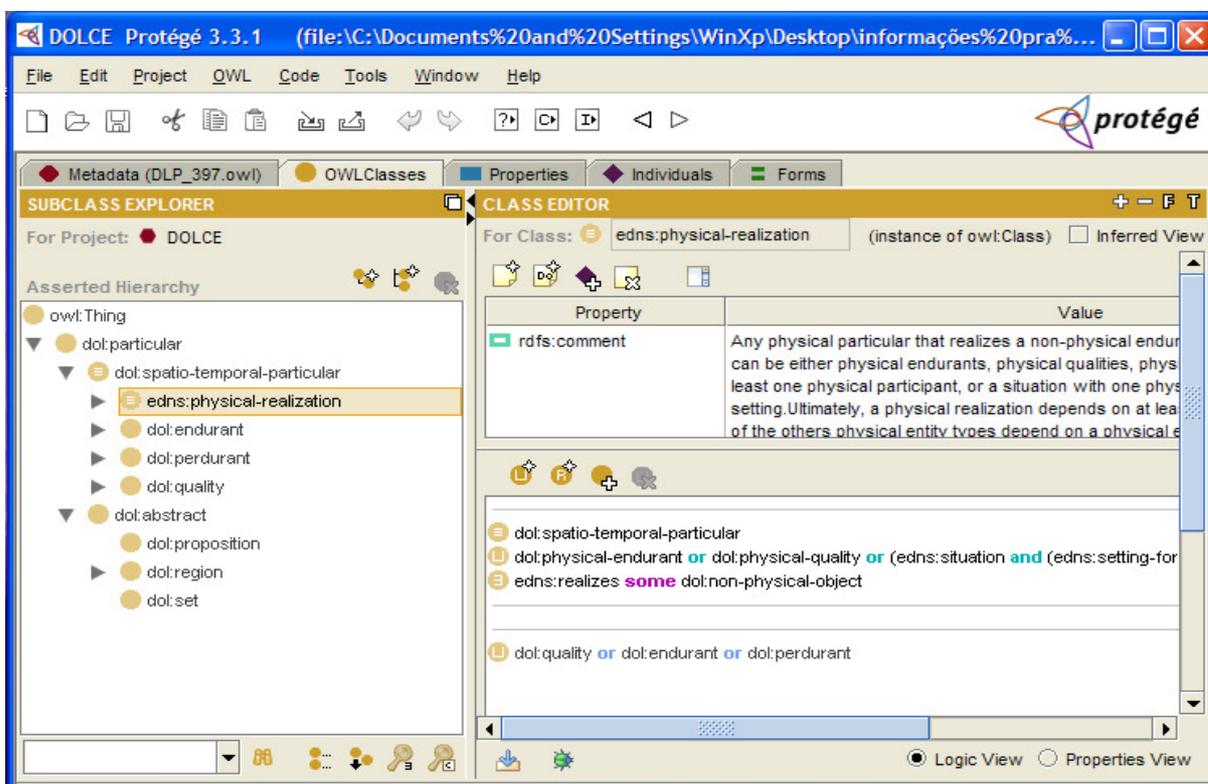


Figura 7. A Aba ‘OWL Classes’.

⁸⁹ A ontologia DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*), utilizada nesta pesquisa e discutida a partir da subseção 2.4, é uma ontologia de nível superior disponível para *download* no site do Laboratory for Applied Ontology: <<http://www.loa-cnr.it/DOLCE.html>>. Acesso em: 04 jun. de 2011.

⁹⁰ Quando falamos de ontologias, “conceitos” e “classes” (ou “categorias”) são termos intercambiáveis (HORRIDGE et al., 2004, p.14). Nesta dissertação, o que viemos chamando até então (e continuaremos a chamar) de “conceito” é uma categoria. Assim, na nossa terminologia: o conceito RICHELIEU (uma categoria, ou melhor, uma sub-categoria) faz parte do conceito TIPO DE BORDADO (outra categoria, ou melhor, uma supercategoria); com essa nota pretendemos deixar claro, então, que entendemos que conceitos e categorias denominam o mesmo tipo de construto teórico. Às vezes poderemos fazer esta distinção, conforme o grau de abrangência: **conceitos** (para sub-categorias / conceitos mais específicos) e **categorias** (para categorias / conceitos mais gerais).

⁹¹ Ao nos referirmos aos nomes das abas e a outros elementos do Protégé, procuraremos manter, entre aspas simples, a mesma forma escrita visualizada no editor: ‘OWL Classes’, ‘Value’, ‘CLASS EDITOR’, etc.

A Figura 8 mostra a aba ‘Properties’, selecionando a propriedade ‘part’, que, por sua vez, conecta classes da ontologia por meio da relação lógica ‘parte-de’.

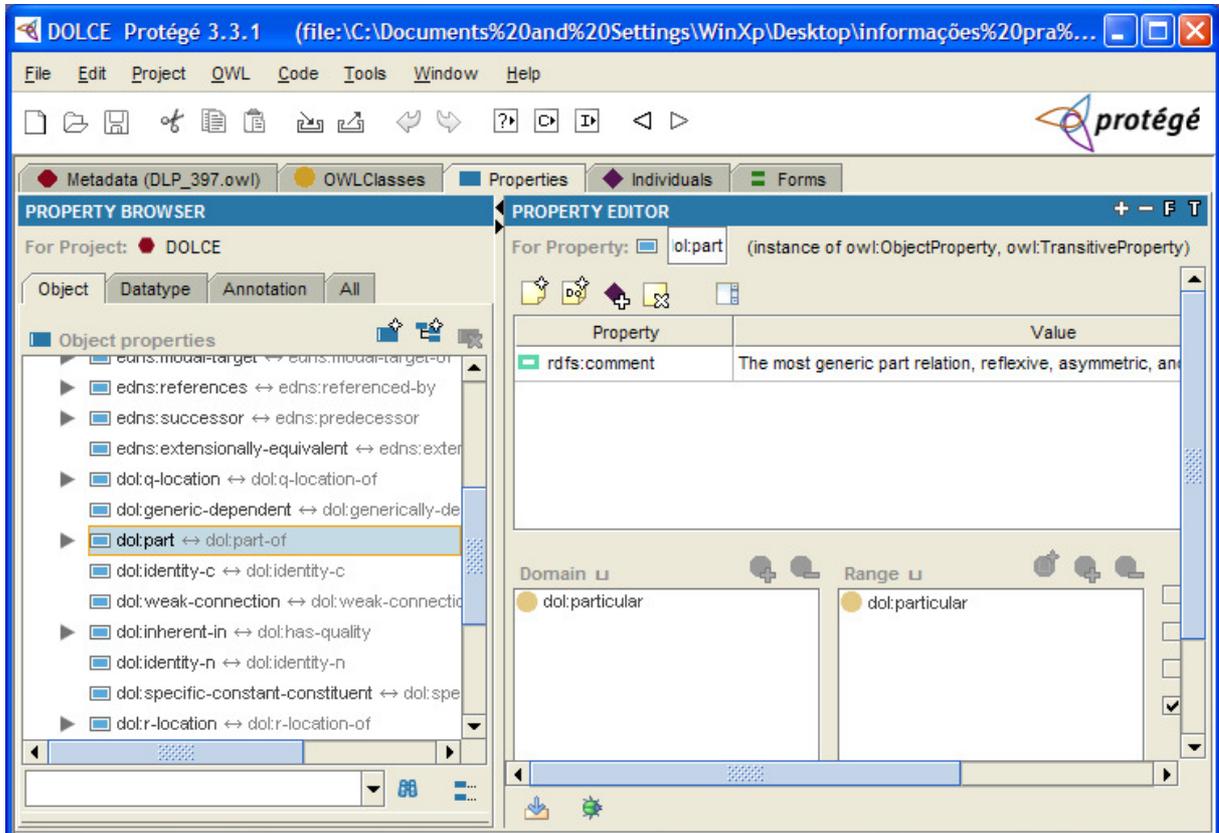


Figura 8. A Aba ‘Properties’.

Não exibimos a aba ‘Individuals’ porque a ontologia DOLCE, carregada (“crua”) como foi desenvolvida, não traz informações para esse componente, que deve, pois, ser populado de acordo com os objetivos do usuário. Passemos, então, para as abas dos *plug-ins* acionados no nosso trabalho com o editor de ontologias.

O primeiro *plug-in* é o OntoLing, com o qual trazemos léxicos computacionais, como wordnets e dicionários, para dentro do Protégé; a Figura 9 mostra a janela que seleciona o carregamento da versão WN 2.0 da WorNet de Princeton e a Figura 10 mostra, na parte esquerda, as informações e funcionalidades dessa wordnet já disponíveis para as atividades de consulta (da glosa, de sinônimos, etc.) ou de enriquecimento (semi-)automático da ontologia (que aparece na parte direita do editor).

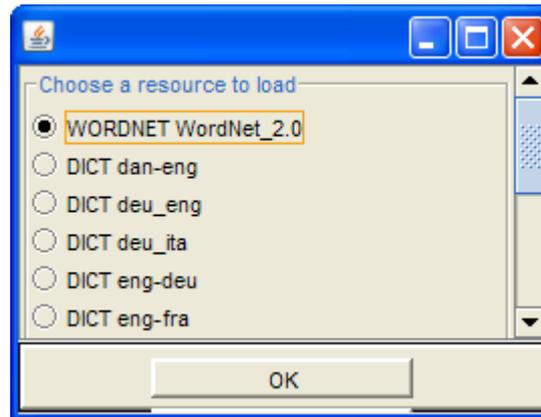


Figura 9. Léxicos computacionais estruturados e acessíveis através do *plug-in* OntoLing.

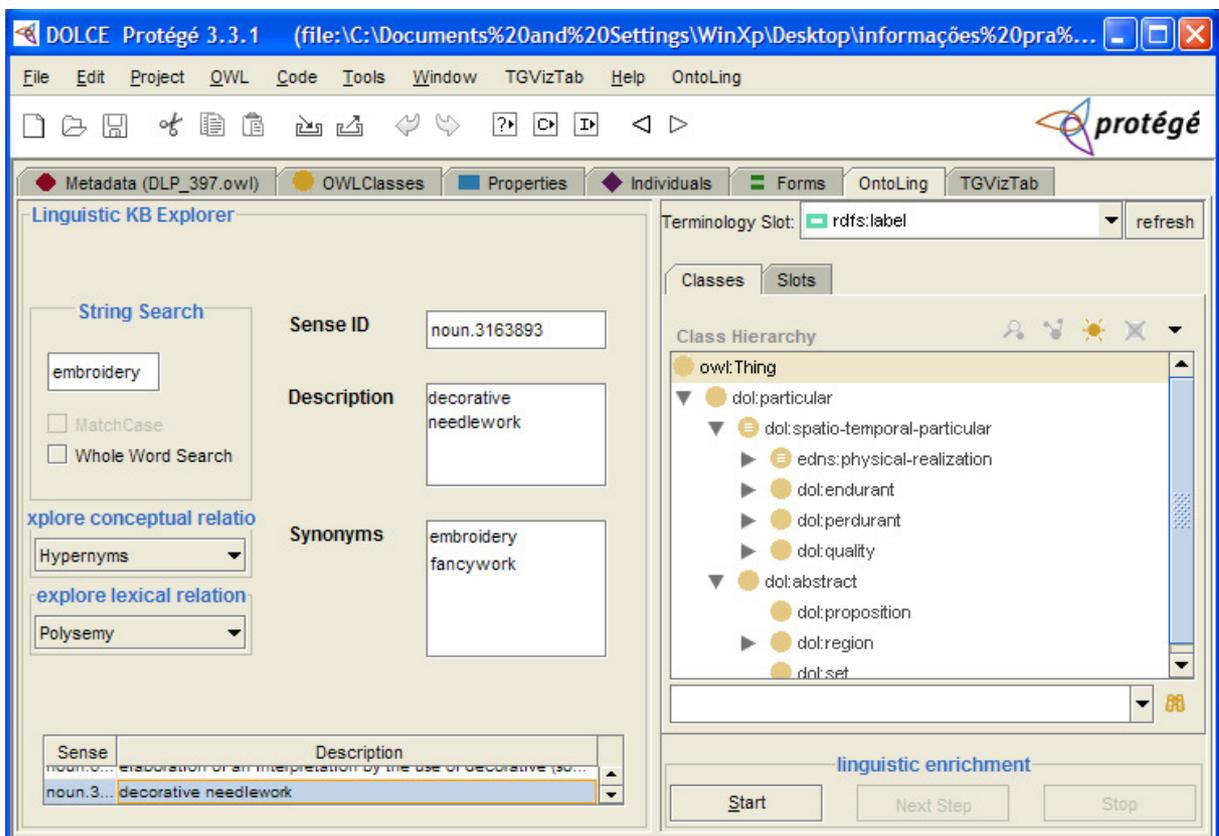


Figura 10. A versão WN 2.0 da WorNet de Princeton carregada pelo *plug-in* 'OntoLing'.

O segundo *plug-in* é o TGVizTab, que, conforme mostra a Figura 11, gera a representação em forma de gráfico da ontologia DOLCE. As classes ontológicas (os nós da rede gráfica) são conectadas entre si por meio das relações (os arcos da rede gráfica) estabelecidas no Protégé. Conforme seus interesses, o usuário pode alterar a forma como visualiza a ontologia usando o *plug-in* TGVizTab, pois é possível: (i) escolher quantos níveis da ontologia serão representados, (ii) focar partes da ontologia usando as barras de rolagem ou a opção 'zoom', (iii) modificar as cores de quaisquer elementos do gráfico, etc.

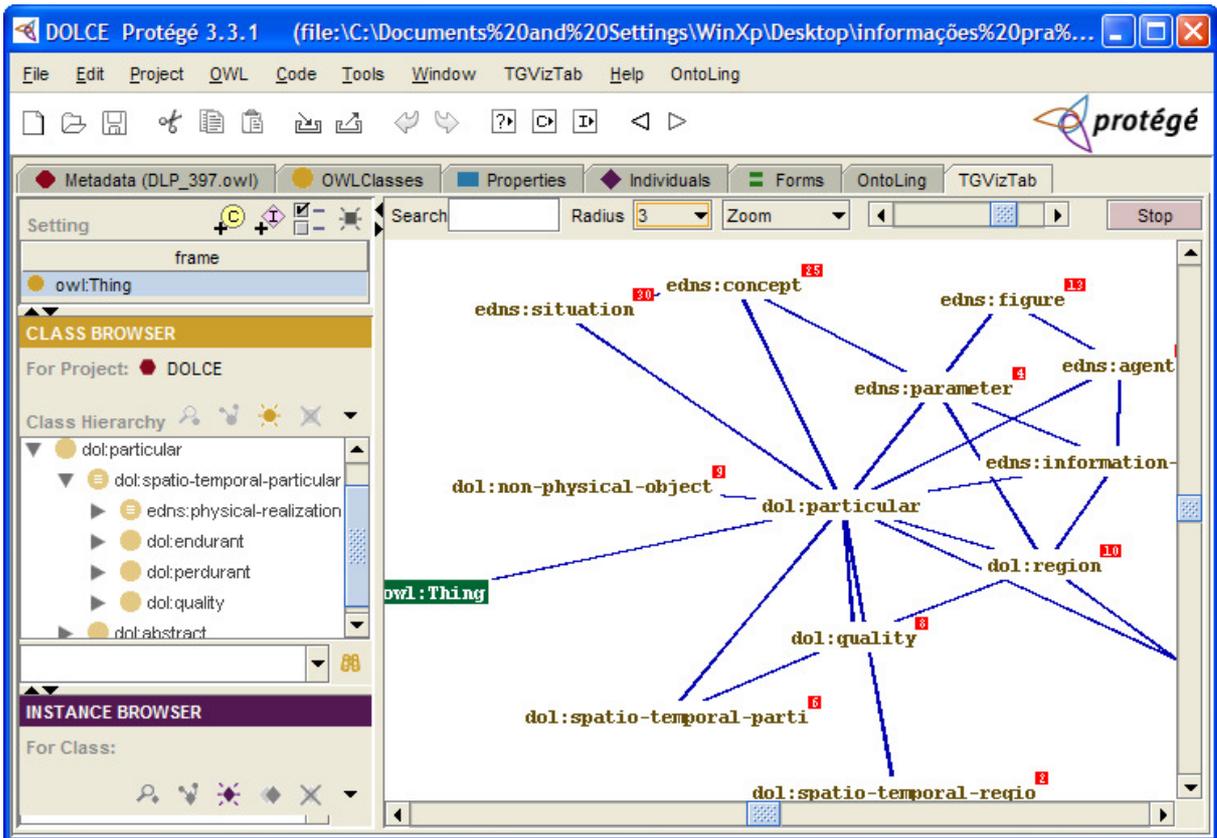


Figura 11. Representação gráfica da ontologia gerada pelo *plug-in* 'TGVizTab'.

Mais detalhes sobre o Protégé serão apreciados na subseção 3.4, na qual descreveremos como esse editor de ontologias é utilizado na concretização de parte significativa dos objetivos propostos nesta pesquisa.

3 CONSTRUÇÃO DO ONTOLÉXICO: SYNSETS, FRAMES E O PROTÉGÉ

Nesta seção, apresentaremos os desenvolvimentos para a realização dos objetivos da pesquisa, desmembrados, conforme já anunciamos, nestas atividades específicas, que às vezes se entrecruzam: identificação dos conceitos da ontologia em construção concomitantemente à coleta dos itens lexicais do português e do inglês (cf. Ia, IIa’); descrição dos conceitos em termos das glosas e das relações a serem estabelecidas entre eles (cf. Ib); e estruturação dos itens lexicais por meio da ancoragem deles na ontologia, isto é, do estabelecimento da correlação: itens(s) léxico(s) ↔ conceito da ontologia (cf. IIc), e alinhamento entre os fragmentos dos dois léxicos, português e inglês, produzindo assim um ontoléxico bilíngue do domínio léxico-conceitual IB(I) (cf. IId).

Na subseção 3.1, mostraremos a construção de synsets como forma de representação e registro das informações relevantes para o ontoléxico, sendo essas informações identificadas e analisadas levando-se em consideração as formas paradigmática e sintagmática de organização semântica, conforme acompanharemos em 3.2, na qual também apresentaremos uma sistematização, agora mais detalhada, dos principais conceitos (isto é, das categorias) da ontologia IB(I). Em 3.3, investigaremos como recursos podem ser utilizados, com as devidas correções e adaptações, na implementação do ontoléxico que estamos construindo. Finalmente, em 3.4, implementamos um fragmento do ontoléxico.

3.1 Sistematização preliminar

Informações – que eventualmente embasarão e integrarão o ontoléxico IBI – em coleta no *corpus* da pesquisa, com base no ferramental e na metodologia descritos ao longo da seção 2, estão previamente analisadas e formalizadas em termos de synsets. A opção por esse tipo de sistematização é duplamente motivada: além de o synset já ser um constructo, com o qual estamos bem familiarizados, eficaz para o **registro das informações**, ele também é uma peça-chave para o desenvolvimento de wordnets, ou seja, ao construirmos synsets para o domínio IB(I), contribuímos também para a **ampliação da base WN.Br** e, por conseguinte, com o seu alinhamento a outras wordnets, em especial à WN.Pr.

Apresentamos, a seguir, um exercício preliminar de sistematização de itens lexicais e conceitos em termos de synsets. Para isso, partimos da organização de itens lexicais (substantivos) coletados manualmente no “Jornal da Cidade”. Dos 39 artigos, selecionamos os itens lexicais dispostos em ordem alfabética no Quadro 5 e, conforme ocorrem nos textos, sem lematização.

acabamento; algodão; almofadas; aplicações; Arranjo Produtivo Local (APL) de cama, mesa e banho e enxoval de bebê; arremate dos bordados; artesã; artesanato; Associação do Comércio Ambulante em Produtos Artesanais ou Semi-Industrializados; aventais; babado; bastidores de madeira; bichinhos; bichos de pelúcia; bolsas; bonés; bordadeira; bordadeira digital; bordadeira eletrônica; bordadeira tradicional; bordado; bordado artesanal; bordado inglês; bordado vazado; bordados contínuos; bordados industriais; camisetas; camisões; capa para galão de água; capas de almofadas; Capital do Bordado; Capital Nacional do Bordado; centro de mesa; cobertor de micro fibra; coberta; cobertor; cobre-leitos / cobre leito; cobre leito casal; colchas; colcha de algodão; colcha de matelassê; confecção industrial; confecção; conjuntos de cama, mesa e banho; corte; cortina; costura; costureira; criação; crivo; crochês; desenho; designer; dobra das peças; edredom / edredom; enxovais; enxovais de bebês; enxovais de noivas; esquema de impressão tipográfica; estamparias; fabricação; fábrica de bordado; fábrica de enxovais; FAI; FEBI; feira; feira da praça; feira de negócios; feira de sábado; Feira do Artesanato de Ibitinga; Feira do Bordado; Feira do Bordado de Ibitinga; feira popular; feirinha; finalização; fio; fio ecológico; fronhas; guardanapos; Ibitinga; imagem; indústria do bordado; indústria familiar; jogo de camisola; jogos com sete panos de prato; jogos de cama; jogos de lençóis; jogos de toalha; kit casal; lençóis; lençol de fibra de bambu; lese; linha de poliéster; linhas de costura; malha com fibra de garrafa pet; manta de pé; mantas; mantas decorativas; mantas em matelassê; máquina de costura / máquina de costurar; máquina industrial; máquina pantográfica; máquina semi-industrial; máquinas computadorizadas; máquinas de bordar; máquinas de bordar automáticas; máquinas de pedal; máquinas de produção em série; máquinas elétricas; maquininha; maquininha de bordar; matérias-primas; modelos (de bordado); Operação Bordado; pano de prato; paramentos religiosos; peça artesanal; peça semi-artesanal; peças em fibra de bambu; peças; peso para porta; pijamas; pontos; porta-sacos; porta-travesseiros; produção; produtos; proprietários das fábricas; quadros; rendas; richelieu; robes; roupa de cama; roupa para dormir; roupas; roupas para animais de estimação; roupinhas de bebê; salão; salão de bordado; software de criação de bordados; tapete; tecido decorado; tecido estampado; tecido liso; tecido matelado; tecidos; tecidos de algodão; tecidos de malha; tecidos matelados; tecnologias têxteis; toalha de mesa; toalhas; toalhas de banho; toalhinhos; travesseiros; tricô; viscose; zigui-zague

Quadro 5: Itens lexicais coletados do “Jornal da Cidade”.

Podemos agrupar os itens lexicais do Quadro 5 segundo os tipos semânticos da rede WN.Pr, conforme estes exemplos mostram:

- <noun.act> = *acabamento, corte, Operação Bordado*
- <noun.artifact> = *linha, máquina industrial, feira da praça*
- <noun.group> = *Associação do Comércio Ambulante em Produtos Artesanais ou Semi-Industrializados*
- <noun.location> = *Capital Nacional do Bordado, Ibitinga*
- <noun.person> = *bordadeira, costureira, designer*

Disponer os itens lexicais em tipos semânticos, de acordo com a proposta da rede WN.Pr, já é realizar uma pré-categorização; identificar os tipos semânticos dos itens lexicais coletados como exemplificamos aqui ou herdá-los através do alinhamento de synsets nos permite reunir informações valiosas para a determinação das categorias ontológicas do domínio léxico-conceitual da IBI.

Como synsets são constituídos de itens lexicais sinônimos, estes synsets, por exemplo, apontam para conceitos que cabe a nós especificar:

1. {*máquina de costura, maquininha*}
2. {*bichinho, bicho de pelúcia*}
3. {*feira da praça, feira popular, feira de sábado, feirinha*}

O conceito representado pelo synset construído no exemplo 1 ocorre na rede WN.Pr, o que permite o alinhamento entre os synsets para as duas línguas: {*máquina de costura, maquininha*} *EQ_SYNONYM* {*sewing machine*}; o mesmo se dá no exemplo 2: {*bichinho, bicho de pelúcia*} *EQ_SYNONYM* {*teddy, teddy bear*}; no entanto, o exemplo 3 nos traz um conceito não presente na base da WN.Pr. O Quadro 6 ilustra nossa seleção na rede WN.Pr.

{03702760} <noun.artifact> open-air market, open-air marketplace, market square -- (a public marketplace where food and merchandise is sold)

=> {03201450} <noun.artifact> farmer's market, green market, greenmarket -- (an open-air marketplace for farm products)

=> {04098160} <noun.artifact> souk -- (an open-air market in an Arabian city)

Quadro 6. Synsets selecionados da rede WN.Pr para análise de alinhamento.

O synset {*souk*} também expressa um conceito bem particular e insere-se na base como hipônimo do synset que pode nos oferecer uma solução de alinhamento indireto para o exemplo 3: {*feira da praça, feira popular, feira de sábado, feirinha*} *EQ_HAS_HYPERONYM* {*open-air market, open-air marketplace, market square*}.

Exemplos de synsets construídos para o domínio e alinhados à WN.Pr podem ser observados no Apêndice 2.

3.2 Ampliação da sistematização

O Quadro 7 sintetiza um exercício que fizemos para explorar uma forma de estruturação do ontoléxico do domínio léxico-conceitual da IBI. Nosso ponto de partida foram informações que extraímos e analisamos dos recursos lexicogramaticalmente organizados do *corpus*, ou seja, das redes WN.Pr e FrameNet.

Coluna 1	Coluna 2	Coluna 3
FrameNet	WN.Br	WN.Pr
FES ⁹²	Synsets do português	Synsets do inglês
<i>Creator</i>	{bordador}	{embroiderer}
<i>Instrument</i>	{agulha, agulha de bordar}	{darning needle, embroidery needle}
<i>Artifact</i>	{bordado}	{embroidery}
<i>Type</i>	↓ {bordado ponto-cruz}	↓ {cross-stitch}
<i>Location_of_representation</i>	{almofada}	{cushion}
<i>Type</i>	⋮ ↓ {almofadrom}	⋮ ↙

Quadro 7. Exercício exploratório de construção do ontoléxico do domínio léxico-conceitual da IBI.

Informações extraídas e analisadas a partir da rede WN.Pr e WN.Br → os synsets na mesma linha nas colunas 2 e 3, por estarem alinhados de modo direto, conectam-se pela relação *EQ_SYNONYM*, ou seja, {bordador} *EQ_SYNONYM* {embroiderer}, {agulha, agulha de bordar} *EQ_SYNONYM* {darning needle, embroidery needle} e assim sucessivamente. A

⁹² Traduzimos, da FrameNet, a definição de cada um dos FEs que usamos na exemplificação: *Creator* = indivíduo ou indivíduos que trazem o Artefato à existência, *Instrument* = a entidade que o Criador usa para criar a peça de arte, *Artifact* = FE que identifica a entidade ou substância que é projetada para ter um Uso particular, *Type* = uma indicação do subtipo do Artefato e *Location_of_representation* = o local físico que sustenta o bordado (isto é, o suporte do bordado).

relação de hiponímia está representada pelas setas cheias. Por exemplo, em {bordado} EQ_HAS_HYPONYM {bordado ponto-cruz}, o conceito expresso no primeiro synset é mais genérico que o expresso no segundo. A relação de sinonímia aproximada está representada pelas setas tracejadas, já que, devido a uma lacuna no léxico do inglês, dois synsets do português alinham-se a um único synset do inglês: {almofada} EQ_NEAR_SYNONYM {almofadrom} e {cushion} EQ_NEAR_SYNONYM {almofadrom}.

Os conceitos estão representados em forma de synsets (cada item lexical do synset pode ser uma LU do frame em que se insere), e esses estão conectados por meio das relações semânticas previstas nas duas bases:

Informações extraídas e analisadas a partir da rede FrameNet → a coluna 1 enumera os FEs (*Frame Elements*) de dois frames, o *Physical_artworks* e o *Create_physical_artwork*, em que o primeiro é usado pelo segundo, devido à especificação da relação entre frames “é usado por” (*Is Used By*). Esses FEs fornecem rica informação contextual para os itens lexicais do ontoléxico, posto que indicam as relações temáticas entre os participantes da cena descrita pelo frame, a saber, a criação de um bordado em uma almofada ou em um almofadrom: o bordador (*Creator*), usando uma agulha (*Instrument*), faz um bordado (*Artifact*) da especialidade ponto-cruz (*Type*) em uma almofada ou em um almofadrom (*Location_of_representation*).

Esse exercício exploratório exemplifica bem a complementaridade entre as relações sintagmáticas (de redes como a FrameNet) e paradigmáticas (de redes como a WordNet), além de motivar a investigação aprofundada das relações sintagmáticas e do potencial que elas têm para, estruturadas nos moldes da FrameNet, auxiliar na categorização ontológica do domínio IBI e no estabelecimento de relações entre categorias e conceitos. Essa investigação é o nosso foco nos parágrafos que seguem.

De acordo com o que já indicamos na seção 2.1.5.2 e mostramos no Quadro 7, estamos atentos às informações contidas na FrameNet que podem contribuir para nossos objetivos; até agora citamos dois frames, *Create_physical_artwork* e *Physical_artworks*, bem como algumas relações entre eles e alguns FEs que eles abrangem, e é partindo de ambos que buscamos outros frames, e seus elementos afins, que também auxiliem na obtenção de novas informações e na proposição de um frame essencial, o *Bordar*⁹³, para o domínio conceitual que escolhemos.

⁹³ Seguindo o padrão de outros frames da rede, como *Hunting* ou *Killing*, o frame *Bordar* poderia ser nomeado em inglês como *Embroidering*.

Dissemos que há uma relação entre os frames *Create_physical_artwork* e *Physical_artworks*, que, por sua vez, relacionam-se também ao frame *Create_representation*, conforme ilustramos na esquematização da Figura 14.

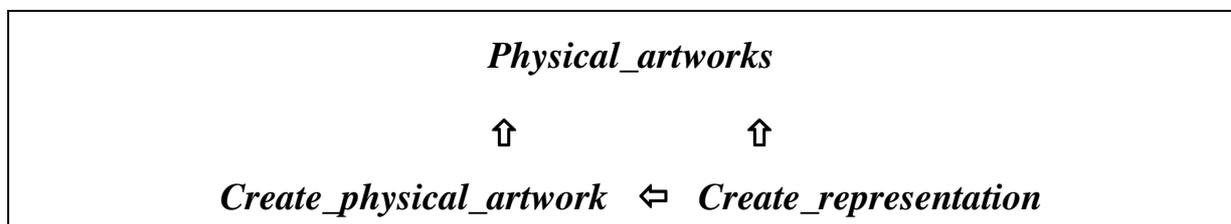


Figura 14: Relações entre os frames; as setas ⇔ e ↑ indicam a relação *uses* (utiliza).

A esses três frames, aos quais se associam LUs como *photograph.v*, *sculpt.v* e *drawing.n*, podemos relacionar o conceito BORDADO ARTEFATO (ou BORDADO), já que tratamos do bordado enquanto item físico criado. Com a delimitação do conceito de BORDADO localizamos e desenvolvemos, paralelamente, o conceito relativo ao bordado enquanto atividade, a esse conceito reservamos as notações BORDADO ATIVIDADE ou BORDAR. Pesquisas no *corpus* proporcionaram a montagem do Quadro 8, que traz definições, de dicionários variados (do português e do inglês), para os itens lexicais *bordado* e *embroidery*.

BORDADO ARTEFATO	BORDADO ATIVIDADE
<ul style="list-style-type: none"> • obra de bordadura • ornato assim feito • ornamentation of fabric with needlework • a design of coloured stitches on cloth • a design or decoration formed by or as if by embroidery • an object decorated with embroidery • embroidered work or fabric; ornamental needlework 	<ul style="list-style-type: none"> • ato de bordar ('ornamentar com fios') • the act or art of embroidering • the activity of decorating cloth with coloured stitches • the art or process of forming decorative designs with hand or machine needlework • the art or work of ornamenting fabric with needlework; embroidering
<ul style="list-style-type: none"> • trabalho sobre tecido (ou outro material), feito à mão ou à máquina, em que se criam figuras ou ornatos com fios de diferentes tipos, introduzidos por meio de agulha, às vezes com aplicação de outros materiais, como miçangas, lantejoulas, pérolas, pedras etc. • lavor em relevo, feito na roupa, a linha, fio de lã, prata, ouro etc. • ação ou resultado de bordar • trabalho feito em tecido ou tela, ger. à mão, utilizando agulha e linhas ou fios coloridas para criar ornatos em relevo • the art or pastime of embroidering cloth 	

Quadro 8. Definições dicionarísticas relativas aos conceitos BORDADO ARTEFATO e BORDADO ATIVIDADE.

Em se tratando de atividades (BORDAR), em especial aquelas de extensão industrial como o ramo dos bordados, somos levados ao frame *Craft*, que, por sua vez, engloba (relação entre frames *uses*) o frame *Fields*. Informações relativas a esses dois frames, juntamente com os três já citados, foram retiradas da FrameNet e agrupadas no Quadro 9, colocado no final desta subseção.

De acordo com a recém-explorada sistematização da rede FrameNet e com os dados da rede WN.Pr, deduzimos que:

- o conceito “BORDADO ARTEFATO” relaciona-se
 - ao FE *artifact* do frame *Physical_artworks* e
 - ao *synset {03673332} <noun.artifact> needlework, needlecraft -- (a creation created or assembled by needle and thread)* da WN 2.0; e que
- o conceito “BORDADO ATIVIDADE” relaciona-se
 - ao FE *activity* do frame *Craft* e
 - ao *synset {00855701} <noun.act> needlework, needlecraft -- (work (such as sewing or embroidery) that is done with a needle)* da WN 2.0.

A ação criadora representada nos frames *Create_representation* e *Create_physical_artwork* e o que está envolvido com o que é criado (*Physical_artworks*) são destacados em trechos de textos coletados no *corpus*. Assinalamos essas informações em conformidade com os FEs⁹⁴ desses três frames que se mostram mais significativos para o domínio léxico-conceitual focado:

1. Para fazer o richelieu [*Type*] eu [*Creator*] tenho que fazer na maquininha [*Instrument*].
2. Vai colocando linha [*Material*] conforme as cores, a flor [*Represented*], por exemplo, vermelha, a cor que quiser.
3. O que eu [*Creator*] bordo é só no linho [*Location_of_representation*].

⁹⁴ Traduzimos, da rede FrameNet, a definição de cada um dos FEs que usamos nessa exemplificação: *Material* = qualquer indicação do que constitui o Artefato, incluindo componentes, ingredientes, etc., *Reference* = representação do Representado que o Criador usa como referência na criação de sua obra de arte e *Represented* = a entidade – que pode ser uma coisa, uma ação ou um estado – que é representada. As demais definições já foram explicadas na nota de rodapé n° 92.

4. A bordadora [*Creator*] foi procurar por toda a parte aquela rosa branca [*Reference*] perfeitíssima, em cuja semelhança bordasse a sua.
5. Neste ponto, o bordado [*Artifact*] passa para outro setor da indústria, a arrematação, onde pessoas munidas de pequenas tesouras [*Instrument*] se encarregam de cortar as pontas de fiões [*Material*] excedentes.
6. Em uma fábrica, para cada riscadeira [*Creator*] eram necessárias cinco bordadeiras [*Creator*] para que ninguém ficasse parado.

Assinalamos ainda estas outras informações, agora levando em consideração FEs⁹⁵ dos frames *Craft* e *Fields*:

7. O bordado Madeira [*Type*], verdadeira obra de arte, testemunha a sensibilidade e o saber dos artesãos [*Practitioner*] de uma ilha do Atlântico – Madeira [*Culture/Place*].
8. Os seus bordados [*Salient_entity*] se tornaram conhecidos e, com os negócios prosperando, Dioguina passou a contratar outras bordadeiras [*Practitioner*].

Demonstraremos, a seguir, como as informações provenientes da rede FrameNet podem ser estudadas, paralelamente à análise atenta do *corpus*, de forma a determinar e desenvolver as categorias ontológicas⁹⁶ a partir dos quais estruturaremos a ontologia do domínio IBI.

⁹⁵ Traduzimos, da rede FrameNet, a definição de cada um dos FEs que usamos nessa exemplificação: *Activity* = esse FE marca expressões que indicam a Atividade que é executada habitualmente ou convencionalmente, *Place* = esse FE marca expressões que indicam o local onde a Atividade é praticada, *Practitioner* = esse FE marca expressões que indicam de quem é a Atividade em questão, *Salient_entity* = entidade física ou abstrata que é caracteristicamente associada com a Atividade e *Type* = subcampo de Atividade ao qual o Praticante está engajado.

⁹⁶ Lembrando que para nós as categorias referem-se aos conjuntos de subcategorias, sendo que as subcategorias são denominadas conceitos.

FRAMES	DEFINIÇÃO	UNIDADES LEXICAIS	ELEMENTOS DO FRAME	
			NUCLEARES	NÃO NUCLEARES
<i>Create_ representation</i>	A Creator produces a physical object which is to serve as a Representation of an actual or imagined entity or event, the Represented.	carve.v, cast.v, draw.v, paint.v, photograph.v, sketch.v	Creator, Represented	Depictive, Depictive_of_represented, Instrument, Iteration, Manner, Material, Means, Place, Purpose, Reference, Representation, Role, Time
<i>Create_ physical_ artwork</i>	A Creator creates an artifact that is typically an iconic Representation of an actual or imagined entity or event. The Representation may also be evocative of an idea while not based on resemblance.	artist.n, cast.v, draw.v, paint.v, sculpt.v, take_((picture)).v	Creator, Representation	Depictive, Descriptor, Explanation, Instrument, Location_of_representation, Manner, Means, Purpose, Reference, Role, Time
<i>Physical_ artworks</i>	A physical object, the Artifact, is produced by a Creator to stimulate the perceptions, emotions, or cognition of an audience. Though it need not, in many cases, the created Artifact is intended to represent an actual or imagined entity or event, the Represented. The Represented may be specifically shown in a particular state (State_of_represented).	bronze.n, diagram.n, drawing.n, image.n, painting.n, photo.n, photograph.n, picture.n, poster.n, sculpture.n, statue.n, statuette.n, trace.n	Artifact	Creator, Descriptor, Material, Name, Place, Represented, State_of_represented, Time_of_creation, Type, Use
<i>Craft</i>	A particular Activity, performed conventionally or habitually by more than one Practitioner within a Culture, is described in terms of its method.	art.n, craft.n, science.n	Activity, Culture, Practitioner	Descriptor, Place
<i>Fields</i>	An individual or group (Practitioner) or a part of their work life (Work) is defined professionally by the Activity that they typically engage in. Alternatively, the Activity may be indicated by a Salient_entity conventionally associated with it.	area.n, art.n, artistic.a, field.n, history.n, in.prep, industrial.a, industry.n, literary.a, literature.n, mathematics.n, sector.n, theater.n	Activity, Practitioner, Salient_entity, Work	Place, Time, Type

Quadro 9. Informações de frames para o desenvolvimento do ontológico do domínio léxico-conceitual da IBI.

3.3 Consolidação da sistematização

O Quadro 10 agrupa os FEs anteriormente selecionados e introduz algumas categorias ontológicas por eles motivadas.

FRAMES		CATEGORIAS ONTOLÓGICAS POTENCIAIS
(i) <i>Create_representation</i> (ii) <i>Create_physical_artwork</i> (iii) <i>Physical_artworks</i>	(iv) <i>Craft</i> (v) <i>Fields</i>	
FEs (Frame Elements)		
Artifact	Salient_entity	BORDADO
Creator	Practitioner	PROFISSIONAL
Instrument		INSTRUMENTO
	Activity	BORDAR
Location_of_representation		SUPORTE
	Place	LOCAL
Material		MATERIAL
	Type	TIPO DO BORDAR
Reference		REFERÊNCIA
Represented		REPRESENTADO
Type		TIPO DOS BORDADOS

Quadro 10. Categorias ontológicas propostas a partir da análise dos FEs.

Entendemos que os FEs do Quadro 10 são os constituintes básicos dos frames *Bordar1* e *Bordar2*, e, para eles, propomos as seguintes descrições:

→ Frame *Bordar1*

Um profissional [*Creator*] produz, com determinado material [*Material*], o bordado [*Artifact*] em um dado suporte [*Location_of_representation*], usando para isso um instrumento [*Instrument*]. Esse bordado pode ser de um tipo específico [*Type*] e visa a representar algo [*Representation*] a partir de uma referência [*Reference*].

→ Frame *Bordar2*

Profissionais [*Practitioner*], ligados a dada localidade [*Place*], envolvem-se em uma atividade [*Activity*] que pode ser de um tipo específico [*Type*] e que tem como produto peças que chamamos de bordado [*Salient_entity*].

Bordar1 e *Bordar2* são frames que apresentam conceitos que se entrecruzam e na maioria das vezes são complementares, já que, de acordo com o que verificamos, não é raro os discursos tratarem simultaneamente do bordado ação criativa e do bordado ocupação, o que fica explicitado, por exemplo, com o paralelismo entre *Artifact* e *Salient_entity* (3ª linha do Quadro 10) e entre *Creator* e *Practitioner* (4ª linha); desse modo, consideramos que os conceitos evocados por esses dois frames relacionam-se à mesma atividade, cujo conceito rotulamos BORDAR, e o foco de nosso trabalho é o produto dessa atividade, o bordado (que, conforme já anunciamos, tem um conceito por nós representado por BORDADO).

Não vemos necessidade de especificar quais FEs são nucleares e quais são não nucleares, pois constatamos que todos têm algum grau de relevância para a caracterização dos conceitos da ontologia e, principalmente, do seu conceito-chave, o BORDADO. É importante explicar que entendemos caracterizar como estabelecer, dentro da ontologia, relações entre os conceitos e seus atributos, sendo assim, por exemplo, um conceito como RICHELIEU, que se refere a um tipo de bordado, caracteriza-se pela combinação de diversos atributos descritos na ontologia. Algumas das categorias propostas no Quadro 10, que se envolverão na caracterização da ontologia IB(I), são apresentadas a seguir (A a I), juntamente com dois outros elementos caracterizadores presentes nos frames:

A TIPO DE BORDADO: identifica os tipos do bordado enquanto um artefato.

B PROFISSIONAL DO BORDADO: identifica os profissionais envolvidos, direta ou indiretamente, com o bordado. Por exemplo: a bordadeira, o arrematador, etc.

C INSTRUMENTO: identifica os instrumentos envolvidos, direta ou indiretamente, com o bordado. Por exemplo: máquina de bordar, bastidor, etc.

D TIPO DE SUPORTE: identifica quais os tipos de suportes que sustentam o bordado. Por exemplo: linho, étamine, etc.

E LOCAL: identifica os locais de referência para o bordado. Por exemplo: bordado Madeira, bordado francês, etc.

F MATERIAL DO BORDADO: identifica os tipos de materiais com os quais são feitos o bordado. Por exemplo: linhas de algodão, fios de ouro, etc.

G ETAPA DE CONFECÇÃO: identifica as atividades envolvidas, direta ou indiretamente, com o bordado. Por exemplo: bordar, lavar, etc.

H MODALIDADE DE CONFECÇÃO: identifica as modalidades do bordado. Por exemplo: manual, mecânico, etc.

I COR: identifica as cores predominantes do bordado. Por exemplo: azul, preto, etc.

- **Referência:** identifica o que é referido no bordado. Por exemplo: uma paisagem, a memória, etc.
- **Representado:** identifica o que está representado no bordado. Por exemplo: uma figura geométrica, uma ave, etc.

Baseados, então, nas investigações de várias naturezas a partir do *corpus*, nos procedimentos metodológicos e nas teorias descritas nas seções anteriores e nos exercícios exploratórios até então exemplificados, pudemos propor estas categorias ontológicas, sobre as quais fazemos algumas considerações:

- como bordar envolve outras atividades paralelas e complementares, seja no tratamento do próprio bordado ou de seu suporte, identificamos também a categoria ETAPA, que reúne os conceitos relativos a essas atividades, tanto os identificados no *corpus* lexicalizados na forma de verbos (*riscar, lavar, etc.*) quanto na forma de substantivos (*riscadura, lavagem, etc.*); tal categoria, como ficou demonstrado, não é prevista pelos frames selecionados e sua identificação foi feita a partir de várias informações coletadas no *corpus*, como se vê, por exemplo, na frase-exemplo número 5, na qual o conceito ARREMATAR se lexicaliza (*arrematação*);
- *Referência* e *Representado* são FEs que destacamos durante a análise das informações provenientes da FrameNet e a análise das categorias em potencial para nossa ontologia; eles se aplicam bem enquanto elementos caracterizadores dos frames, mas não enquanto categorias ontológicas essenciais para nossa ontologia (talvez para alguma outra ontologia mais genérica sobre representação em geral), apontamos suas ocorrências em frases do *corpus* a título de exemplificação e para mostrar que esse, a representação, é um dos pontos nos quais a ontologia do domínio IB(I) poderia dialogar com outras ontologias;
- por ser esta uma pesquisa de fundamentação do ontolético, por ora não objetivamos certos detalhamentos da estrutura ontológica do nosso constructo, por exemplo: além do diálogo, citado no item anterior, motivado pelos elementos *Referência* e *Representado*, seria possível também integrar informações de outras ontologias à nossa no que diz respeito à especificação da matéria-prima para os suportes e materiais do bordado (algodão, linho e outras substâncias sendo associadas a linhas e tecidos,

por exemplo) ou à especificação das confecções que produzidas com os suportes bordados (com o étamine bordado se faz toalhas ou com a cambraia bordada se faz almofadas, por exemplo), sendo que em qualquer um dos casos estaríamos criando mais oportunidades para o desenvolvimento de um ontoléxico bem rico que contemplasse conceitos e itens lexicais peculiares como os que viemos apresentando (SEMANTINHA/*semaninha*, ALMOFADROM/*almofadrom*, etc.).

O domínio léxico-conceitual da IBI, que, na ontologia, está sendo estruturado em termos das categorias ontológicas destacadas, abarca conceitos dentre os quais apenas uma minoria lhe é particular, e certamente deixa de abarcar tantos outros que fazem parte do domínio léxico-conceitual mais geral da IB. Informações de natureza lexical e conceitual sobre no domínio léxico-conceitual da IB(I), que atesta a seleção das categorias e irão popular o ontoléxico, puderam ser obtidas no *corpus* conforme exemplificam os trechos que seguem:

→ Em sua dissertação sobre os bordados portugueses, SILVA (2006) dá uma definição de bordado:

Podemos considerar “bordado” todos os trabalhos decorativos exercidos por meio de uma agulha sobre qualquer tipo de suporte pré-existente. Os tipos de suporte podem variar em conformidade com o tipo de trabalho que se pretenda obter, sendo o mais comum o tecido; não podemos esquecer os trabalhos lindíssimos, por exemplo, das chinelas dos trajés à Vianesa, entre outros. (p.8).

A partir dessa definição atestamos as categorias ontológicas que destacamos. Por exemplo: “exercidos por meio de agulha” permite apreender **C** e “sobre qualquer tipo de suporte pré-existente”, **D**.

→ CINTRÃO (1990), quando se refere às atividades relacionadas ao bordado em Ibitinga, também oferece informações sobre as categorias ontológicas do domínio: “[...] muitos homens que exerciam profissões que demandavam maior esforço físico passaram a se dedicar ao bordado. Alguns na produção direta e outros, indiretamente, prestando serviços de lavanderia, desfiadura, riscadura.” (p.30) evoca a categoria **G**, e “É importante ressaltar que as atividades paralelas, relacionadas ao bordado, compõem

o total da força-de-trabalho. São elas: riscadeira, cortadora de bicos, costureira, passadeira, embaladeira e vendedora [...]” (p.31) evoca a categoria **B**.

→ Baseados na *Grande enciclopédia Delta Larousse* (1970), formamos o Quadro 11, que detalha as categorias **A**, **C**, **D**, **F** e **H** e itens lexicais nelas ancorados (cf. Ia, IIa’).

CATEGORIA ONTOLÓGICA	EXEMPLOS DE ITENS LEXICAIS
A	<i>bordado renascença, bordado de imaginação, Saint-Gall, bordado Cornely</i>
C	<i>bastidor, quadro, tear industrial, tear suíço</i>
D	<i>linho, seda, cambraia, morim, tecido fino, pano grosseiro, tecido transparente</i>
F	<i>seda, lã, fio de metal</i>
H	<i>bordado de agulha, bordado a máquina, bordado mecânico</i>

Quadro 11. Categorias ontológicas e itens lexicais extraídos da ‘*Grande enciclopédia Delta Larousse*’.

→ O *Macmillan English dictionary for advanced learners* (2007), além de ser um tradicional dicionário legível com computador, que nos dá acesso a vários itens lexicais, apresenta também um pequeno thesaurus, no qual, através de busca, pela chave ‘embroidery’, por exemplo, atestam-se pelo menos duas das categorias ontológicas: o domínio rotulado ‘Sewing and Needlework’ engloba as expressões ‘To Sew, or to Sew Something’, com verbos relacionados às ações de costurar, além de ‘Things Used for Sewing and Needlework’ (**C**) e ‘People Who Sew’ (**B**).

→ Na fábula de Fernando Pessoa (1986, p.69) também destacamos algumas categorias e, a título de exemplificação, os FEs *Referência* e *Representado*:

A uma bordadora [**B**] dum país longínquo foi encomendado pela sua rainha que bordasse, sobre seda [**D**] ou cetim [**D**], entre folhas [*Representado*], uma rosa branca [*Representado*]. [...] Gastou dias sobre dias, chorosas horas, buscando a rosa [*Referência*] que imitasse com seda [**F**] [...] não tendo melhor remédio, bordou de memória [*Referência*] a rosa [*Representado*] que lhe haviam exigido.

→ Das definições das entradas de dicionários também podemos extrair importantes informações. Por exemplo, no *Random House unabridged electronic dictionary* (1996), *black work* = “embroidery done with black, usually silk, thread on white

fabric, esp. linen” e *crewelwork* = “decorative embroidery done with crewel yarn on cotton or linen, using simple stitches traditionally worked in floral or pastoral designs” e, no *Dicionário eletrônico Houaiss da língua portuguesa* (HOUAISS, 2001), *galão* = “tira de tecido bordado com fios de ouro, prata, seda, algodão etc.” e *rendendê* = “tipo de bordado sobre linho, com pontos cheios e abertos, formando desenhos geométricos” permitem atestarmos as categorias ontológicas **A**, **D** e **F** e o FE *Representado* (“floral/pastoral designs” e “desenhos geométricos”).

Essas categorias discutidas nos últimos parágrafos são as que mais apropriadamente abarcam os conceitos mais relevantes no domínio léxico-conceitual da IB(I). Guiamo-nos, então, por essas categorias centrais para buscarmos novas informações, por exemplo: por meio do método de geração de ontologia *middle-out* (DING e FOO, 2002) é possível identificarmos conceitos mais específicos e mais gerais que nos ajudam na delimitação do nível de domínio e do nível superior da ontologia. Como auxílio a essa tarefa de delimitação recorreremos a uma ontologia pronta disponível na Web, a ontologia DOLCE, que é trabalhada com o editor Protégé, conforme acompanhamos na subseção seguinte (3.4).

3.4 Implementação

Com base nas informações provenientes do *corpus*, parte delas registradas em synsets (como vimos na subseção 3.1), um esboço da ontologia do domínio IBI foi traçado levando-se em consideração principalmente os conceitos mais importantes identificados, as relações entre eles embasadas na rede FrameNet, a estruturação hierárquica da rede WN.Pr e a ontologia geral DOLCE, que, como ontologia de nível superior, permite integrar a nossa ontologia a uma ontologia mais geral.

Levando em conta que a rede WN.Pr traz consigo uma ontologia subjacente (FELLBAUM, 1998, citada em 2.1.5.1) podemos, feitas as devidas análises e adaptações, aproveitar sua estruturação, que foi, em parte, trilhada e registrada durante as atividades de construção de synsets para os domínios IB(I). Retomemos, por exemplo, o arquivo do synset {*bordador*} (Figura 2, p.23): a cadeia de synsets hiperônimos de {*embroiderer*} nos leva do conceito BORDADOR até o conceito mais geral ENTIDADE, conforme acompanhamos na

sequência (1) que segue (a seta \rightarrow significa ‘é-um-tipo-de’ na relação conceitual de subsunção presente no esquema).

(1) BORDADOR \rightarrow OPERADOR DE AGULHA \rightarrow FABRICANTE \rightarrow CRIADOR
 \rightarrow PESSOA \rightarrow ORGANISMO \rightarrow COISA VIVA \rightarrow OBJETO \rightarrow ENTIDADE

Comparemos (1) a outras duas cadeias de conceitos (2) e (3), respectivamente presentes nos synsets *{finisher}* e *{cutter}* da rede WN.Pr:

(2) ARREMATADOR \rightarrow TRABALHADOR \rightarrow PESSOA \rightarrow ORGANISMO \rightarrow
 COISA VIVA \rightarrow OBJETO \rightarrow ENTIDADE

(3) CORTADOR \rightarrow TRABALHADOR TREINADO \rightarrow TRABALHADOR \rightarrow
 PESSOA \rightarrow ORGANISMO \rightarrow COISA VIVA \rightarrow OBJETO \rightarrow ENTIDADE

Observamos que, a partir de PESSOA, a organização dos conceitos mais específicos necessita passar por uma reformulação, pois, apesar de TRABALHADOR TREINADO não constar de (2), um arrematador também precisa de treino e um bordador é, afinal, um trabalhador, mesmo sem o conceito TRABALHADOR estar presente em (1). Sob a luz dos conhecimentos provenientes principalmente do *corpus* e fornecidos por informantes da área, analisar como a estrutura da ontologia presente na rede WN.Pr pode ser rearranjada faz parte da investigação que culmina na proposição de uma ontologia mais precisa para o domínio IB(I). Para as cadeias (1), (2) e (3), chegamos à estruturação da Figura 15.

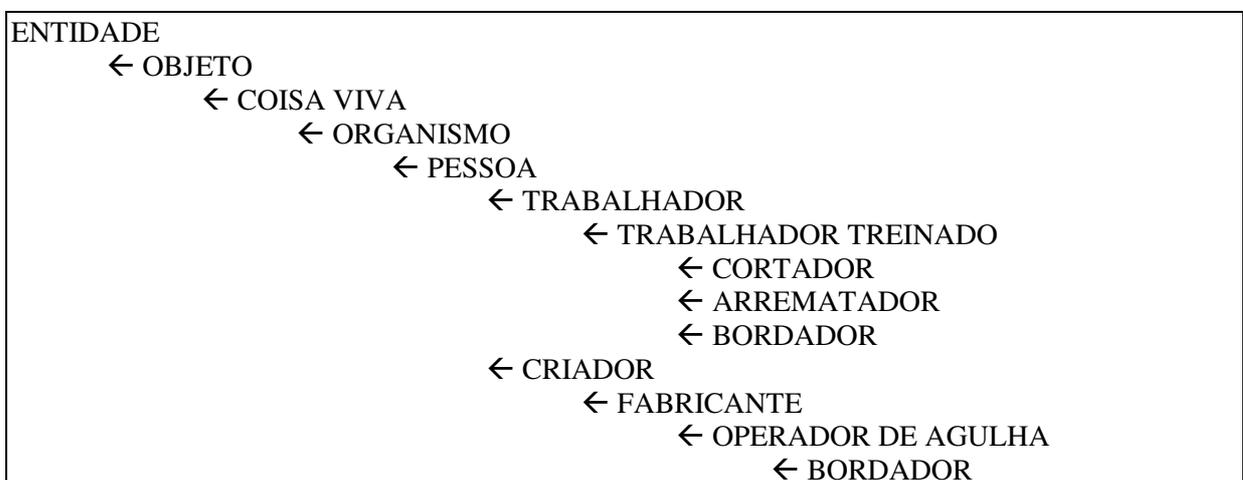


Figura 15: Reestruturação de parte da ontologia da rede WN.Pr (a seta \leftarrow indica ‘é-um-tipo-de’).

Como vemos na Figura 15, à estrutura foram especificadas as informações de que um arrematador é um profissional treinado e, mais importante, de que um bordador é tanto um criador quanto um trabalhador⁹⁷: ao propormos que BORDADOR tem dois “pais” (dois hiperônimos distintos) na nossa ontologia, estamos contornando uma lacuna (lexical)⁹⁸ da base da rede WN.Pr.

Esse tipo de análise dos últimos parágrafos, envolvendo identificação e “correção” de lacunas, por exemplo, gera e aperfeiçoa conhecimentos e métodos para nosso trabalho, em especial no que diz respeito à organização dos conceitos nas estruturas motivadas pela rede WN.Pr, que, afinal, estarão conectadas à ontologia DOLCE, que, por sua vez, tem parte da sua estrutura reproduzida na Figura 16: do lado esquerdo, a imagem da ontologia carregada no Protégé; do direito, uma esquematização paralela com nossas sugestões de tradução para os rótulos das categorias.

Observamos que a categoria mais geral da ontologia é denominada PARTICULAR (ou ENTIDADES PARTICULARES na tradução). Logo, a ela pertencerão todas as entidades do domínio a ser descrito. De acordo com o detalhamento estipulado por seus desenvolvedores, as categorias conceituais da ontologia DOLCE subdivide-se: primeiro, em quatro grandes categorias (COISAS, EVENTUALIDADES, ATRIBUTOS e ABSTRAÇÕES); em seguida, em categorias cada vez mais específicas. Conforme for necessário, no decorrer da subseção, forneceremos informações sobre as categorias com as quais trabalhamos diretamente; para detalhes sobre como a ontologia DOLCE foi desenvolvida, como e por que ela abarca tais categorias, indicamos Gangemi et al. (2010).

⁹⁷ Uma vez fixada essa estrutura torna-se possível dar encaminhamento a outros conceitos relacionados à categoria PROFISSIONAL. Por exemplo, LAVADOR é inserido na hierarquia como conceito mais específico de TRABALHADOR, mas não de CRIADOR.

⁹⁸ Trabalhos anteriores já discutiram os tipos de lacunas e as dificuldades por elas postas durante o alinhamento entre wordnets (MARCELLINO; DIAS-DA-SILVA, 2007; MARCELLINO, 2008; MARCELLINO; DIAS-DA-SILVA, 2009).

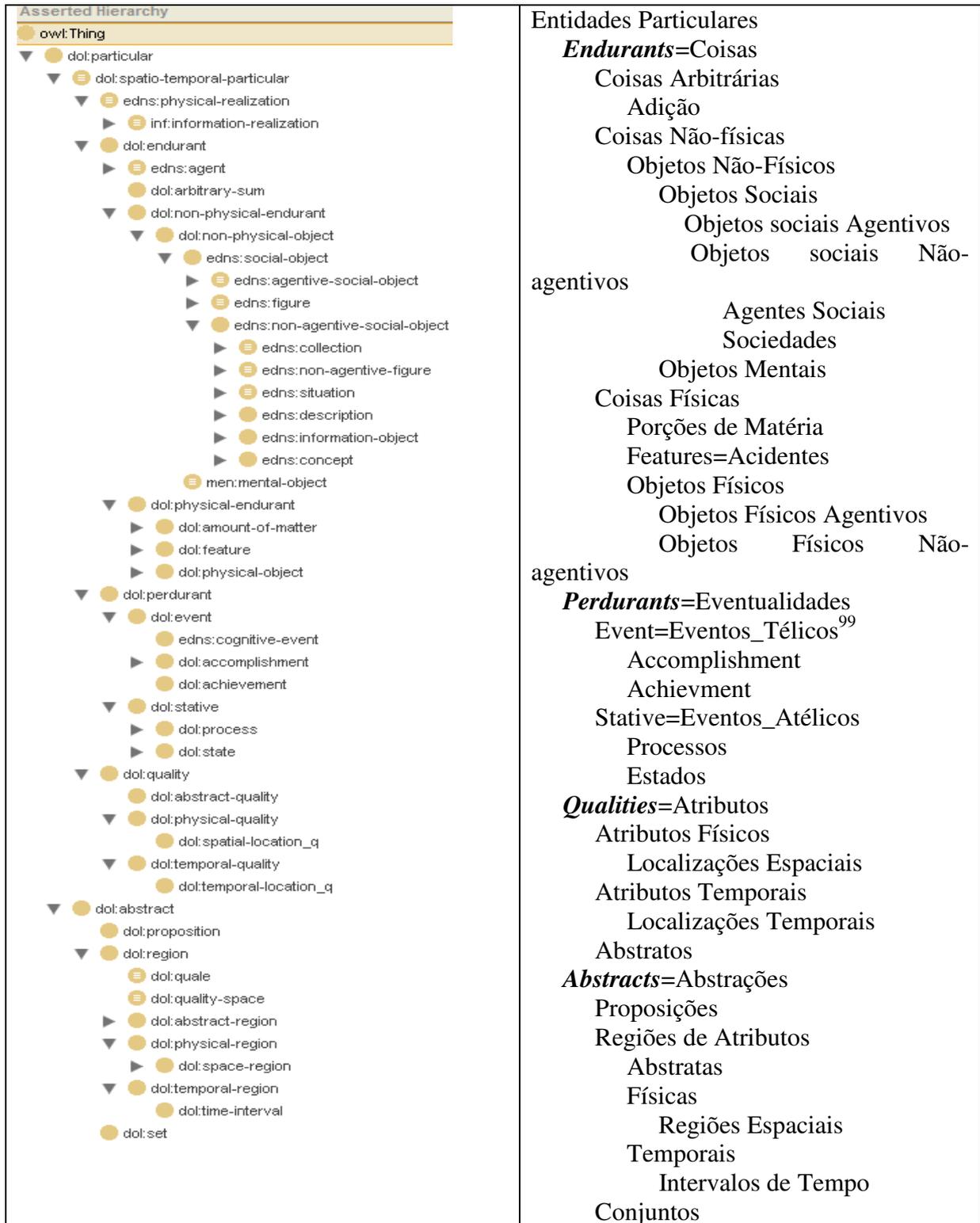


Figura 16. Categorias da ontologia de nível superior DOLCE.

É importante que tenhamos, como um de nossos embasamentos, uma ontologia geral de domínio como a ontologia DOLCE, que se afina com a rede WN.Pr, pois, partir do zero¹⁰⁰,

⁹⁹ A caracterização télico/atélico refere-se à in/determinação do evento: um evento télico tem início ou fim bem determinados, um atélico não.

certamente é um procedimento mais demorado e menos afim com o princípio da reutilização; para utilizá-la produtivamente no Protégé, no entanto, vamos estipular alterações necessárias e com o propósito de facilitar a implementação da contrapartida linguística (dos itens lexicais), sem a qual não seria possível constituirmos um ontoléxico. Essas alterações, motivadas pelo resultado eficiente da metodologia apresentada em Di Felippo (2008), são referentes, pois, aos usos destes componentes do Protégé:

- A aba ‘Individuals’, originalmente designada para a representação dos indivíduos, ou seja, para a instanciação dos objetos do domínio de interesse da ontologia, abrigará os itens lexicais que lexicalizam os conceitos da nossa ontologia¹⁰¹;
- A aba ‘Properties’, originalmente designada para a representação das relações entre os indivíduos da ontologia, abrigará as relações entre os conceitos da nossa ontologia;
- A aba ‘OWL Classes’, originalmente designada para representar classes (no sentido de conjunto de indivíduos) da ontologia, abrigará as categorias, ou seja, os conceitos da nossa ontologia.

Em analogia à Figura 17, que mostra classes, indivíduos e propriedades (relações) de uma ontologia, construímos, na Figura 18, uma representação ilustrativa de parte do ontoléxico do domínio léxico-conceitual da IB(I). Os grandes círculos, os pequenos quadriláteros e as setas, na Figura 17, representam, respectivamente, nomes de classes (“person”), indivíduos pertencentes às classes (Matthew”) e relações entre classes (“hasPet”); na Figura 18, os grandes círculos, os pequenos quadriláteros e as setas, representam, respectivamente, conceitos (BORDADO_RUSSO), e lexicalizações desses conceitos (*Russian embroidery*; *bordado russo*) e relações entre conceitos (TCS).

¹⁰⁰ De acordo com Ding e Foo (2002), uma ontologia pode ser criada do zero (do inglês: *from scratch*), embora seja um trabalho altamente custoso, do ponto de vista dos recursos humanos e materiais.

¹⁰¹ Nas ontologias formais, os indivíduos instanciam classes (por exemplo, os indivíduos ‘Ibitinga’ e ‘Araraquara’ são instâncias da classe ‘Cidade’); no nosso trabalho, os itens lexicais instanciam conceitos (por exemplo, *bordado*, *embroidery* e *fancywork* são instâncias do conceito BORDADO). Vale, entretanto, lembrar que estamos diante de instanciações de naturezas diferentes, sendo que o segundo tipo, chamado de lexicalização, é o que particularmente nos interessa.

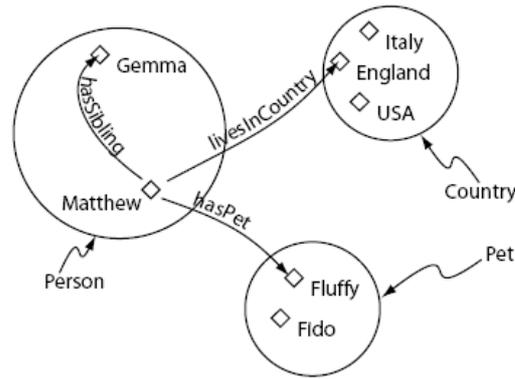


Figura 17. Esquematização de uma ontologia e dos seus componentes centrais (HORRIDGE et al., 2004, p.14).

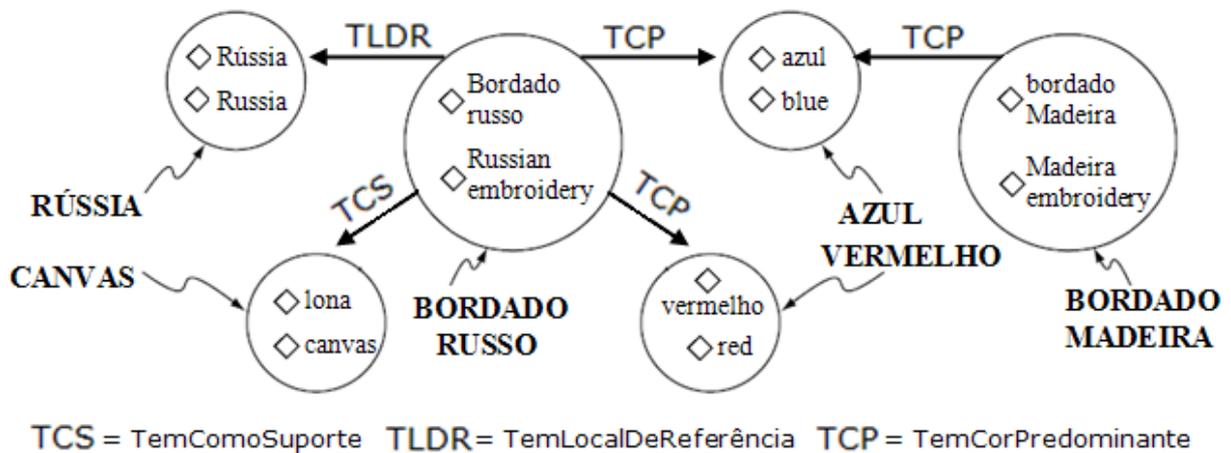


Figura 18. Esquematização de parte do ontológico domínio léxico-conceitual da IB(I) e dos seus componentes.

Passamos, a partir deste ponto, aos procedimentos práticos da implementação da ontologia no Protégé. Lembramos que a estruturação dos conceitos de nível superior da ontologia do domínio léxico-conceitual IB(I) é fornecida pela ontologia DOLCE e que o nível intermediário e parte do nível desse domínio estruturam-se a partir das análises que fizemos da rede WN.Pr, como a ilustrada na Figura 15 (p.76), sendo que novas análises podem se mostrar necessárias quando da conexão entre a ontologia DOLCE e a rede WN.Pr: Gangemi et al. (2010) atentam para o fato de a rede WN.Pr apresentar imprecisões estruturais que precisam ser avaliadas e corrigidas antes da sua utilização¹⁰². Algumas dessas correções foram feitas antes da implementação, outras foram demandadas no decorrer das atividades, já que, durante o processo de implementação, podemos fazer constatações e promover reanálises até

¹⁰² Por exemplo, no projeto OntoWordNet a rede WN.Pr foi replanejada como uma ontologia formal, seguindo-se basicamente (GANGEMI et al., 2010, p.50-52): a distinção entre synsets que podem ser formalizados como classes dos que podem ser formalizados como indivíduos, a interpretação das relações lexicais da rede WN.Pr como relações ontológicas e o alinhamento dos níveis superiores da rede WN.Pr à ontologia DOLCE (com reformulação da hiperonímia quando necessário).

então não possibilitadas, o que vem comprovar que trabalhar em um editor de ontologias é mais do que apenas implementar, é determinar/revisar conhecimentos.

Começamos a exemplificar como se dão as atividades de desenvolvimento do ontolético partindo de dois conceitos-chave: BORDADO e BORDAR.

Para o conceito BORDADO,

- a rede WN.Pr apresenta a seguinte hierarquia: BORDADO → TRABALHO DE AGULHA → CRIAÇÃO → ARTEFATO → OBJETO FÍSICO → ENTIDADE; e
- a ontologia DOLCE: ENTIDADE PARTICULAR ← COISA ← COISA FÍSICA ← OBJETO FÍSICO ← OBJETO FÍSICO NÃO-AGENTIVO ← ARTEFATO MATERIAL ← SISTEMA COMO ARTEFATO.

Como vemos, a ontologia DOLCE e a rede WN.Pr coincidem no conceito ARTEFATO, que é o nosso ponto de intersecção e cuja aplicação foi: BORDADO → TRABALHO DE AGULHA → CRIAÇÃO → ARTEFATO MATERIAL → OBJETO FÍSICO NÃO-AGENTIVO (...).

Apesar de a base da WN.Pr, em outra ocorrência de lacuna, não contemplar o conceito BORDAR (bordado enquanto atividade/ocupação), podemos prever que, se existisse, o synset <noun.act> {*embroidery2*} seria hipônimo de <noun.act> {*needlework, needlecraft*}, como mostra a Figura 19.

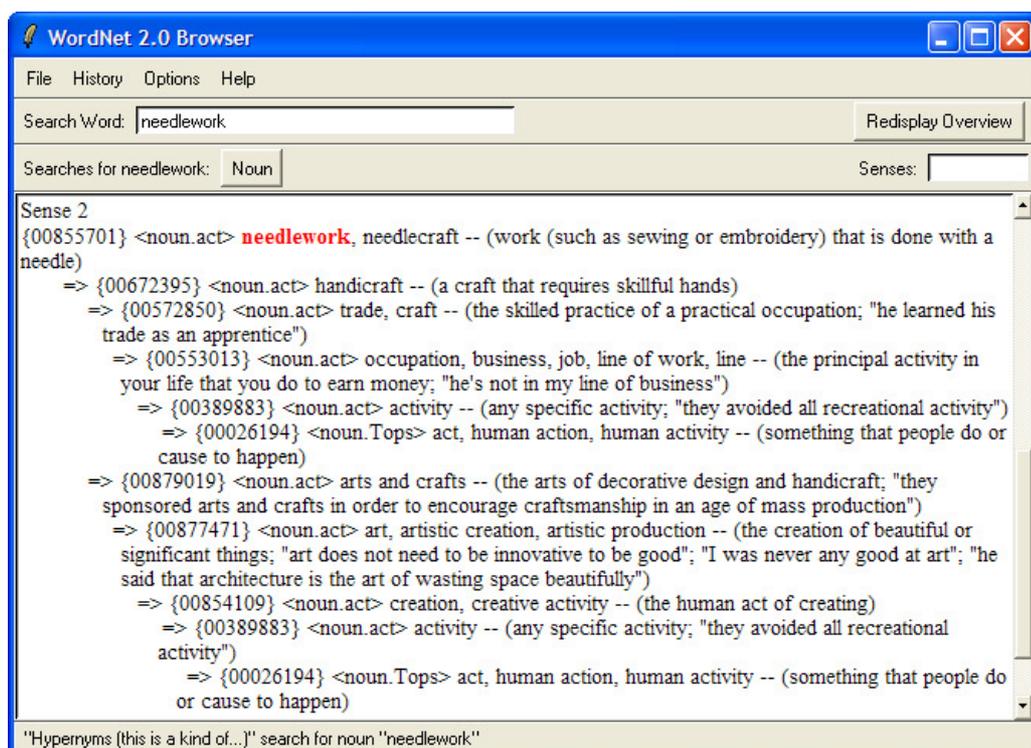


Figura 19. Cadeia de hiperônimos para o synset {*needlework, needlecraft*}.

Com a Figura 19, verificamos que o conceito TRABALHO DE AGULHA¹⁰³ tem relação com (‘é-um-tipo-de’) ARTESANATO, e este último, por sua vez, relaciona-se a duas diferentes cadeias de conceitos mais gerais, de modo que podemos enriquecer nossa ontologia ao especificarmos que o conceito de BORDAR refere-se tanto a uma atividade criativa quanto a uma ocupação, conforme constatamos na Figura 20, que ilustra também, no Protégé, a estruturação de BORDADO explicada anteriormente.



Figura 20. Encaixe dos conceitos BORDADO e BORDAR na ontologia geral DOLCE.

É importante notar que BORDAR relaciona-se ao conceito ATIVIDADE incluído sob EVENTO ATÉLICO (stative); embora a ontologia DOLCE já abarque, em sua estrutura, um conceito ATIVIDADE (EVENTUALIDADE \leftarrow EVENTO TÉLICO \leftarrow REALIZAÇÃO \leftarrow AÇÃO \leftarrow ATIVIDADE), ela contempla o aspecto télico das atividades, enquanto optamos por considerar que as atividades (bordar, a arrematação, etc.) são atélicas.

¹⁰³ A partir deste ponto, precisamos especificar, então, que temos também dois tipos de trabalhos de agulha, um que é um artefato (citado na análise de BORDADO) e outro que é uma atividade, logo, adotamos a notação numérica para esses casos: TRABALHO DE AGULHA1 e TRABALHO DE AGULHA2.

Uma vez encaixado o conceito na parte já estruturada da ontologia (conforme o previsto no objetivo Ia), passamos para a descrição dele (cf. Ib) e para as atividades de ancoragem (cf. II), que acompanhamos nos próximos parágrafos.

A Figura 21 ilustra nosso trabalho de edição da ontologia na aba ‘OWL Classes’ do editor Protégé ativa. Nela, temos:

- à esquerda, a estrutura hierárquica da ontologia na qual conferimos, por exemplo, que o conceito (ou subcategoria, como também podemos chamá-lo) LINHA DE SEDA relaciona-se (‘é-um-tipo-de’) com o conceito (ou categoria) LINHA, que, por sua vez, em níveis mais genéricos ainda, relaciona-se com a categoria ‘system-as-artifact’ da ontologia DOLCE.
- à direita, em ‘CLASS EDITOR’, há um campo que podemos usar para explicitar o conceito em termos de uma glosa, trata-se da propriedade ‘comment’, à qual podemos atribuir como valor (‘Value’) uma glosa que pode ser criada no momento da implementação com a ajuda do *plug-in* OntoLing (recurso extra do editor Protégé). Com esse *plug-in*, é possível consultar a WN.Pr, localizar o conceito expresso na forma de synset e copiar ou traduzir a sua glosa (nosso procedimento tem sido a tradução e/ou adaptação para o português da glosa do synset em inglês).

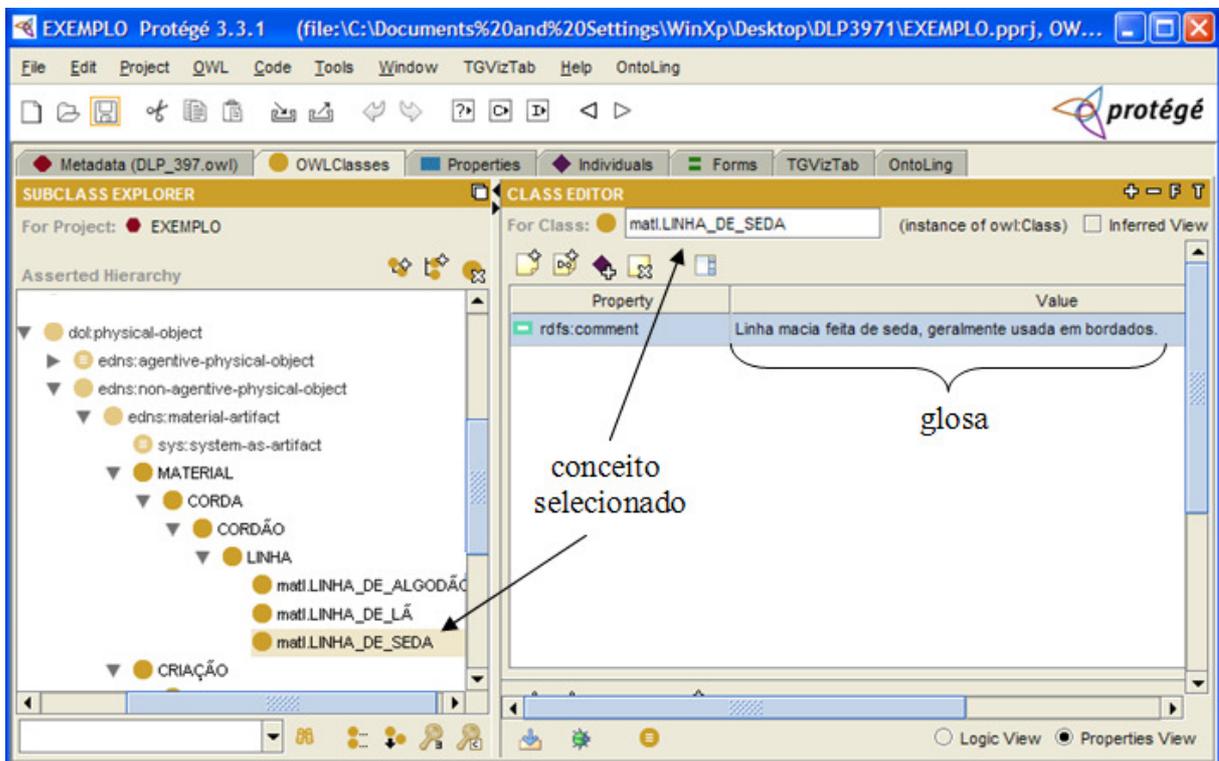


Figura 21. Especificação da glosa para um conceito (Aba ‘OWL Classes’).

A Figura 22 ilustra o trabalho na aba ‘Individuals’, que é subdividida em ‘INSTANCE BROWSER’, no centro, onde acrescentamos (opção ‘Create instance’) as lexicalizações do conceito selecionado na aba ‘OWL Classes’ (no exemplo, AGULHA DE BORDAR), e ‘INDIVIDUAL EDITOR’, do lado direito, onde registramos uma frase-exemplo, retirada do *corpus*, para cada item lexical (as frases-exemplos são os valores da propriedade ‘comment’) e explicitamos (em ‘Lang’) se são itens do português (PB, de português brasileiro) ou do inglês (EN, de *English*).

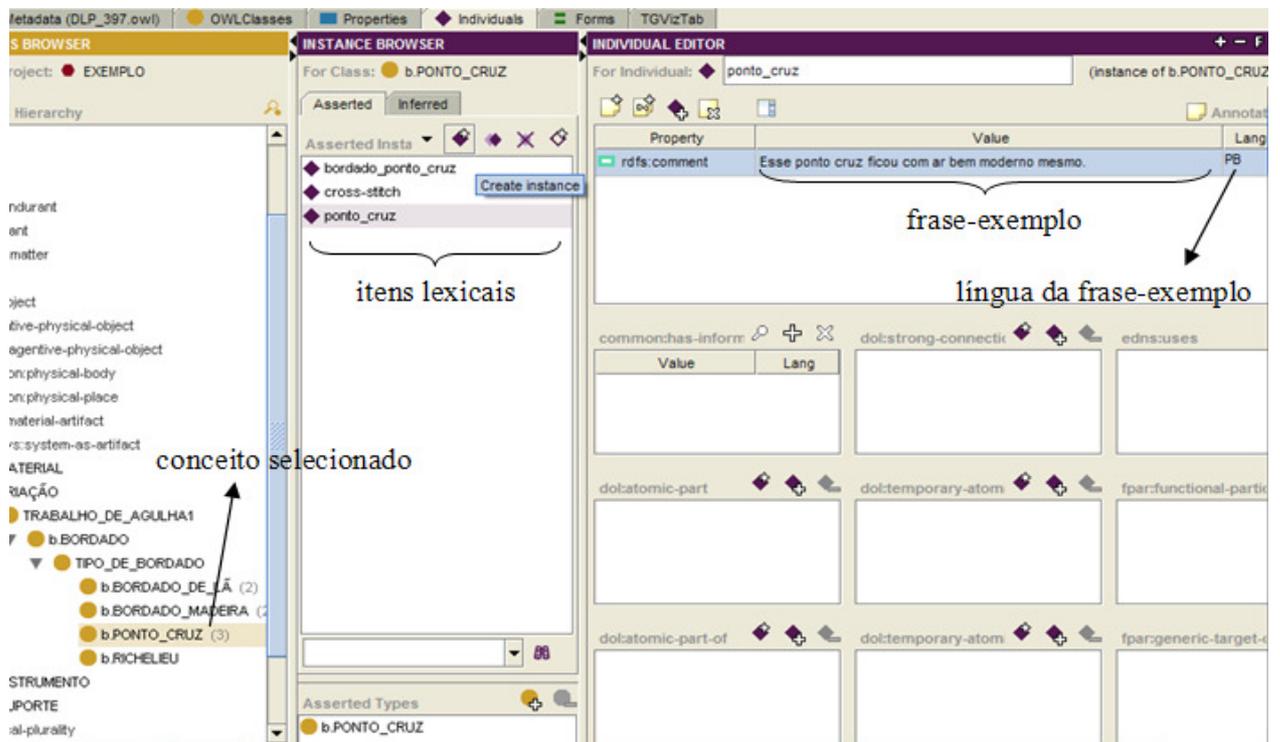


Figura 22. Especificação das lexicalizações, frases exemplo e idioma (Aba ‘Individuals’).

Notamos, nas Figuras 21 e 22, que há nomes de conceitos que vêm precedidos de notações¹⁰⁴, por exemplo: “matl.LINHA_DE_SEDA” (figura 21) e “b.PONTO_CRUZ” (Figura 22), nos quais “matl.” indica que esse conceito insere-se na categoria MATERIAL e “b.” (de bordado) indica que esse conceito é particular ao domínio IB(I), sendo que pode haver a associação de “b.” com outras indicações. O Quadro 12 sintetiza nossas convenções para as notações referentes às categorias.

¹⁰⁴ Uma vez que os conceitos estão etiquetados, fica fácil recuperar, usando as ferramentas de busca do Protégé ou de seu *plug-in* TGVizTab, todos aqueles que, por algum interesse, desejamos listar. Por exemplo, buscando por “b.” localizamos todos os conceitos que estão diretamente relacionados ao domínio de nosso estudo, estejam eles localizados em qualquer parte da ontologia.

Categoria	Notação	Exemplo
PROFISSIONAL DO BORDADO	prfs.	prof.b.BORDADOR
INSTRUMENTO	inst.	inst.TESOURA
TIPO DE SUPORTE	supr.	supr.CÂNHAMO
LOCAL	locl.	locl.ILHA_DA_MADEIRA
MATERIAL DO BORDADO	matl.	matl.LINHA_DE_OURO
ETAPA DE CONFECÇÃO	etap.	etap.ENGOMAR
MODALIDADE DE CONFECÇÃO	modl.	modl.MANUAL
COR	cor.	cor.BRANCO

Quadro 12. Notações referentes aos nomes das categorias do ontológico.

Descrevemos, agora, o processo de estabelecimento das relações entre conceitos. Para isso, tomamos, como exemplo, uma importante relação que identificamos e rotulamos de ‘éDaModalidade’. Sabemos que, por exemplo, o bordado ponto-cruz, segundo as nossas fontes de informação, só pode ser feito à mão, logo, faz parte da caracterização do conceito PONTO-CRUZ indicar que ele se relaciona a essa modalidade do bordar; uma vez que outros bordados podem se caracterizar por essa mesma modalidade ou por uma modalidade diferente, ou ainda por uma combinação delas, dizemos que o domínio da relação ‘éDaModalidade’ é TIPO_DE_BORDADO, e o seu contra-domínio MODALIDADE_DE_CONFECÇÃO.

A Figura 23 exemplifica a criação da relação ‘éDaModalidade’. Observamos que a aba ‘Properties’ do Protégé apresenta, à esquerda, um navegador (‘PROPERTY BROWSER’), no qual criamos a relação/propriedade (‘Create object property’), e, à direita dele, o editor da relação/propriedade (‘PROPERTY EDITOR’), onde inserimos o domínio (‘Domain’) e o contra-domínio (‘Range’) da relação/propriedade e onde fornecemos uma descrição informal para a relação/propriedade (no campo ‘Value’).

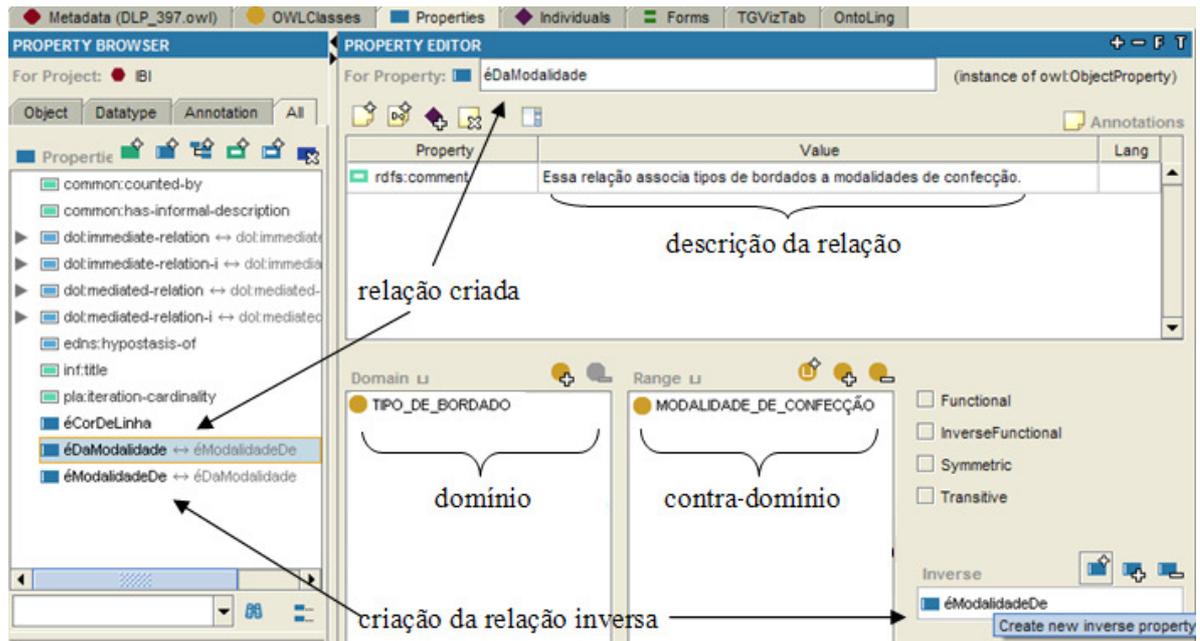


Figura 23. Implementação/edição das relações entre os conceitos.

É importante atentar para a necessidade de se criarem restrições que especifiquem exatamente quais são os caracterizadores do conceito. O conceito PONTO-CRUZ, além ser caracterizado como um elemento do **domínio** TIPO_DE_BORDADO, é também caracterizado com a seleção de uma série de restrições (propriedades). Selecionando-se esse conceito no editor, passa-se às especificações dessas restrições como, por exemplo, o conceito MANUAL, ilustrado na Figura 24. Nessa figura, está selecionada aba 'OWL Classes' do editor; na janela 'CLASS EDITOR', selecionamos a visualização das relações (no editor Protégé, denominadas "properties"), clicando em 'Properties View' (destacado no canto inferior direito da figura) referentes ao conceito (nesse exemplo, 'éDaModalidade') e clicando também no botão 'Create restriction' (também destacado na figura). A partir desse ponto, temos acesso à janela da Figura 25. Nela podemos indicar exatamente a quais elementos da categoria MODALIDADE_DE_CONFECÇÃO o conceito PONTO-CRUZ se relaciona: no exemplo, como vemos na figura, ao conceito PONTO-CRUZ relaciona-se apenas a restrição/propriedade, modl.MANUAL.

Outros tipos de restrições estão previstos no editor Protégé. Por exemplo, se um dado bordado pode ser confeccionado manualmente ou em máquinas, é possível selecionar, no contra-domínio da relação 'éDaModalidade', modl.MANUAL e modl.MECÂNICO, separando-os pelo restritor (conectivo lógico) 'or' (há também os restritores 'and' e 'not', para outras situações de restrições na caracterização dos conceitos, conforme pode ser acompanhado mais adiante na Figura 26).

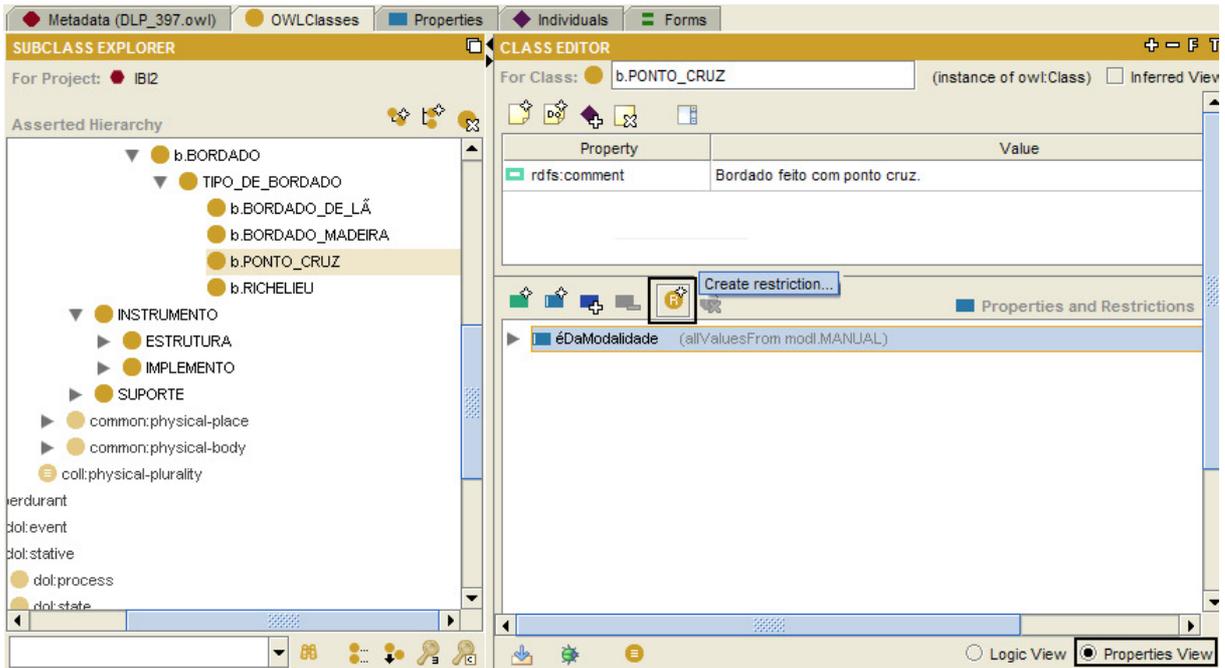


Figura 24. Caracterização dos conceitos com restrições (isto é, propriedades).

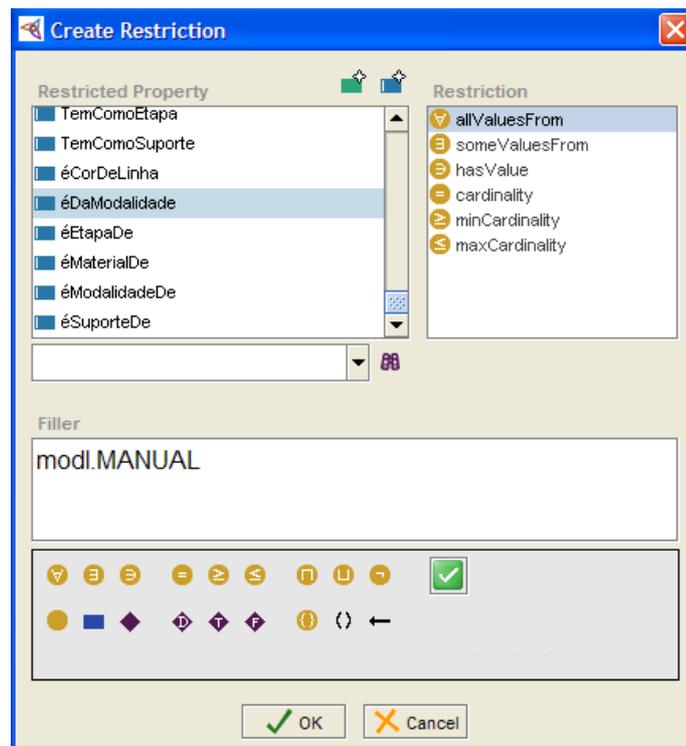


Figura 25. Janela para a implementação/edição e seleção de restrições (propriedades dos conceitos).

Verificamos que a Figura 23 apresenta também a relação 'éModalidadeDe', que foi implementada como uma relação inversa ('Inverse', no canto inferior esquerdo) à relação 'éDaModalidade'. Assim, registramos automaticamente que, por exemplo, se

“PONTO_CRUZ ‘éDaModalidade’ MANUAL”, então “MANUAL ‘éModalidadeDe’ PONTO_CRUZ”.

O Quadro 13 lista todas as relações (as restrições) e as suas inversas, bem como os seus respectivos os domínios e contra-domínios.

relação = éDaModalidade domínio da relação = TIPO_DE_BORDADO contra-domínio da relação = MODALIDADE_DE_CONFECCÃO relação inversa = éModalidadeDe	relação = TemComoEtapa domínio da relação = TIPO_DE_BORDADO contra-domínio da relação = ETAPA_DE_CONFECCÃO relação inversa = éEtapaDe
relação = TemComoSuporte domínio da relação = TIPO_DE_BORDADO contra-domínio da relação = TIPO_DE_SUORTE relação inversa = éSuporteDe	relação = FeitoComMaterial domínio da relação = TIPO_DE_BORDADO contra-domínio da relação = MATERIAL_DO_BORDADO relação inversa = éMaterialDe
relação = EnvolveOInstrumento domínio da relação = TIPO_DE_BORDADO contra-domínio da relação = INSTRUMENTO relação inversa = InstrumentoEnvolvidoEm	relação = TemCorPredominante domínio da relação = TIPO_DE_BORDADO contra-domínio da relação = COR relação inversa = CorPredominanteDe
relação = EnvolveOProfissional domínio da relação = ETAPA_DE_CONFECCÃO contra-domínio da relação = PROFISSIONAL_DO_BORDADO relação inversa = ProfissionalEnvolvidoEm	relação = TemLocalDeReferência domínio da relação = TIPO_DE_BORDADO contra-domínio da relação = LOCAL relação inversa = éLocalDeReferência

Quadro 13. Relações implementadas na ontologia IB(I).

Essas relações foram propostas, conforme anunciamos, com base principalmente na organização sintagmática dos elementos da rede FrameNet e recebem nomes intuitivos para facilitar o manuseio do ontoléxico via língua natural. A maioria dessas relações é utilizada na caracterização dos conceitos, mas há aquelas que se aplicam apenas a casos mais específicos como, por exemplo, a relação ‘TemCorPredominante’, que é útil para indicarmos que o *bordado blackwork* é feito com linha de determinada cor (preta), mas outros tipos de bordado podem não requerer essa informação; o mesmo se dá com a relação ‘TemLocalDeReferência’, utilizada para relacionarmos o bordado Madeira à Ilha da Madeira.

Para finalizar, comentamos a implementação de uma representação mais detalhada do conceito PONTO-CRUZ para ilustrar a sua caracterização com todas as relações possíveis embasadas em informações coletadas do *corpus* da pesquisa. A Figura 26, com a aba ‘OWL Classes’ do editor Protégé acionada, exibe o conjunto dessas relações

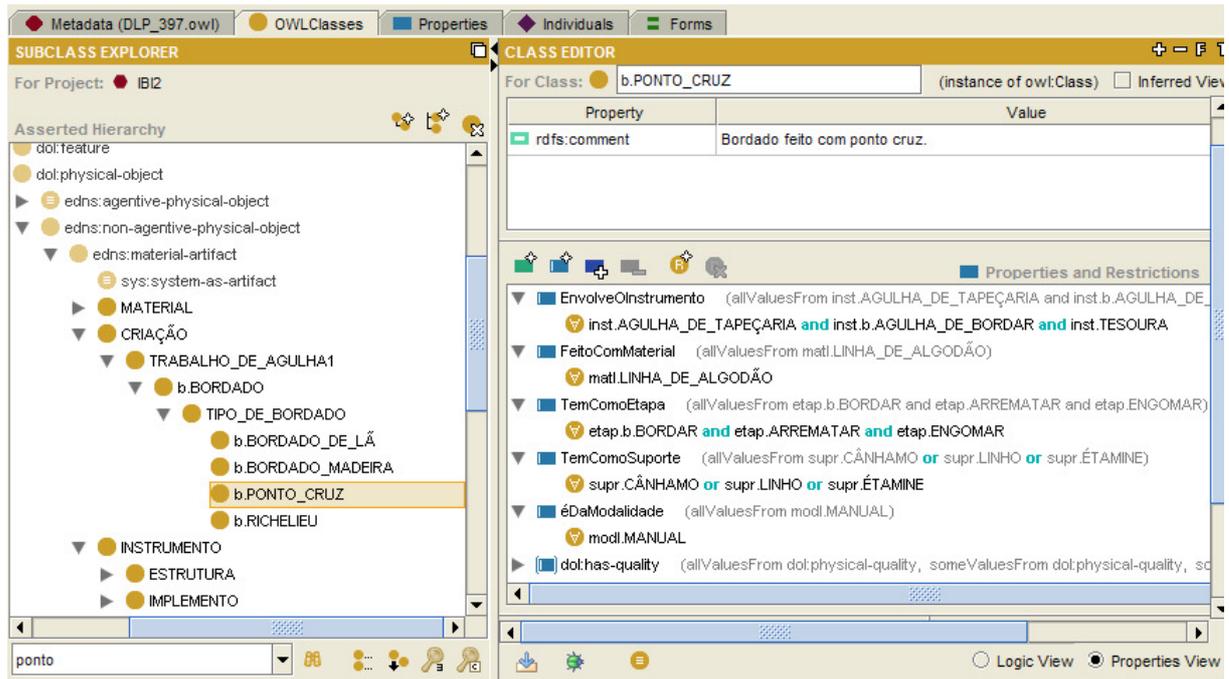


Figura 26. Caracterização do conceito PONTO-CRUZ (formalmente: ‘OWL Classes’).

A Figura 27, centrada na categoria (no conceito) TIPO_DE_BORDADO, mostra a visualização de um pequeno fragmento do ontológico do domínio léxico-conceitual da IB(I), em que estão exibidos os itens lexicais do português *bordado madeira*, *bordado de lã*, *richelieu* e *ponto cruz* e os itens lexicais correspondentes do inglês: *Madeira embroidery*, *crewelwork*, *richelieu* (ainda não implementado no editor Protegé) e *cross-stitch*. Essa visualização, conforme adiantamos, é feita através do *plug-in* TGVizTab, que permite “navegar” pela ontologia e verificar, fácil e interativamente, a estrutura hierárquica da ontologia que estrutura o ontológico, os diferentes tipos de relações que foram codificadas entre os seus elementos e os itens lexicais (de diferentes línguas) já inseridos.

Fica evidente, na Figura 27, a proposta de construção de ontológicos, ou seja, a estruturação dos itens lexicais de uma língua ou mais línguas naturais em termos dos conceitos que esses itens lexicais codificam (lexicalizam). Por exemplo: *ponto cruz* e *cross-stitch*, itens lexicais pertencentes aos léxicos do português do inglês, respectivamente, codificam o conceito PONTO CRUZ da ontologia do domínio léxico-conceitual da IB(I); a relação de “instanciação lexical” que se estabelece entre esses itens lexicais (do léxico) e o conceito (da ontologia) constitui, assim, a relação constitutiva essencial do ontológico.

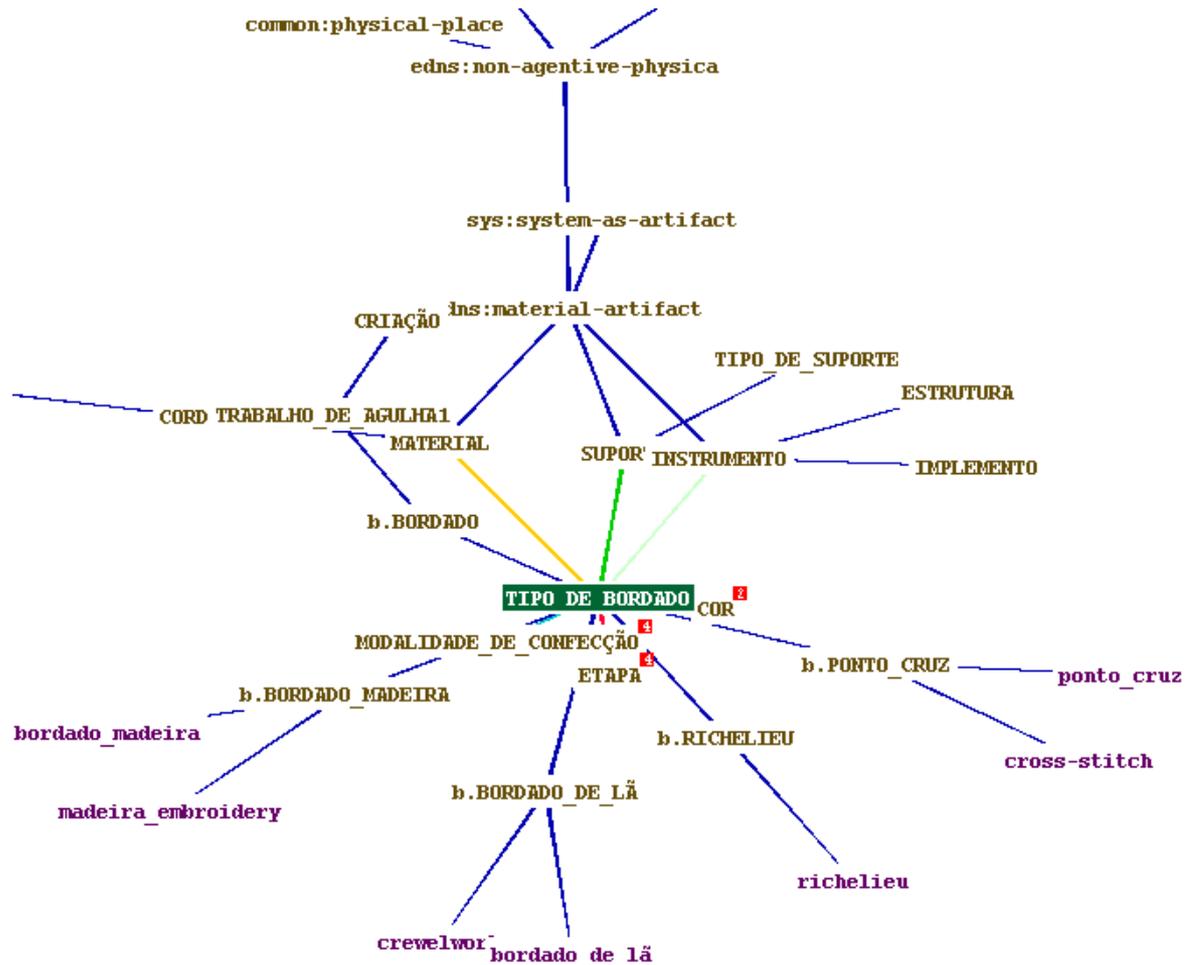


Figura 27. Fragmento do ontológico domínio léxico-conceitual da IB(I) (visualização com o TGVizTab).

Assim, concluindo o percurso de nossa pesquisa, ressaltamos que formalizar o máximo possível de informações léxico-conceituais do domínio léxico-conceitual da IB(I), tornando-as acessíveis e manipuláveis para os mais diversificados usos no campo das ontologias (ou seja, uso nas tecnologias de Recuperação de Informação, na Terminótica e na Web Semântica), é um objetivo pretensioso que não perseguimos neste estudo monográfico. No entanto, progressos nessa direção podem ser alcançados se forem estudados, desenvolvidos e aplicados os procedimentos de sistematização e implementação de ontológicos como os que relatamos nesta nossa incursão pelos caminhos dos Estudos do Léxico, da Semântica Lexical, do Processamento Automático de Línguas Naturais e das Ontologias.

4 CONSIDERAÇÕES FINAIS

Reacendemos, nestas considerações, a valorização do domínio escolhido para a pesquisa, pois a indústria do bordado, em especial a sua especialização inovativa e particular em Ibitinga, oferece muito a ser apreciado e estudado. Compor um *corpus* para o domínio léxico-conceitual da IB(I) foi uma realização importante, bem como foi a coleta de conceitos e de itens lexicais do português e do inglês, fundamentais para a representação exploratória do ontoléxico. Por meio da familiarização e da experimentação com ontologias e recursos lexicais, pudemos traçar uma metodologia que nos guiou pela nada trivial tarefa de construção da intrincada interface que nos propusemos construir entre léxicos e ontologias, ou seja, a construção de ontoléxicos, que se alicerçam a cobiçada tecnologia do século XXI: a *Web Semântica*.

Entretanto, para que a coleta das informações léxico-conceituais relevantes para a modelagem do ontoléxico se torne significativa e para que este estudo exploratório se estenda para a implementação de um ontoléxico abrangente e o mais completo possível, será ainda preciso atentar para os recursos (semi)automáticos de processamento de dados, principalmente os assentados na Linguística de *Corpus*, além da necessidade de acompanharmos os desenvolvimentos teóricos e metodológicos sempre atualizados pelos pesquisadores do PLN, com destaque para os estudiosos do ontoléxico, como Huang et al. (2010).

A expansão (isto é, a inserção de mais conceitos, gerais ou específicos, e das relações entre eles) e a alimentação (isto é, a inserção dos itens lexicais a relacionados a esses conceitos) do ontoléxico esboçado para o domínio léxico-conceitual da IB(I), conforme explanamos no parágrafo anterior, se mostram factíveis e necessárias, ficando a cargo de pesquisas futuras que se disponham a partir de onde paramos, propondo novos avanços e novas aplicações. De imediato, podemos destacar um ponto que, ao longo da nossa pesquisa, mostrou ter bastante potencial para esse desenvolvimento aspirado: trata-se do refinamento da parte ontológica (dos conceitos e das relações entre os conceitos) e da parte lexical (o acréscimo de itens lexicais de outras línguas, por exemplo) e do diálogo do ontoléxico com outros de domínios afins como, por exemplo, o das confecções em geral (contemplando a infinidade de produtos constituídos de tecidos bordados ou detalhes bordados) ou o das substâncias que são a matéria-prima para alguns elementos envolvidos no bordado, como as linhas e os tecidos, que podem ser de vários materiais.

Para exemplificar, de modo concreto, parte do conhecimento coletado e sistematizado neste trabalho, implementamos, num editor de ontologias versátil e gratuitamente acessível na Web, um pequeno fragmento do ontolético, alvo de toda a investigação – tarefa que poderá ser mais um alvo de trabalhos futuros, sobretudo, com as expansões e detalhamentos referidos nos parágrafos anteriores.

REFERÊNCIAS BIBLIOGRÁFICAS

ALLAN, K. **Natural language semantics**. Cornwall: Blackwell, 2001.

ALMEIDA, G. M. de B.; OLIVEIRA, L. H. M. de; ALUISIO, S. M.. A terminologia na era da informática. **Ciência e Cultura**, São Paulo, v. 58, n. 2., p. 42-45, 2006. Disponível em: <http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S000967252006000200016&lng=en&nrm=iso>. Acesso em: 04 jun. 2011.

ALUÍSIO, S. M.; ALMEIDA, G. M. B. **O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa linguística**. Calidoscópico. v. 4, n. 3, p. 155-177, set/dez 2006. Disponível em: <http://www.unisinos.br/publicacoes_cientificas/images/stories/pdfs_calidoscopio/vol4n3/art04_aluisio.pdf>. Acesso em: 04 jun. 2011.

ARAUJO, J. B.; GALDAMEZ, E. V. C.; CARPINETTI, L. C. R.; OLIVEIRA, J. F. G. Introdução da Prática de Medição de Desempenho para Produção Sustentável no Cluster Industrial de Ibitinga. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, n. 28, 2008, Rio de Janeiro. **Anais...** Rio de Janeiro: [s.n.], 2008. p. 1-14.

ATKINS, S.; RUNDELL, M. **The Oxford Guide to Practical Lexicography**. Oxford: Oxford University Press, 2008.

BABINI, M. Do conceito à palavra: os dicionários onomasiológicos. **Cienc. Cult.**, São Paulo, v. 58, n. 2, p. 38-42, 2006. Disponível em: <http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S000967252006000200015&lng=en&nrm=iso>. Acesso em: 04 jun. 2011.

BACIC, M. J.; SOUZA, M. C. A. F. Um estudo sobre os mecanismos de coordenação na determinação de custos e formação de preços no aglomerado de pequenas empresas de Ibitinga. **Revista del Instituto Internacional de Costos**, [S.l.], n. 3, p. 7-31, jul./dez. 2008. ISSN 1646-6896. Disponível em: <http://www.revistaic.org/articulos/num3/articulo1_esp.pdf>. Acesso em: 04 jun. 2011.

BORBA, F. S. **Pequeno vocabulário de linguística moderna**. 2ª ed. São Paulo: Companhia Editora Nacional, 1976.

_____. **Dicionário de usos do português do Brasil**. São Paulo: Ática, 2002.

_____. **Organização de dicionários: uma introdução à lexicografia**. São Paulo: Editora UNESP, 2003.

CABRÉ, M. T. **La terminología: representación y comunicación - elementos para una teoría de base comunicativa y otros artículos**. Barcelona: Universitat Pompeu Fabra, 1999.

CANANI, A. S. K. B. **De bonecas, flores e bordados: investigações antropológicas no campo do artesanato em Brasília**. 2008, 322 f. Tese (Doutorado em Antropologia Social) – Departamento de Antropologia – Universidade de Brasília, Brasília, 2008.

CHAGAS, L. L. **Do arado ao bordado: mudança no trabalho do homem do Sertão.** 2007, 141 f. Dissertação (Mestrado em Psicologia) – Centro de Ciências Humanas – Universidade de Fortaleza, Fortaleza, 2007.

CHISHMAN, R. L. O. Integrando léxicos semânticos e ontologias: uma aproximação a favor da Web Semântica. **Informação & Informação**, Londrina, v. 14, n. esp., p. 103-124, 2009. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/2159/3226>> Acesso em: 04 jun. 2011.

CINTRÃO, J. F. F. **O trabalho das bordadeiras de Ibitinga: fragmentação e subordinação.** 1990. 108 f. Dissertação (Mestrado em Ciências Sociais) – Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1990.

CORREIA, M.; LEMOS, L. S. P. A inovação lexical: como e porquê. In: _____. **Inovação lexical em português.** Lisboa: Colibri, 2005.

COSTA, M. P. Glossário de termos têxteis e afins. **Revista da Faculdade de Letras**, Porto, v. 3, p. 137-161.2004

CROFT, W. **Radical construction grammar** – Syntactic theory in typological perspective. Oxford: Oxford University Press, 2001.

CRUSE, D. A. **Lexical Semantics.** Cambridge: Cambridge University Press, 1986.

CRUZ, V. P. Entrevista com bordadeiras do Morro São Bento de Santos: uma reminiscência dos bordados da Ilha da Madeira. **Cadernos de Psicologia Social do Trabalho**, São Paulo, v. 10, n. 1, p. 121-136. 2007. Disponível em: <<http://www.revistasusp.sibi.usp.br/pdf/cpst/v10n1/v10n1a09.pdf>>. Acesso em: 04 jun. 2011.

DAY, L. F. **Art in needlework.** London: Bradbury, Agnew & Co. Ld., 1900.

DIAS-DA-SILVA, B. C. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais.** Araraquara, 1996. 272 p. Tese (Doutorado em Letras) – Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara. 1996

_____. Human language technology research and the development of the Brazilian Portuguese wordnet. In: INTERNATIONAL CONGRESS OF LINGUISTICS, n. 17., 2003, Prague. **Proceedings...** Prague: Matfyzpress, 2003. 12 p. 1 CR-ROM.

_____. O estudo linguístico-computacional da linguagem. **Letras de Hoje**. Porto Alegre, v. 41. p. 103-138. 2006. ISSN 0101-3335.

_____. Wordnet.Br: an exercise of human language technology reserch. **Palavra**, Rio de Janeiro, v. 12, p. 15-24. 2007. ISSN 1413-7763.

DI FELIPPO, A. **Delimitação e Alinhamento de Conceitos Lexicalizados no Inglês Norteamericano e no Português Brasileiro.** 2008, 253 f. Tese (Doutorado em Linguística e Língua Portuguesa) – Departamento de Linguística e Língua Portuguesa – Universidade Estadual Paulista, Araraquara, 2008.

DING, Y.; FOO, S. Ontology Research and Development part 1 – A review of ontology generation. **Journal of Information Science**, [S.l.] v. 28, n. 2, p. 123-136, abr. 2002.

DOWNING, D. A.; COVINGTON, M. A.; COVINGTON, M. M.; COVINGTON, C. A. **Dictionary of computer and Internet terms**. 10th ed. China: Barron's, 2009.

Enciclopédia Mirador internacional. São Paulo: Melhoramentos, 1976, v.4.

FELLBAUM, C. (Ed.) **WordNet**: an electronic lexical database. Cambridge, Massachusetts: Cambridge University Press, 1998.

FERREIRA, A. B. de H. **Miniaurêlio eletrônico**. Versão 5.12. [S.l.]: Positivo Informática, 2004. 1 CD-ROM.

FIESP – Departamento de Competitividade e Tecnologia. **A experiência da FIESP junto ao APL de enxovais de cama, mesa e banho e enxovais para o bebê de Ibitinga**. set. 2006. Disponível em: <<http://www.fiesp.com.br/competitividade/regional/apl%20ibitinga%202006.pdf>>. Acesso em: 04 jun. 2011.

FILLMORE, C. J. Frame semantics and the nature of language. In: CONFERENCE ON THE ORIGIN AND DEVELOPMENT OF LANGUAGE AND SPEECH, 1976, New York. **Annals of the New York Academy of Sciences**, v. 280. New York: [s.n.], 1976. p. 20-32.

FILLMORE, C. J.; JOHNSON, C. R.; PETRUCK, M. R. L. Background to FrameNet. **International Journal of Lexicography**, Oxford, v. 16, n. 3, p. 235-250, 2003.

FINATTO, M. J. B. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, A. N.; KRIEGER, M. G. (Org.) **As ciências do léxico**: lexicologia, lexicografia, terminologia. v. 2. Campo Grande: UFMS, 2004, p. 341-357.

FLORIAN, F.; LORENZO, H. C. Território e ambiente institucional: o arranjo produtivo local (APL) “Bordados de Ibitinga-SP”. **Revista Brasileira de Gestão e Desenvolvimento Regional**, Taubaté, v. 4, n. 4, p. 25-53, set./dez. 2008. Disponível em: <<http://www.rbgdr.net/032008/artigo2.pdf>>. Acesso em: 04 jun. 2011. ISSN: 1809-239X.

FRANÇA, P. C. Conceitos, classes e/ou universais: com o que é que se constrói uma ontologia? **LinguaMÁTICA**, [S.l.], v. 1, n. 1, p. 105-122, maio 2009. Disponível em: <<http://linguamatica.com/index.php/linguamatica/article/view/10/13>>. Acesso em: 04 jun. 2011. ISSN: 1647-0818.

GANGEMI, A.; GUARINO, N.; MASOLO, C.; OLTRAMARI, A. Interfacing WordNet with DOLCE: towards OntoWordNet. In: HUANG, C.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A.; PRÉVOT, L. (Eds.) **Ontology and the lexicon**: a natural language processing perspective. Cambridge: Cambridge University Press, 2010, p.36-52.

GELLER, J.; PERL, Y.; LEE, J. Editorial: Ontology Challenges: A Thumbnail Historical Perspective. **Knowledge and Information Systems**, London, v. 6, n. 4, p. 375-379, 2004.

GEROLAMO, M. C.; CARDOZA, E.; VICARI, F.; CARPINETTI, L. C. R. Aglomeração de pequenas e médias empresas como ambiente propício à melhoria de desempenho – caracterização do arranjo produtivo local do setor de confecção de bordados de Ibitinga.

Revista do Centro de Ciências Administrativas, Fortaleza, v. 11, n. 1, p. 82-92. Disponível em: <http://www.unifor.br/images/pdfs/pdfs_notitia/1381.pdf>. Acesso em: 04 jun. 2011.

ISSN (impresso): 1414-0896.

Grande enciclopédia Delta Larousse. Rio de Janeiro: Delta, 1970, v. 2.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, Stanford, v. 5, n. 2, p. 199-220, jun. 1993.

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

HIGGINS, L. **Handbook of embroidery**. London: Sampson Low, Marston, Searle and Rivington, 1880.

HIRST, G. Ontology and the Lexicon. In: STAAB, S.; STUDER, S. (Ed.). **Handbook on Ontologies**. Berlin: Springer-Verlag, 2004, p.209-229.

HORRIDGE, M. et al. **A practical guide to building OWL ontologies using the Protégé-OWL plugin and CO-ODE tools**. The University Of Manchester, 2004. Disponível em: <http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resources/ProtegeOWLTutorialP3_v1_0.pdf>. Acesso em: 04 jun. 2011.

HOUAISS, A. **Dicionário eletrônico Houaiss da língua portuguesa**. (Versão 1.0). Rio de Janeiro: FL Gama Design Ltda., 2001.

HUANG, C.; HSIAO, P.; SU, I.; KE, X. Paronymy: enriching ontological knowledge in WordNets. In: GLOBAL WORDNET CONFERENCE, n. 4, 2008, Szeged. **Proceedings...** Szeged: University of Szeged, 2008. p. 220-228.

HUANG, C.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A.; PRÉVOT, L. (Eds.) **Ontology and the lexicon: a natural language processing perspective**. Cambridge: Cambridge University Press, 2010.

IDE, N.; GREENSTEIN, D.; VOSSEN, P. Special Issue on EuroWordNet. **Computers and the Humanities**, Netherlands, v. 32, n. 2-3, 1998.

JACKENDOFF, R. **Foundations of language**. Oxford: Oxford University Press, 2002.

KATZ, J. J., FODOR, J. A. The structure of semantic theory. **Language**, [S.l.], n. 39, p. 170-210, 1963.

KRIEGER, M. G. Tipologias de dicionários: registros de léxico, princípios e tecnologias.

Calidoscópio, v.4, n.3, p.141-147, 2006. Disponível em:

<http://www.unisinos.br/publicacoes_cientificas/images/stories/pdfs_calidoscopio/vol4n3/art02_krieger.pdf>. Acesso em: 04 jun. 2011.

KRIEGER, M. G.; FINATTO, M. J. B. **Introdução à Terminologia**: teoria e prática. São Paulo: Contexto, 2004.

LEITE, M. P. As bordadeiras de Ibitinga: trabalho a domicílio e prática sindical. **Cadernos Pagu**, Campinas, n. 32, p. 183-214, 2009. Disponível em: <<http://www.scielo.br/pdf/cpa/n32/n32a07.pdf>>. Acesso em: 04 jun. 2011. ISSN 0104-8333.

LENCI, A. The life cycle of knowledge. In: HUANG, C.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A.; PRÉVOT, L. (Eds.) **Ontology and the lexicon**: a natural language processing perspective. Cambridge: Cambridge University Press, 2010, p.241-257.

Macmillan English dictionary for advanced learners. Versão 2.1. França: Black Publishers, 2007. CD-ROM.

MARCELLINO, E. R. A representação das lacunas pragmáticas no alinhamento de wordnets. In: SEMINÁRIO DO GEL, n. 56, 2008, São José do Rio Preto. **Programação...** São José do Rio Preto: [s.n.], 2008.

MARCELLINO, E. R.; DIAS-DA-SILVA, B. C. As lacunas e o alinhamento semântico de wordnets. In: XIX CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UNESP, 2007, Araraquara. **Anais...** Araraquara: [s.n.], 2007.

_____. A aplicação de relações de equivalência complexas na co-indexação entre wordnets. In: CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UFSCAR, n. 16, 2008, São Carlos. **Anais...** São Carlos: [s.n.], 2008, p. 729, v. 4. 1 CD-ROM

_____. Construção de uma ontologia de domínio e a busca da solução do tennis problem. In: ENCONTRO DE LINGUÍSTICA DE CORPUS, n. 8, 2009, Rio de Janeiro. **Livro de Resumos...** Rio de Janeiro: DGRAF UERJ, 2009. p. 29-32. Disponível em: <<http://www.corpuslg.org/elc/programa/resumos/Erasmus.html>>. Acesso em: 04 jun. 2011.

MATUSHIMA, M. K. **Especialização produtiva e aglomeração industrial: uma análise da indústria de confecções de Ibitinga-SP**. 2005. 197 f. Tese (Doutorado em Geografia) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, 2005.

NÉDELLEC, C.; NAZARENCO, A. Ontology and Information Extraction: a necessary symbiosis. In: BUITELAAR, P.; CIMIANO, P.; MAGNINI, B. (Eds.) **Ontology learning from text: methods, evaluation and applications**. Amsterdam: IOS, 2005. p.155-170. ISBN: 1-58603-523-1.

NILES, I.; PEASE, A. Towards a Standard Upper Ontology. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS, n. 2., 2001. **Proceedings...** Ogunquit: ACM, 2001. 8 p.

OLIVEIRA, M. L. G. **Bordado como assinatura**: tradição e inovação do artesanato na comunidade de Barateiro - Itapajé/CE. 2006, 167 f. Dissertação (Mestrado em Design e Sociedade) – Departamento de Artes e Design – PUC-Rio, Rio de Janeiro.

PEASE, A.; FELLBAUM, C. Formal ontology as interlíngua: the SUMO and WordNet linking Project and global WordNet. In: HUANG, C.; CALZOLARI, N.; GANGEMI, A.;

LENCI, A.; OLTRAMARI, A.; PRÉVOT, L. (Eds.) **Ontology and the lexicon: a natural language processing perspective**. Cambridge: Cambridge University Press, 2010, p.25-35.

PESSOA, F. **Ficção e teatro** (Introdução, organização e notas de António Quadros). Mem Martins: Europa-América, 1986, p.69. Disponível em: <<http://arquivopessoa.net/textos/2128>> Acesso em: 28 maio 2011.

PINTO, H. S.; MARTINS, J. P. Ontologies: how can they be built? **Knowledge and Information Systems**, London, v. 6, n. 4, p. 441-464, 2004.

PRÉVOT, L.; BORGIO, S.; OLTRAMARI, A. Interfacing ontologies and lexical resources. In: HUANG, C.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A.; PRÉVOT, L. (Eds.) **Ontology and the lexicon: a natural language processing perspective**. Cambridge: Cambridge University Press, 2010, p.185-200.

PRÉVOT, L.; HUANG, C.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A. Ontology and the lexicon: a multidisciplinary perspective. In: HUANG, C.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A.; PRÉVOT, L. (Eds.) **Ontology and the lexicon: a natural language processing perspective**. Cambridge: Cambridge University Press, 2010, p.3-24.

Random House unabridged electronic dictionary. (Versão 2.0). New York: WordPerfect, 1996. CD-ROM.

RODRIGUES, J. O.; MARCELLINO, E. R.; DIAS-DA-SILVA, B. C. Co-Indexação léxico-semântica de synsets de substantivos entre wordnets. In: SIMPÓSIO INTERNACIONAL DE INICIAÇÃO CIENTÍFICA DA USP, n. 16, 2008, São Paulo. **Anais...** São Paulo: [s.n.], 2008.

ROSA, R. **Porque acreditar em Ibitinga**. Ibitinga: Predolim, 1991.

ROSCH, E. Natural categories. **Cognitive psychology**, [S.l.], n. 4, p. 328-350, 1973.

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

SANDMANN, A. J. Produtividade lexical. In: _____. **Competência lexical – produtividade, restrições e bloqueio**. Curitiba: UFPR, 1991.

SARDINHA, T. B. Linguística de corpus: histórico e problemática. **Revista de Documentação em Linguística Teórica e Aplicada**, São Paulo, v. 16, n. 2, p. 323-367, 2000. Disponível em: <<http://www.scielo.br/pdf/delta/v16n2/a05v16n2.pdf>>. Acesso em: 04 jun. 2011. ISSN 0102-4450

SCHEFFCZYK, J.; BAKER, C. F.; NARAYANAN, S. Reasoning over natural language text by means of FrameNet and ontologies. In: HUANG, C.; CALZOLARI, N.; GANGEMI, A.; LENCI, A.; OLTRAMARI, A.; PRÉVOT, L. (Eds.) **Ontology and the lexicon: a natural language processing perspective**. Cambridge: Cambridge University Press, 2010, p.53-71.

SERVIÇO Brasileiro de Apoio às Micro e Pequenas Empresas. **Bordados e rendas – cama, mesa e banho: estudos de mercado SEBRAE/ESPM (relatório completo)**. SEBRAE, 2008.

SILVA, P. F. T. L. **Bordados tradicionais portugueses**. 2006. 133 f. Dissertação (Mestrado em Design e Marketing) – Departamento de Engenharia Têxtil, Universidade do Minho, Guimarães, 2006.

SINGLETON, D. **Language and the lexicon**. London: Hodder Headline Group/Oxford University Press, 2000.

TEUBERT, W.; ČERMÁKOVÁ, A. **Corpus linguistics: a short introduction**. London: Continuum, 2007.

TEIXEIRA-DA-MATA, J. V. Introdução às categorias. In: ARISTÓTELES. **Categorias**. Goiânia: UFG, Alternativa, 2005, p.7-74.

The new encyclopaedia Britannica in 30 volumes. 15. ed., v. III. USA, 1974.

TURCO, L. **A Experiência no APL de enxovais bordados de cama, mesa e banho de Ibitinga**. Ibitinga: Sindicato das Indústrias e Comércio de Bordados de Ibitinga, 2007. Disponível em: <http://www.mdic.gov.br/arquivos/dwnl_1198780394.pdf>. Acesso em: 04 jun. 2011. Apresentação em PowerPoint.

VIEIRA, A. P. P. N. **Bordados tradicionais portugueses**. 2002. 159 f. Dissertação (Mestrado em Design e Marketing) – Departamento de Engenharia Têxtil, Universidade do Minho, Guimarães, 2002.

VOSSSEN, P. **EuroWordNet: a multilingual database with lexical semantic networks for European Languages**. Dordrecht: Kluwer, 1998.

VOSSSEN, P. Ontologies. In: MITKOV, R. (Ed.). **The Handbook of Computational Linguistics**. Oxford: Oxford University Press, 2003. p. 464-482.

WEISZFLOG, W. (Ed.) **Michaelis português** – moderno dicionário da língua portuguesa. Versão 1.0. São Paulo: DTS Software Brasil Ltda, 1998. CD-ROM.

WILKS, Y., FASS, D.; GUO, C.; McDONALD, J.; PLATE, T.; SLATOR, B. Providing machine tractable dictionary tools. In: PUSTEJOVSKY, J. (Ed.) **Semantics and the Lexicon**. Dordrecht: Kluwer Academic Publishers, 1993, p.341-401.

APÊNDICE 1

Este apêndice registra uma amostra dos itens lexicais ancorados no domínio léxico-conceitual da IB(I), a partir de buscas em parte do *corpus* disponível. As buscas foram feitas manualmente, com exceção às feitas no dicionário *Random House*, que permitiu a utilização da ferramenta de pesquisa reversa.

As bordadeiras de Ibitinga: trabalho a domicílio e prática sindical (LEITE, 2009)

arremate; arte de bordar; Associação Profissional dos Empregados em Empresas de Artesanatos em Geral de Ibitinga; bordadeiras; bordado; corte; costura; fabricação; fábricas; Feira do Bordado de Ibitinga; indústria do bordado; lavanderia; máquina de; bordado industrial; máquina de costura; máquinas; máquinas de bordar; operários; produtos de cama, mesa e banho; risco; salão; salão de bordado; Sindicato dos Bordadores de Ibitinga; Sindicato dos Empregados em Empresas de Artesanatos em Geral de Ibitinga; Sindicato dos Trabalhadores nas Indústrias de Confecção e Bordado de Ibitinga e Região; técnica

O trabalho das bordadeiras de Ibitinga: fragmentação e subordinação (CINTRÃO, 1990)

acabamento; agulha da máquina de bordar; arrematação; bastidores; bicos; bordadeiras; bordado artesanal; bordado industrial; cortadora de bicos; costureira; desfiadura; embaladeira; fabricação; fios; linha; máquina de bordar; máquinas de costura; máquinas industriais; máquinas semi-automáticas; matéria-prima; modelos; passadeira; riscadeira; riscadura; riscos; salões de bordado; tecidos; tesouras; vendedora

BORDADO. *Enciclopédia Mirador Internacional (1976).*

(pontos) cegos; (pontos) corrediços; (pontos) simples; agulha de gancho; agulhas; algodão; aplicações; arabescos; bastidor com pés; bastidor quadrado; bastidor redondo; bastidor retangular; bastidor simples; bastidores; batista; bordado; bordado a branco; bordado a cheio; bordado a matiz; bordado árabe; bordado branco; bordado de cadeia; bordado de cor; bordado de passagem; bordado de recorte; bordado de relevo; bordado de renda; bordado de transporte; bordado em relevo; bordado inglês; bordado manual; bordado mosaico; bordado plano; bordado recortado; bordado rendado; bordado sobre tapeçaria; bordado sobre tela; cambraia; canutilhos; carreiras; cartão; conchinhas; contas de madeira; contas de vidro; cordões; cortinas; couro; crivo; estofos; feltro; festão; fibra sintética; fio de algodão; fio de linho; fio de ouro; fio de prata; fio de ráfia; fio de seda; fios metálicos; fios tintos; galões; gazes; guipure; ilhós; lantejoulas; lãs; linha de algodão; linho; materiais preciosos; materiais rústicos; musselinas; nylon; opala; ornamentos; ouro; palha; paramentos litúrgicos; passamanaria; pedaços de tecidos; pedras preciosas; pergaminho; pérolas; pintura a agulha; ponto cheio; ponto de alinhavo; ponto de barraca; ponto de cruzeta; ponto de nós; ponto florentino; ponto horizontal; ponto real; ponto renascença; ponto Richelieu; ponto veneziano; pontos contados; prata; recamo; rendas; reposteiros; rococó; roseta; roupas de confecção; roupas interiores; sedas; sementes; talagarça; tambor; tapeçarias; tapetes; tecido; tecido branco; tecido de trama frouxa; tecido-base; tecidos espessos; tecidos fechados; tecidos leves; tecidos transparentes; tela; toalhas de mesa; torçal; túnicas; vestes

TUDO_ARTE . <www.tudoarte.com.br/>, website em que foram identificadas 8 lojas virtuais que comercializam produtos fabricados em Ibitinga: </jrsemaninha>, </pesosdeporta>, </panosdeprato>, </regianeartesanato>, </ldbonecasdepano>, </arasolpijamas>, </veralucia> e </coisasdecozinha>.

avental; babydoll; baihna; barrado; caminho; caminho de mesa; camisola; capa de assadeira; capa de batedeira; capa de botijão; capa de galão de água; capa de liquidificador; capa de microondas; capa para fogão; capa para tampa; cesta de pão; cesta de prato; chinelo; gorgurão; gorgurinho; juta; lese; linhão; overloque; pano de prato; passa fita; peso de porta; pijama; popeline; porta assadeira; porta fralda; porta papel toalha; porta prato; porta talheres; protetor de vidro de fogão; puxa-saco; robe; sacaria; tapete; tapete para o assento; tapete para o vaso sanitário; tecido 100% algodão; tecido atoalhado; tecido estampado; tecido liso; tecido misto; tecido tergal; textolen; tira de algodão; toalha de fogão; toalha de mesa; toalhina para fogão; viés

Especialização produtiva e aglomeração industrial: uma análise da indústria de confecções de Ibitinga-SP (MATUSHIMA, 2005)

almofadas; amostra; aplicações; apliques de fitas; babados; bordado; bordado computadorizado; bordados em tecido; capas de colchão; capas de sofá; colchas; cortar (as peças); cortinas; costura reta; costureiras; edredons; enxovais de bebê; enxovais de noivas; fronhas; indústria de bordados; lençóis; linhas; máquina de costura manual; máquinas; máquinas computadorizadas; molde; mosquiteiros; overloque; panos de prato; peças; ramo de confecções bordadas de cama, mesa e banho; rendas; riscadeira; riscar (os desenhos a serem bordados); roupas de bebê; tapetes; tecido riscado; tecidos; toalhas; toalhas de mesa; travesseiros

APPAREL, APPENZELL, BERLIN WOOL,... *Random House* (1996) (Aqui disponibilizamos as palavras-entrada e as respectivas definições, localizados através de buscas por palavras-chave. As marcadas com o asterisco (*) foram também identificadas nos synsets da rede WN.Pr).

Palavra-chave: **embroidery**

apparel

5. Eccles. a piece of embroidery, usually oblong, on certain vestments, esp. on the alb or amice.

Appenzell

3. (often l.c.) a type of embroidery, used chiefly on fine handkerchiefs.

Berlin wool

a soft woolen yarn for embroidery or knitting. Also called berlin.

black work

embroidery done with black, usually silk, thread on white fabric, esp. linen. Also, black work.

bolting cloth

a sturdy fabric, usually of fine silk or nylon mesh, used chiefly in serigraphy, embroidery, and as a foundation fabric for wigs.

Bonnaz

embroidery made by machine.

brede

2. braiding or embroidery.

broderie anglaise

fine white needlework done on fine cloth, typically on eyelet. Also, broderie Anglaise. Also called Madeiraembroidery.

bullion

4. embroidery or lace worked with gold wire or gold or silver cords.

cannetille

a gold or silver thread with a spiral twist, formerly much used in embroidery

canvas *

7. any fabric of linen, cotton, or hemp of a coarse loose weave used as a foundation for embroidery stitches, interlining, etc.

chenille *

1. a velvety cord or yarn of silk or worsted, for embroidery, fringes, etc.

crewel

1. Also called crewel yarn. a worsted yarn for embroidery and edging.

crewelwork *

decorative embroidery done with crewel yarn on cotton or linen, using simple stitches traditionally worked in floral or pastoral designs.

cross-stitch *

2. embroidery or needlepoint done with this stitch.

cutwork *

1. openwork embroidery in which the ground fabric is cut out about the pattern.

4. See point coupé (def. 1).

embroidery needle *

a needle with a long eye, used for embroidery and fine darning.

eyelet *

1. a small hole, usually round and finished along the edge, as in cloth or leather for the passage of a lace or cord or as in embroidery for ornamental effect.

featherstitch

1. an embroidery stitch producing work in which a succession of branches extend alternately on each side of a central stem.

floss *

2. silk filaments with little or no twist, used in weaving as brocade or in embroidery.

gros point *

1. a large stitch used in embroidery. Cf. petit point (def. 1), tent stitch.

hardanger

embroidery openwork having elaborate symmetrical designs created by blocks of satin stitches within which threads of the embroidery fabric are removed.

herringbone stitch *

a type of cross-stitch in embroidery similar to the catch stitch in sewing, consisting of an overlapped V-shaped stitch that when worked in a continuous pattern produces a twill-weave effect.

ladder stitch

an embroidery stitch in which crossbars at equal distances are produced between two solid ridges of raised work.

needlepoint *

1. embroidery upon canvas, usually with uniform spacing of stitches in a pattern.

needlework *

1. the art, process, or product of working with a needle, esp. in embroidery, needlepoint, tapestry, quilting, and appliqué.

2. the occupation or employment of a person skilled in embroidery, needlepoint, etc. Also called needlecraft, stitchery.

openwork *

any kind of work, esp. ornamental, as of embroidery, lace, metal, stone, or wood, having a latticelike nature or showing openings through its substance.

orphrey *

1. an ornamental band or border, esp. on an ecclesiastical vestment.

2. gold embroidery.

3. rich embroidery of any sort.

4. a piece of richly embroidered material.

overhand *

3. (in sewing and embroidery) with close, shallow stitches over two edges.

petit point *

1. a small stitch used in embroidery. Cf. gros point (def. 1), tent stitch.

2. embroidery done on a canvas backing and resembling woven tapestry.

picot *

one of a number of ornamental loops in embroidery, or along the edge of lace, ribbon, etc.

pinwork

(in the embroidery of needlepoint lace) crescent-shaped stitches raised from the surface of the design.

rope stitch

(in embroidery) a stitch formed from the entwining of stitches.

runner

17. a long, narrow strip of line, embroidery, lace, or the like, placed across a table.

satin stitch *

a long, straight embroidery stitch worked closely parallel in rows to form a pattern that resembles satin.

sewing cotton

cotton thread used for sewing, embroidery, etc.

sewing silk

finely twisted silk thread used for sewing, embroidery, etc.

smocking *

2. embroidery stitches used to hold gathered cloth in even folds.

stitch *

1. one complete movement of a threaded needle through a fabric or material such as to leave behind it a single loop or portion of thread, as in sewing, embroidery, or the surgical closing of wounds.

stitchwork

embroidery or needlework.

stumpwork

a type of embroidery popular in the 17th century, consisting of intricate, colorful designs padded with horsehair to make them stand out in relief.

taboret

2. a frame for embroidery.

tambour *

3. Also called tabaret. a circular frame consisting of two hoops, one fitting within the other, in which cloth is stretched for embroidering.

4. embroidery done on such a frame.

tent stitch *

a short, slanting stitch used in embroidery. Cf. gros point (def. 1), petit point (def. 1).

tracery *

2. any delicate, interlacing work of lines, threads, etc., as in carving or embroidery; network.

trapunto

quilting having an embossed design produced by outlining the pattern with single stitches and then padding it with yarn or cotton.

branch (verbo)

15. to adorn with needlework; decorate with embroidery, as in textile fabrics.

work (verbo)

45. to make or decorate by needlework or embroidery: She worked a needlepoint cushion.

Palavra-chave: **embroidered**

bargello *

1. a straight stitch worked in a high and low relief pattern to form a variety of zigzag or oblique designs.

2. needlepoint work or a design done in such stitches, esp. the traditional needlepoint created by a classic stitch (Florentine stitch) done in diagonal lines.

bis2

a sheer, often embroidered linen, used in the manufacture of altar cloths. Also, bisso.

clock2

1. a short embroidered or woven ornament on each side or on the outer side of a sock or stocking, extending from the ankle upward.

<p>ephod a richly embroidered, apronlike vestment having two shoulder straps and ornamental attachments for securing the breastplate, worn with a waistband by the high priest.</p> <p>infula one of the two embroidered lappets of the miter of a bishop.</p> <p>lappet weaving weaving into which an embroidered pattern produced by additional warp threads has been introduced with the aid of a lappet.</p> <p>sampler * 2. a piece of cloth embroidered with various stitches, serving to show a beginner's skill</p>
<p>Palavra-chave: embroidering</p>
<p>couching 2. a method of embroidering in which a thread, often heavy, laid upon the surface of the material, is caught down at intervals by stitches taken with another thread through the material. 3. work so made. schiffli 1. a large, loomlike machine for embroidering textiles and working patterns in lace.</p>
<p>Palavra-chave: needlework</p>
<p>course 17. the row of stitches going across from side to side in knitting and other needlework (opposed to wale).</p> <p>drawn work * ornamental work done by drawing threads from a fabric, the remaining portions usually being formed into lacelike patterns by needlework.</p> <p>hemstitch * 3. the stitch used or the needlework done in hemstitching.</p> <p>needlewoman * a woman who does needlework.</p> <p>sewing table a worktable for holding sewing materials, often supplied with a bag or pouch for needlework. Cf. bag table.</p> <p>stiletto 2. a pointed instrument for making eyelet holes in needlework.</p> <p>white work needlework done in white on fine white cloth, esp. linen. Also, white work.</p> <p>workbag * a bag for holding implements and materials for work, esp. needlework.</p> <p>workbasket * a basket used to hold needlework paraphernalia.</p> <p>workbox * a box to hold instruments and materials for work, esp. needlework.</p>

APÊNDICE 2

Neste apêndice, apresentamos amostras do conteúdo de synsets que foram analisados para a seleção das informações que integram o ontoléxico IB(I). Cada synset foi analisado em um arquivo padrão, em conformidade com a metodologia de alinhamento entre wordnets, relatada na subseção 2.1.1.

Synset para o conceito ARREMATADOR.

ANÁLISE DA GLOSA ARREMATADOR

A ANÁLISE RESULTA EM:

S='09013755' = "trabalhador que executa as últimas etapas de um processo manufaturado"

{arrematador: A plasticidade divide a trabalhadora domiciliar em costureira, bordadeira, arrematadora de bordado, etc. }

Chave: ARREMATADOR <noun.person>

ILI: 09013755

Sense 3

{09013755} <noun.person> finisher -- (a worker who performs the last step in a manufacturing process)

=> {09025575} <noun.person> worker -- (a person who works at a specific occupation; "he is a good worker")

=> {00006026} <noun.Tops> person, individual, someone, somebody, mortal, human, soul -- (a human being; "there was too much for one person to do")

=> {00003226} <noun.Tops> organism, being -- (a living thing that has (or can develop) the ability to act or function independently)

=> {00003009} <noun.Tops> living thing, animate thing -- (a living (or once living) entity)

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00005598} <noun.Tops> causal agent, cause, causal agency -- (any entity that causes events to happen)

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

Synset para o conceito BASTIDOR.

ANÁLISE DA GLOSA BASTIDOR

A ANÁLISE RESULTA EM:

S='04217323' = "armação com dois arcos usada no bordado"

{bastidor: Depois arma com o bastidor e põe a máquina pra funcionar. }

Chave: BASTIDOR <noun.artifact>

ILI: 04217323

Sense 1

{04217323} <noun.artifact> tambour1, embroidery frame, embroidery hoop -- (a frame made of two hoops; used for embroidering)

=> {03267657} <noun.artifact> framework, frame2, framing -- (a structure supporting or containing something)

=> {04192033} <noun.artifact> supporting structure -- (a structure that serves to support something)

=> {04174544} <noun.artifact> structure, construction -- (a thing constructed; a complex construction or entity; "the structure consisted of a series of arches"; "she wore her hair in an amazing construction of whirls and ribbons")

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

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00002645} <noun.Tops> whole, whole thing, unit -- (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, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Synset para o conceito BORDADO.

ANÁLISE DA GLOSA BORDADO

A ANÁLISE RESULTA EM:

S='03163893' = "trabalho de agulha decorativo"

{bordado: O bordado tem seus segredos e a finalização dele é que o diferencia dos demais.}

Chave: BORDADO <noun.artifact>

ILI: 03163893

Sense 2

{03163893} <noun.artifact> embroidery, fancywork -- (decorative needlework)

=> {03673332} <noun.artifact> needlework, needlecraft -- (a creation created or assembled by needle and thread)

=> {03015859} <noun.artifact> creation -- (an artifact that has been brought into existence by someone)

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

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00002645} <noun.Tops> whole, whole thing, unit -- (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, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Synset para o conceito

Synset para o conceito BORDADO DE LÃ.

ANÁLISE DA GLOSA BORDADO DE LÃ

A ANÁLISE RESULTA EM:

S='03017889' = "bordado feito com estame de lã frouxamente retorcido"

{bordado de lã: Elas aprenderam a pintar e demarcar os pontos, desenhando flores e pássaros para posteriormente preenchê-los com o bordado de lã.}

Chave: BORDADO DE LÃ <noun.artifact>

ILI: 03017889

Sense 1

{03017889} <noun.artifact> crewelwork -- (embroidery done with loosely twisted worsted yarn)

=> {03163893} <noun.artifact> embroidery, fancywork -- (decorative needlework)

=> {03673332} <noun.artifact> needlework, needlecraft -- (a creation created or assembled by needle and thread)

=> {03015859} <noun.artifact> creation -- (an artifact that has been brought into existence by someone)

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

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00002645} <noun.Tops> whole, whole thing, unit -- (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, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Synset para o conceito CONTROLE DE QUALIDADE.

ANÁLISE DA GLOSA CONTROLE DE QUALIDADE

A ANÁLISE RESULTA EM:

S='00756926' = "manutenção dos padrões de qualidade de bens manufaturados"

{controle de qualidade: O bordado passa por um rigoroso controle de qualidade, responsável pela detecção de eventuais defeitos em quaisquer das fases de produção.}

Chave: CONTROLE DE QUALIDADE <noun.act>

ILI: 00756926

Sense 1

{00756926} <noun.act> quality control -- (maintenance of standards of quality of manufactured goods)

=> {00756518} <noun.act> internal control -- (an accounting procedure or system designed to promote efficiency or assure the implementation of a policy or safeguard assets or avoid fraud and error etc.)

=> {00754095} <noun.act> control -- (the activity of managing or exerting control over something; "the control of the mob by the police was admirable")
 => {00389883} <noun.act> activity -- (any specific activity; "they avoided all recreational activity")
 => {00026194} <noun.Tops> act, human action, human activity -- (something that people do or cause to happen)

Synset para o conceito CORTADOR.

ANÁLISE DA GLOSA CORTADOR

A ANÁLISE RESULTA EM:

S='09914482' = "alguém cujo trabalho é cortar"

{cortador: Ele tem como atividade Auxiliar o cortador no enfiado e corte de tecidos para a confecção de cuecas, pijamas e camisetas.}

Chave: CORTADOR <noun.person>

ILI: 09914482

Sense 3

{09346554} <noun.person> cutter -- (someone whose work is cutting (as e.g. cutting cloth for garments))

=> {09914482} <noun.person> skilled worker, trained worker -- (a worker who has acquired special skills)

=> {09025575} <noun.person> worker -- (a person who works at a specific occupation; "he is a good worker")

=> {00006026} <noun.Tops> person, individual, someone, somebody, mortal, human, soul -- (a human being; "there was too much for one person to do")

=> {00003226} <noun.Tops> organism, being -- (a living thing that has (or can develop) the ability to act or function independently)

=> {00003009} <noun.Tops> living thing, animate thing -- (a living (or once living) entity)

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00005598} <noun.Tops> causal agent, cause, causal agency -- (any entity that causes events to happen)

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

Synset para o conceito COSTURA.

ANÁLISE DA GLOSA COSTURA

A ANÁLISE RESULTA EM:

S='00673125' = "o ato de ligar ou prender com pontos"

{costura: A costura de enxovais fez parte do universo feminino.}

Chave: COSTURA <noun.act>

ILI: 00673125

Sense 1

{00673125} <noun.act> sewing, stitching -- (joining or attaching by stitches)

=> {00672395} <noun.act> handicraft -- (a craft that requires skillful hands)

=> {00572850} <noun.act> trade, craft -- (the skilled practice of a practical occupation; "he learned his trade as an apprentice")

=> {00553013} <noun.act> occupation, business, job, line of work, line -- (the principal activity in your life that you do to earn money; "he's not in my line of business")

=> {00389883} <noun.act> activity -- (any specific activity; "they avoided all recreational activity")

=> {00026194} <noun.Tops> act, human action, human activity -- (something that people do or cause to happen)

=> {00879019} <noun.act> arts and crafts -- (the arts of decorative design and handicraft; "they sponsored arts and crafts in order to encourage craftsmanship in an age of mass production")

=> {00877471} <noun.act> art, artistic creation, artistic production -- (the creation of beautiful or significant things; "art does not need to be innovative to be good"; "I was never any good at art"; "he said that architecture is the art of wasting space beautifully")

=> {00854109} <noun.act> creation, creative activity -- (the human act of creating)

=> {00389883} <noun.act> activity -- (any specific activity; "they avoided all recreational activity")

=> {00026194} <noun.Tops> act, human action, human activity -- (something that people do or cause to happen)

Synset para o conceito ENGOMAR.

ANÁLISE DA GLOSA ENGOMAR

A ANÁLISE RESULTA EM:

S='00405741' = "endurecer usando goma"

{engomar: Eu bordo do jeito que eu quero, aí eu mando lavar, passar e engomar, aí eu recorto.}

Chave: ENGOMAR <verb.change>

ILI: 00405741

Sense 1

{00405741} <verb.change> starch -- (stiffen with starch; "starch clothes")

=> {00405481} <verb.change> stiffen1 -- (make stiff or stiffer; "Stiffen the cream by adding gelatine")

=> {00121430} <verb.change> change1, alter1, modify10 -- (cause to change; make different; cause a transformation; "The advent of the automobile may have altered the growth pattern of the city"; "The discussion has changed my thinking about the issue")

Synset para o conceito LAVADOR.

ANÁLISE DA GLOSA LAVADOR

A ANÁLISE RESULTA EM:

S='10066346' = "alguém cujo trabalho é lavar coisas"

{lavador: Sempre que as peças chegavam num cesto maior, o lavador era obrigado a transferi-las para outro menor.}

Chave: LAVADOR <noun.person>

ILI: 10066346

Sense 1

{10066346} <noun.person> washer -- (someone who washes things for a living)

=> {09025575} <noun.person> worker -- (a person who works at a specific occupation; "he is a good worker")

=> {00006026} <noun.Tops> person, individual, someone, somebody, mortal, human, soul -- (a human being; "there was too much for one person to do")

=> {00003226} <noun.Tops> organism, being -- (a living thing that has (or can develop) the ability to act or function independently)

=> {00003009} <noun.Tops> living thing, animate thing -- (a living (or once living) entity)

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence

(living or nonliving))

=> {00005598} <noun.Tops> causal agent, cause, causal agency -- (any entity that causes events to happen)

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

Synset para o conceito LAVAR.

ANÁLISE DA GLOSA LAVAR

A ANÁLISE RESULTA EM:

S='01492749' = "limpar com água e algum agente limpante como sabão"

{lavar: Eu bordo do jeito que eu quero, aí eu mando lavar, passar e engomar.}

Chave: LAVAR <verb.contact>

ILI: 01492749

Sense 3

{01492749} <verb.contact> wash, launder -- (cleanse with a cleaning agent, such as soap, and water; "Wash the towels, please!")

=> {01490246} <verb.contact> clean, make clean -- (make clean by removing dirt, filth, or unwanted substances from; "Clean the stove!"; "The dentist cleaned my teeth")

=> {00121430} <verb.change> change1, alter1, modify10 -- (cause to change; make different; cause a transformation; "The advent of the automobile may have altered the growth pattern of the city"; "The discussion has changed my thinking about the issue")

Synset para o conceito LINHO.

ANÁLISE DA GLOSA LINHO

A ANÁLISE RESULTA EM:

S='03535711' = "tecido feito com fibras de linho"

{linho: Originalmente, o bordado blackwork era feito sobre um linho branco com fios de seda pretos.}

Chave: LINHO <noun.artifact>

ILI: 03535711

{03535711} <noun.artifact> linen1 -- (a fabric woven with fibers from the flax plant)

=> {03189674} <noun.artifact> fabric, cloth, material, textile -- (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, artefact -- (a man-made object taken as a whole)

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00002645} <noun.Tops> whole, whole thing, unit -- (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, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Synset para o conceito PONTO CRUZ.

ANÁLISE DA GLOSA PONTO CRUZ

A ANÁLISE RESULTA EM:

S='03023957' = "bordado feito com ponto cruz"

{ponto cruz: No ponto cruz já está contado aquilo que você vai bordar.}

Chave: PONTO CRUZ <noun.artifact>

ILI: 03023957

Sense 2

{03023957} <noun.artifact> cross-stitch -- (embroidery done in cross-stitch)

=> {03163893} <noun.artifact> embroidery, fancywork -- (decorative needlework)

=> {03673332} <noun.artifact> needlework, needlecraft -- (a creation created or assembled by needle and thread)

=> {03015859} <noun.artifact> creation -- (an artifact that has been brought into existence by someone)

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

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00002645} <noun.Tops> whole, whole thing, unit -- (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, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Synset para o conceito RISCADOR.

ANÁLISE DA GLOSA RISCADOR

A ANÁLISE RESULTA EM:

S='10087845' = "profissional que desenha o risco a ser bordado em um suporte"

{riscador: Para cada riscadeira eram necessárias cinco bordadeiras para que ninguém ficasse parado. }

Chave: RISCADOR <noun.person>

ILI: EQ_HAS_HYPERONYM = 10087845

Sense 1

{10087845} <noun.person> workman, working man, working person -- (an employee who performs manual or industrial labor)

=> {09410258} <noun.person> employee -- (a worker who is hired to perform a job)

=> {09025575} <noun.person> worker -- (a person who works at a specific occupation; "he is a good worker")

=> {00006026} <noun.Tops> person, individual, someone, somebody, mortal, human, soul -- (a human being; "there was too much for one person to do")

=> {00003226} <noun.Tops> organism, being -- (a living thing that has (or can develop) the ability to act or function independently)

=> {00003009} <noun.Tops> living thing, animate thing -- (a living (or once living) entity)

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00005598} <noun.Tops> causal agent, cause, causal agency -- (any entity that causes events to happen)

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

Synset para o conceito TESOURA.

ANÁLISE DA GLOSA TESOURA

A ANÁLISE RESULTA EM:

S='03991516' = "ferramenta afiada que possui duas lâminas presas por um eixo"

{tesoura: Pessoas munidas de pequenas tesouras se encarregam de cortar as pontas de fios excedentes. }

Chave: TESOURA <noun.artifact>

ILI: 03991516

Sense 1

{03991516} <noun.artifact> scissors, pair of scissors -- (and edge tool having two crossed pivoting blades)

=> {03146315} <noun.artifact> edge tool -- (any cutting tool with a sharp cutting edge (as a chisel or knife or plane or gouge))

=> {03039706} <noun.artifact> cutter, cutlery1, cutting tool -- (a cutting implement; a tool for cutting)

=> {03040079} <noun.artifact> cutting implement -- (a tool used for cutting or slicing)

=> {04279652} <noun.artifact> tool -- (an implement used in the practice of a vocation)

=> {03432640} <noun.artifact> implement -- (instrumentation (a piece of equipment or tool) used to effect an end)

=> {03443493} <noun.artifact> instrumentality, instrumentation -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)

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

=> {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> {00002645} <noun.Tops> whole, whole thing, unit -- (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, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
 => {02970507} <noun.artifact> compound lever -- (a pair of levers hinged at the fulcrum)
 => {03522749} <noun.artifact> lever -- (a rigid bar pivoted about a fulcrum)
 => {02691238} <noun.artifact> bar -- (a rigid piece of metal or wood; usually used as a fastening or obstruction or weapon; "there were bars in the windows to prevent escape")
 => {03432640} <noun.artifact> implement -- (instrumentation (a piece of equipment or tool) used to effect an end)
 => {03443493} <noun.artifact> instrumentality, instrumentation -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
 => {00019244} <noun.Tops> artifact, artefact -- (a man-made object taken as a whole)
 => {00016236} <noun.Tops> object, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
 => {00002645} <noun.Tops> whole, whole thing, unit -- (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, physical object -- (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 -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

