# O delineamento de uma Ontologia com vistas ao tratamento computacional: uma proposta para o subdomínio da Ecologia de Comunidades

Claudia ZAVAGLIA (UNESP-IBILCE/CNPq) Deni Yuzo KASAMA (UNESP-IBILCE/FAPESP)

RESUMO: Este artigo apresenta uma forte vertente da Lingüística Computacional, qual seja, o Processamento de Línguas Naturais por meio de Ontologias. Tratamos aqui, mais especificamente, de ontologias de domínios restritos, no caso o da Ecologia de Comunidades. Tentamos demonstrar, a partir de bases teóricas, desde a Lexicologia e Terminologia até os recentes estudos sobre a pesquisa ontológica, como se dá na prática a construção de uma ontologia e sua formalização. Desse modo, relatamos os passos que seguimos e as decisões tomadas para que o delineamento de tal ontologia atendesse a critérios básicos estabelecidos por pesquisadores atuais da Engenharia Ontológica, a fim de servir a um tratamento computacional eficiente da estruturação arbórea proposta.

ABSTRACT: This paper presents a strong tendency in Computational Linguistics: the design of ontologies to be reused in Natural Language Processing. We discuss specific ontologies of restricted domains, in this case, the domain of Ecology of Communities. We attempt to demonstrate, from a theoretical basis, from Lexicology and Terminology to recent studies on ontological research, how the construction of an ontology and its formalization work in practice. Thus, we relate the steps that we followed and the choices that we made so that the design of this ontology could meet basic criteria standardized by current researches into Ontological Engineering in order to and serve an effective computational treatment of the proposed tree-like structure.

## 0. Introdução

Por meio de uma investigação teórico-prática, desenvolvemos o presente trabalho embasando questões de ordem prática em linhas teóricas concretas que, efetivamente, forneceram subsídios para nossa pesquisa.

A parte teórica consistiu, sobretudo, na busca em livros, artigos e afins sobre a problemática da construção de ontologias (Genéricas e Especializadas), o que hoje se considera eficaz e de relevância para estudiosos que se interessam pelo tratamento

computacional de dados lingüísticos, seja em Lexicografia e Lexicologia, Terminologia e Terminografia, Processamento de Línguas Naturais (doravante PLN) e suas ramificações, tais como: Tradução Automática, Recuperação da Informação, Motores de Busca, Etiquetadores, Desambiguadores, entre outros.

Com relação à parte prática, concentramos nossos esforços no delineamento do esquema arbóreo conceitual do domínio da Ecologia (subdomínio das Ciências Biológicas). Estivemos restritos à subdivisão da "Ecologia de Comunidades" (EC), dado que essa pesquisa faz parte de um projeto maior em desenvolvimento na UNESP/IBILCE que abarca as outras subdivisões da Ecologia, ou seja, "Ecologia de Populações" – EP e "Ecologia de Ecossistemas" – EEc (tais projetos estão sendo fomentados pela FAPESP, concomitantemente a este).

Por meio do agrupamento dos itens lexicais<sup>1</sup>, conseguimos traçar o esquema arbóreo-conceitual da Ecologia de Comunidades. Até o presente momento, pudemos constatar a relevância dos estudos da linguagem em domínios especiais e como a sua demarcação e identificação tem possibilitado o nosso trabalho, cujo escopo é a disponibilização de dados lingüísticos em Bases de Conhecimento Lexical para PLN. Em uma segunda etapa, trataremos da tradução do esquema conceitual para a língua italiana visando preencher uma das muitas deficiências do mercado lexicográfico computacional bilíngüe brasileiro, qual seja, a confecção de esquemas conceituais especiais (máxime bilíngüe italiano) para a produção de bases de dados computacionais para sistemas em PLN. Cumpre ressaltar que a partir do momento que o uso de Ontologias preconiza um entendimento e uma base de informações que seja comum a uma certa comunidade, uma das tarefas mais densas, e que requer um olhar especial sobre ela, é justamente, o momento de se traçar e de se conceituar as classes hierárquicas que farão parte dessa estrutura ontológica: como classificar superclasses, subclasses, instâncias, slots? E ainda: de onde resgatar as informações necessárias para o delineamento desse mapa ontológico e, consequentemente, como proceder, reagir, atuar? Partindo-se do pressuposto de que uma Ontologia é uma fotografia registrada de um dado instante, de uma certa realidade, por um certo alguém, o ontólogo deverá se conscientizar de que a modelagem das estruturas traçadas terá de se adaptar aos limites impostos pela ferramenta computacional utilizada para a modelagem de dados. Há que se relatar ainda, que existe na tarefa de delinear, demarcar, limitar, caracterizar, definir, conceituar conhecimento, seja ele de qualquer universo de discurso, muito mais arte do que ciência (Mangan, s.d.). Com efeito, o "colocar a mão na massa" ainda é o ponto

crucial de qualquer pesquisa que requer modelagem de conhecimento, uma vez que é feita, com resultados positivos, essencialmente por humanos.

## 1. Objetivos

Tencionamos apresentar uma proposta para o tratamento de itens lexicais especializados, com o objetivo específico de traçar a Ontologia do subdomínio da Ecologia de Comunidades, e implementá-la na ferramenta computacional Protégé-2000<sup>2</sup>.

## 2. Investigação teórica: da Lexicologia e Terminologia às Ontologias

Uma vez que na experiência ontológica busca-se definir os conceitos, procuramos incluir na bibliografia textos que serviriam de base para a fundamentação teórica do trabalho que tratassem da lexicologia. Relacionamos "definição" com "lexicologia", em que a primeira é o escopo do trabalho lexicológico, sendo, contudo, mais antiga que a lexicologia, propriamente dita. Há uma capacidade inerente ao homem de "perceber diferentes níveis de abstração e passar facilmente de um a outro" (Picoche, 1992:138). É espontâneo o trabalho realizado pela mente humana com conjuntos de sinônimos, bem como a percepção de seus equivalentes. De natureza cognitiva, temos ainda a (i) expansão e (ii) a condensação, em que (i) equivale a "exprimir em várias palavras o que acaba de ser dito em uma palavra", ao passo que (ii) corresponde a "resumir em uma palavra o que acaba de ser dito em várias palavras". Aristóteles dissertou sobre a arte de definir há mais de dois milênios, e há mais de quatro séculos são feitos dicionários unilíngües, isto é, recolha de definições.

Com relação às definições nos dicionários, essas possuem uma finalidade prática que serve para atender um usuário que não possui um domínio completo do vocabulário de um discurso. Para aqueles que definiam no século XVII, havia uma exigência de coerência e totalidade, não se preocupando com a forma de suas definições. Simulavam situações de total incompetência por parte do usuário do dicionário em relação ao definido. Ao que não se conhecia, a tendência do usuário era dar uma equação sêmica, que era uma paráfrase sinonímica composta por outras palavras da mesma língua. Em outras palavras, o dicionário unilíngüe é uma vasta rede de palavras cambiáveis entre si; um sistema fechado e circular. Richelet (século XVII) diz que uma definição é "um discurso que explica claramente a natureza de algo" (apud Picoche, 1992). Convém relatar ainda a questão da natureza da definição. Comumente, tratamos da definição

enciclopédica e da definição lingüística, bem como do questionamento sobre qual delas seria mais útil para o delineamento de um esquema conceitual que representará, de maneira formal, o conhecimento. A definição enciclopédica é da ordem da significação, enquanto a definição lingüística é da ordem do valor. Esta trata somente daquilo que é útil para o funcionamento correto da língua e não daquilo que é necessário para o conhecimento exaustivo do tipo de referência a qual remete à palavra. Aquela, por sua vez, não é feita por lingüistas, mas por especialistas ligados ao termo em questão.

Por meio da lexicologia e de seus estudos, é possível entender questões importantes como a circularidade das definições nos dicionários, e como isso funciona em Lingüística Computacional (LC). De fato, em LC, as definições devem ser as mais concretas e concisas possível, dada a especificidade de cada definição singular. E ainda: um substantivo se define por um substantivo, um verbo por outro verbo, um advérbio por um outro advérbio e um adjetivo por um adjetivo ou por paráfrases de valor equivalente, segundo os preceitos de Picoche (1992: 142). Para a definição das entradas, nos amparamos, além disso, no conceito de genus chamado em Hathout (1996: 149) de "hiperônimo lexicográfico". Zavaglia (2002) trata da extração do genus terminus como etapa "importante e essencial para que se realize uma organização taxionômica de um repertório lexical, segundo uma estrutura hereditária em termos de hiperonímia" (p. 168). Como exemplo, temos a definição dada para o conceito "Biota" (subclasse de "Ecologia de Comunidades"): Conjunto de todos os seres vivos (flora e fauna) de uma região. O genus terminus "Conjunto" se classifica, segundo Calzolari et al. (1993), como um genus vazio.

Em nosso trabalho, duas concepções importantes e constantes são os conceitos de "hiperonímia" e "hiponímia". Ao tratarmos de "hiponímia" e "hiperonímia", estamos nos referindo a uma relação semântica existente entre significados. Para "hiponímia" temos ainda os nomes "subordinados" ou "subconjunto" e para "hiperonímia" os nomes "superordenados" ou "superconjunto". Um hiperônimo, ou "lexema superordenado", é também chamado de "arquilexema". No âmbito das ontologias, optamos pelos nomes "superclasse" e "subclasse" para hiperônimo e hipônimo, respectivamente, uma vez que essa é a terminologia empregada na literatura especializada. A relação de "hiponímia" e "hiperonímia" é conhecida como *IS-A* (é\_um) que se caracteriza por um sistema de herança em que um hipônimo (termo mais específico) faz parte de um grupo denominado hiperônimo (termo mais genérico). Podem ocorrer vários hipônimos para um mesmo hiperônimo. Picoche (1992: 98) realiza a distinção entre um hipônimo e o

seu hiperônimo da seguinte maneira: para a classe *piano* e a classe *instrumento*, a primeira se chama "espécie" e a segunda "gênero". Sendo assim, *piano* é uma espécie do gênero *instrumento*, sendo que "todos os *pianos* são *instrumentos*", mas "nem todos *instrumentos* são *pianos*". Esse processo de inclusão é feito pelo ponto de vista da extensão. Logo, a palavra *instrumento* nos remete a um maior número e nomes de objetos do que a palavra *piano*. Em suma, no subdomínio da Ecologia de Comunidades essa relação pode ser exemplificada pela classe "Harmônica" (que é um tipo de relação interespecífica) e sua subclasse "Comensalismo". Então, temos "Comensalismo" como subclasse de "Harmônica" e ainda, "Harmônica" como superclasse de "Comensalismo".

Por sua vez, os estudos no campo da Terminologia foram de extrema importância, uma vez que para o delineamento da ontologia foi necessária uma intensa busca em um córpus<sup>3</sup> específico, a fim de reconhecer as classes e subclasses que dão forma à ontologia. Acrescenta-se a isso, a necessidade de se estabelecer os conceitos para essas classes e subclasses, e ainda, determinar as relações existentes entre elas. Tal uso da terminologia é tratado por Sager (1993, *apud* Almeida, 2000) o qual classifica essa abordagem terminológica como terminologia-instrumento<sup>4</sup>.

Ontologias têm sido usadas na representação de informações que possuam um entendimento semântico comum de situações variadas do mundo real. Há que se considerar, porém, que a descrição detalhada de informações gerais do mundo ao nosso redor não deve ser considerada uma tarefa banal, ao contrário. Com isso, tem-se preferido representar o conhecimento de domínios específicos, no qual apenas uma parte de mundo é representada e formalizada, dado que são veiculadas informações restritas, porém com maior riqueza de detalhes e consenso entre os seus falantes. Com efeito, almeja-se que os conceitos-chave do domínio restrito sejam definidos senão diretamente por especialistas na área, com a ajuda e a consultoria deles. Em consonância, Vasconcelos (2003: 16) diz que: "As ontologias se propõem a capturar domínios de conhecimento de forma genérica, para fornecer um entendimento semântico desses domínios que poderá ser utilizado e compartilhado por diversas comunidades e aplicações".

Ao trabalharmos com ontologias, estamos tratando paralelamente de redes semânticas (Quillian, 1966 *apud* Mazzeranghi, 1991), i.e., uma classe de formalismos de representação do conhecimento que utiliza "nós" para representar objetos, conceitos ou situações que são ligados por arcos orientados – representando relacionamentos entre esses nós (Cappelli, 1990: 2). O significado de um conceito depende do modo como ele

é associado a outros conceitos. Podemos dizer que as redes semânticas são baseadas no modelo psicológico explícito da memória associativa humana, ou seja, na identificação de objetos por meio do reconhecimento, i.e., classes. Nesse sentido, os atributos ou características de um objeto ou classe podem conter várias informações para melhor representá-los (op. cit.). Nos sistemas de representação de frames, isto é, uma série de blocos, ditos slots, deparamo-nos com a problemática em relação ao poder terminológico e ao assercional dos formalismos de representação, respectivamente: "a capacidade de representar o conhecimento em relação aos objetos e as classes que constituem as redes complexas de conceitos nas quais se articula o conhecimento de um domínio" e "o poder de especificar asserções complexas do domínio", além de explicitar nelas os termos do conhecimento, quantificação, relações etc. A diferença entre os dois tipos consiste na capacidade de representar o conhecimento vinculado por meio de grupos nominais (terminologia) e por meio de uma frase complexa (asserção). Segundo Cappelli (1990), o modelo clássico da terminologia baseia-se em Redes Semânticas de Hereditariedade Estruturada (SI-Nets) cujas funções permitem criar e utilizar uma base de conhecimento.

Finalmente, relatamos as considerações sobre os estudos ontológicos baseados em trabalhos que têm sido desenvolvidos na literatura e que se encontram aplicados na prática. O uso de ontologias como forma de representar o conhecimento tem se mostrado eficaz, sobretudo, no meio computacional. Em geral, como qualquer aplicativo convencional, os sistemas baseados em conhecimento são gerados a partir de determinadas especificações como linguagens de programação e protocolos de rede. Mas isso não é suficiente, uma vez que sistemas baseados em conhecimento requerem particularidades para operarem, i.e., necessitam de uma representação formal do conhecimento.

Segundo Ortiz (2000: 2), a semântica baseada em ontologia em PLN serve: (i) de suporte para a tradução de lacunas léxicas; (ii) de suporte para a desambiguação, tanto léxica como estrutural; (iii) para um tratamento adequado do fenômeno da sinonímia. Em consonância, Tiscornia (1995: 1) diz que para o desenvolvimento de aplicativos computacionais é necessária a individualização dos modelos dos mecanismos cognitivos humanos e do processo de formação do conhecimento, e que a ontologia formal, uma das mais recentes abordagens da modulação do conhecimento, é, na verdade, uma revisitação de teorias filosóficas e lingüísticas. Nesse sentido, as categorias ontológicas são "subdivisões de um sistema de classificação utilizadas para

catalogar conhecimento, por exemplo, em uma base de dados" (Tiscornia, 1995: 4). Para Gruber (1993), ontologias compartilham e reutilizam o conhecimento de mundo. Com efeito, segundo o autor "o termo ontologia significa uma especificação de conceitos, isto é, uma ontologia é uma descrição formal dos conceitos e das relações existentes entre estes em um determinado domínio" (*apud* Braga et al., 2002).

Ressaltamos que, atualmente, no campo do PLN, principalmente em Sistemas de Bases de Conhecimento Lexical, é consensual que a inclusão desse tipo de repositório semântico, i.e., do tipo ontológico para a representação do significado, é essencial. Existe a necessidade de se oferecer de forma estruturada e organizada um léxico comum utilizado em conformidade por uma determinada comunidade. O uso de ontologias tem sido amplamente empregado em representações do conhecimento de domínios restritos, máxime para sistemas de busca de informação e indexação de documentos, nos quais a sua aplicação pode ser mais eficaz por tratar, justamente, de conjuntos léxicos de número finito. Em uma Base de Conhecimento Lexical, por exemplo, o uso de uma ontologia pode servir como recurso de apoio à informação contida no repositório lexical dessa base, para que seja possível resgatar o significado de um item léxico de forma unívoca. De fato, os recursos lingüístico-classificatórios que a utilização de uma ontologia pode oferecer para um lingüista e/ou lexicógrafo servem para que ele possa dar conta de individualizar singularmente, dentre os diversos significados ou diversas acepções atribuíveis a um mesmo item lexical, o significado pertinente no interior do feixe de sentidos polissêmicos que a palavra comporta, neutralizando, dessa maneira, a polissemia própria a esse mesmo item lexical.

A taxonomia mais comum de uma ontologia é a do tipo hereditária em que classes e subclasses mantêm relações hierárquicas em forma de árvores, uma vez que temos axiomas do tipo:

- (i) "Todo animal terrestre é um animal, que por sua vez, é uma entidade viva, uma entidade concreta e uma entidade": *um cachorro é um animal, ser vivo e concreto*;
- (ii) "Toda área é uma localização, que por sua vez, é uma entidade concreta e é uma entidade": *uma praia localiza-se em uma cidade que é uma entidade concreta*.

Os membros de uma mesma classe ou subclasse carregam algumas propriedades em comum: na subclasse "animal terrestre", por exemplo, seus membros "boi", "cachorro", "coelho" possuem patas, andam, não falam; propriedades em comum são, portanto, herdadas pela inserção de uma palavra em uma ou em outra classe.

Segundo a sua natureza, as ontologias podem ser classificadas em dois grandes tipos:

- (I) Ontologias Genéricas que, em termos gerais, são aquelas empregadas para a representação de grandes e substanciosos conjuntos de objetos e suas relações. Nesse sentido, elas buscam representar o senso comum de uma comunidade sócio-lingüística, na medida que representam o conhecimento enciclopédico existente de maneira intrínseca nesse grupo lingüístico. De fato, o conjunto de opiniões e modos de agir que são impostos aos indivíduos de um grupo social, em uma determinada época, são comumente aceitos de modo passivo como verdades e comportamentos próprios dessa comunidade. O processo de elaboração desse tipo de ontologia é lento e laborioso, uma vez que o número de informações contido no universo é infinito. Por outro lado, a construção desse tipo de ontologia é mais vantajosa, já que abarca diferentes áreas do conhecimento, isto é, unifica variados conhecimentos de diversos domínios simultaneamente. Além disso, essas ontologias podem ser aplicadas a domínios especializados pelo simples acréscimo de axiomas específicos dos domínios em questão. Uma primeira tentativa de construção de uma ontologia genérica para o português do Brasil pode ser vista em detalhes em Zavaglia (2002); e
- (II) Ontologias Específicas que são empregadas para a representação de conjuntos de objetos e suas relações de um domínio reduzido e específico. Com isso, elas buscam representar o consenso de um grupo lingüístico de especialistas de uma área restrita e especial, dado que representam o conhecimento terminológico existente de maneira contundente nesse grupo lingüístico. Desse modo, o conjunto de opiniões desses falantes é moldado e fabricado de uma maneira artificial, na medida que os indivíduos aprendem e apreendem sua maneira de pensar sobre determinado assunto, dentro de um domínio fixo e pré-estabelecido. O processo de elaboração desse tipo de ontologia pode ser considerado mais rápido e simples, já que o número de informações contido em um domínio específico é reduzido e limitado. Entretanto, uma das desvantagens da construção desse tipo de ontologia é, justamente, a restrição de informações, isto é, a limitação do conhecimento a somente uma área do saber. Por outro lado, qualquer ontologia específica poderá ser modificada com o objetivo de se tornar genérica pelo simples acréscimo de axiomas de outros domínios e do senso comum.

No que diz respeito aos modos de construção de ontologias, são levados em consideração dois tipos de elaboração: (i) a abordagem *top-down* que define

primeiramente os conceitos do senso comum e, em uma segunda etapa, abarca o conhecimento especializado; (ii) a abordagem *bottom-up* que parte de um número prédefinido de bases especializadas para, em seguida, integrar os conceitos gerais que fazem parte do senso comum.

A longo prazo, podemos compartilhar uma ontologia por meio de uma rede, uma vez que seu delineamento demanda esforços e tempo (sobretudo em ontologias genéricas), o que permitiria a pesquisadores do mundo todo acesso a essas redes semânticas. Tendo isso em vista, é bom lembrar que para se manter uma padronização formal que possa dar conta do compartilhamento dessas ontologias por meio de tal rede, devem ser estabelecidos critérios confiáveis, consistentes e exeqüíveis para a sua produção. Nesse sentido, "conceitualização" é a palavra chave ao representar o conhecimento de maneira formal. Objetos, conceitos e outras entidades existentes em determinada área do conhecimento (principalmente no caso de ontologias específicas) e as relações entre elas devem ser conceitualizadas. Tais conceitos nada mais são do que uma visão simplificada e resumida de mundo. Logo, podemos dizer que uma ontologia é "uma especificação explícita da conceitualização" (Gruber, 1993: 1).

Vasconcelos (2003: 17) entende que:

"Ao definirmos ontologias poderemos definir acordos entre usuários de informação e os fornecedores de informação para padronização da utilização dos dados na Web, diminuindo os problemas de consistência no intercâmbio de dados nesse ambiente. Esses acordos são chamados Compromissos Ontológicos (*Ontological Commitments*) e permitem que somente determinados significados de um domínio sejam capturados numa dada especificação. O intercâmbio de dados é padronizado e será baseado no modelo de especificação que foi criado".

Para que uma ontologia seja elaborada, decisões devem ser tomadas quanto a forma de esboçá-la. Gruber (1993: 2-3) estabeleceu um conjunto preliminar de critérios de delineamento de ontologias, que reportamos a seguir: (a) **Clareza** (*Clarity*), que trata, justamente, da objetividade da definição, uma vez que "a ontologia deve efetivamente comunicar o significado pretendido dos termos definidos" (p. 2). As definições em uma ontologia devem ser independentes de contextos sociais ou computacionais; dá-se preferência àquelas completas em detrimento das parciais, e a sua documentação deve ser em língua natural; (b) **Coerência** (*Coherence*), i.e., as definições existentes em uma ontologia devem ser coerentes. Ao definirmos, podemos informalmente citar um exemplo. Se a definição e o exemplo estiverem, porventura, em

desacordo, a ontologia estará incoerente; (c) Extensibilidade (Extendibility), ou seja, uma ontologia deve ser capaz de, a qualquer momento, receber novos termos de maneira que não seja necessária a revisão das definições já existentes; (d) Mínima codificação (Minimal encoding bias): "Uma codificação resulta quando escolhas representativas são puramente feitas para a conveniência da notação ou implementação" (p. 3). Tais codificações devem ser minimizadas já que os agentes do conhecimento compartilhado podem ser implementados em diferentes sistemas e estilos de representação; (e) Mínimo compromisso ontológico (Minimal ontological commitment) em que o número de compromissos ontológicos deve ser minimizado, a fim de maximizar o reuso. Eles devem ser suficientes para suportar as atividades de conhecimento compartilhado almejadas.

No que concerne, ainda, à busca por diretrizes na construção de ontologias, a contribuição de Ortiz (2000) revela-se pertinente, na medida que o autor reporta uma série de aplicações práticas quanto ao uso de ontologias, sobretudo no campo da tradução automática. Ressalta que "este é o princípio fundamental da semântica baseada em ontologia: fundamentar os significados das unidades léxicas em um sistema de símbolos estruturados que contém uma interpretação determinada e com uma motivação independente" (Nirenburg, Raskin & Onyshkevych, 1995 apud Ortiz, 2000: 2). Esse autor propõe, assim, algumas recomendações quanto à inclusão de novos conceitos na ontologia. Vejamos: (i) não especificar instâncias como conceitos na ontologia; as regras para distinguir uma instância de um conceito são as seguintes: a) verificar se a entidade pode ter sua própria instância; neste caso não se trata de uma instância, mas sim de um conceito e b) verificar se a entidade tem uma posição fixa no espaço e/ou no tempo; se assim for tem-se uma instância, do contrário tem-se um conceito; (ii) não decompor conceitos em outros conceitos só porque é possível fazê-lo. A menos que se mostre necessário, os conceitos deveriam permanecer os mais englobadores possível; (iii) não acrescentar um conceito se já existe um suficientemente próximo ou ligeiramente mais geral que o conceito em questão; (iv) não acrescentar conceitos que refletem entradas léxicas de uma língua específica, ou seja, manter a independência de uma língua e (v) não acrescentar conceitos muito específicos, a não ser que seja absolutamente necessário. Em geral, um conceito deveria corresponder-se com mais de um lexema.

A partir dessas reflexões, nos dedicamos a estruturar e conceituar uma Ontologia para a Ecologia de Comunidades, como explanado no capítulo seguinte.

### 3. O delineamento da Ontologia

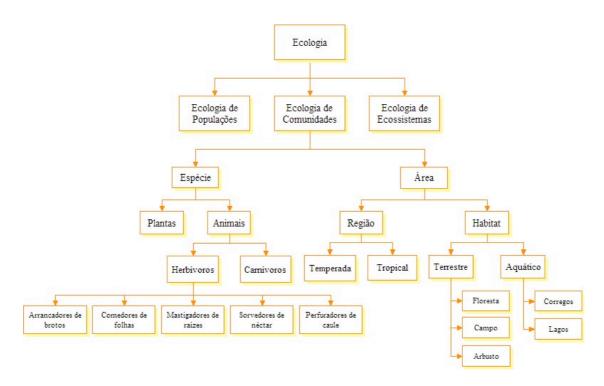
Por meio dos estudos teóricos realizados previamente ao desenvolvimento prático do trabalho, foi possível contemplar os vários conceitos ligados aos estudos ontológicos, o que nos permitiu realizar o trabalho de maneira mais desenvolta e concisa. Contudo, algumas questões importantes precisaram ser, vez ou outra, relembradas no decorrer do trabalho. Exemplo disso, é a questão da hereditariedade dos conceitos que uma subclasse carrega consigo de sua precedente (i.e., que herda de sua superclasse) e como isso é fundamental para a projeção de árvores ontológicas (Zavaglia, 2002: 175). A importância de tal questão foi acentuada, quando atentamos que para demarcar e classificar a subclasse "Y" de uma classe "X", os conceitos de ambas deveriam ser confrontados para que "Y" pudesse herdar os conceitos de "X", em concordância com os critérios que levamos em conta em nossas práticas metodológicas.

A título de ilustrar essa questão, tomemos como exemplo a classe "Interações Intraespecíficas", que se encontrava arrolada em nossa estruturação. Em entrevista com o consultor (Cf. o seu papel na próxima seção), concluímos que tal divisão de significação não fazia parte de "Ecologia de Comunidades", uma vez que a caracterização de ambas não mantêm traços de correspondência. De fato, a Ecologia de Comunidades trata das relações entre os seres de espécies diferentes, ao passo que as "Interações Intraespecíficas" tratam de relações entre seres de uma mesma espécie. Logo, a subdivisão "Interações Intraespecíficas" — que mais corretamente, deve ser abordada pela "Ecologia de Populações" — não herda o conceito que a sua superclasse "Ecologia de Comunidades" traz consigo. Por conseguinte, a excluímos de nossa ontologia.

## 3.1. O esquema arbóreo

Uma vez que estamos no domínio especializado, uma das etapas essenciais para a validação de conceitos e de termos a serem incluídos em uma ontologia é a análise de todo o material formulado por um assessor da área em questão. No nosso caso, esse consultor é o Prof. Dr. Luís Zanini Branco, do Departamento de Zoologia e Botânica, da Universidade Estadual Paulista - UNESP, Câmpus de São José do Rio Preto, especialista em Ecologia. Efetivamente, no processo de delineamento da ontologia em si, várias foram as vezes em que reformulamos o esquema traçado, dado que após as entrevistas com o consultor, as informações digeridas determinavam as alterações de modo decisivo. Exemplo disso são as subclasses "Arrancadores de Brotos",

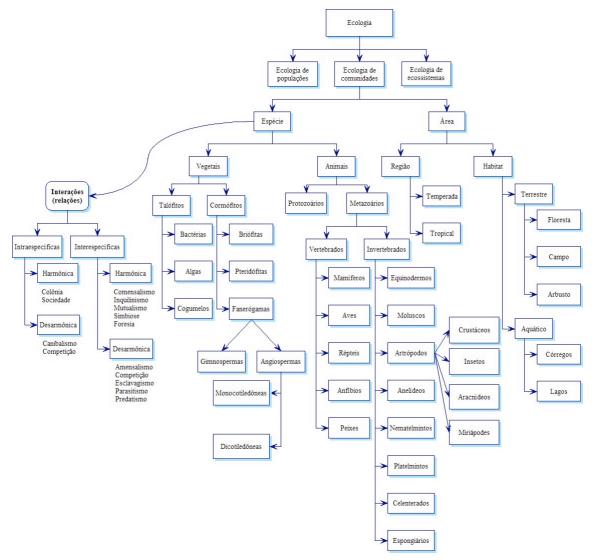
"Comedores de folhas", "Mastigadores de raízes", "Sorvedores de néctar" e "Perfuradores de caules" atribuídas à classe "Herbívoros" (Cf. **Estrutura-arbórea (1): Primeiro esboço**) que eliminamos, uma vez que se revelaram "termos" e não subclasses da Ecologia de Comunidades. Por sua vez, as classes "Região" e "Habitat" ainda estavam por ser ampliadas, passo a passo, conforme desenvolvíamos o trabalho junto ao córpus<sup>5</sup>. Outra alteração fundamental foi a modificação do conceito "Plantas" para "Vegetais", uma vez que esse último abrange tanto as plantas (reino *plantae*) quanto o reino *protista*. Vejamos a estruturação inicial:



Estrutura-arbórea (1): Primeiro esboço

No anseio de dar forma a algumas categorias, sobretudo às categorias "Vegetais" e "Animais", iniciamos uma pesquisa em livros de Botânica e Zoologia, que acabaram resultando no esquema **Estrutura-arbórea** (2). Adicionamos ainda, a classe "Interações", fortemente tratada pela Ecologia de Comunidades. O problema inicial quanto a essa categoria foi como e onde inseri-la na ontologia, de modo que pudéssemos expressar, de maneira fiel, o seu conceito. Uma vez definido o seu campo nocional, qual seja, "Ação exercida mutuamente entre indivíduos", a inserimos como hipônimo de "Espécies", expressando, entretanto, uma relação diferente de "Plantas" e "Animais". Com efeito, enquanto essas duas últimas expressavam a relação

*é\_um\_tipo\_de* (em relação à "Espécie"), "Interações" expressa a relação télica<sup>6</sup> *é\_a\_atividade\_de*. Vejamos a nova configuração:



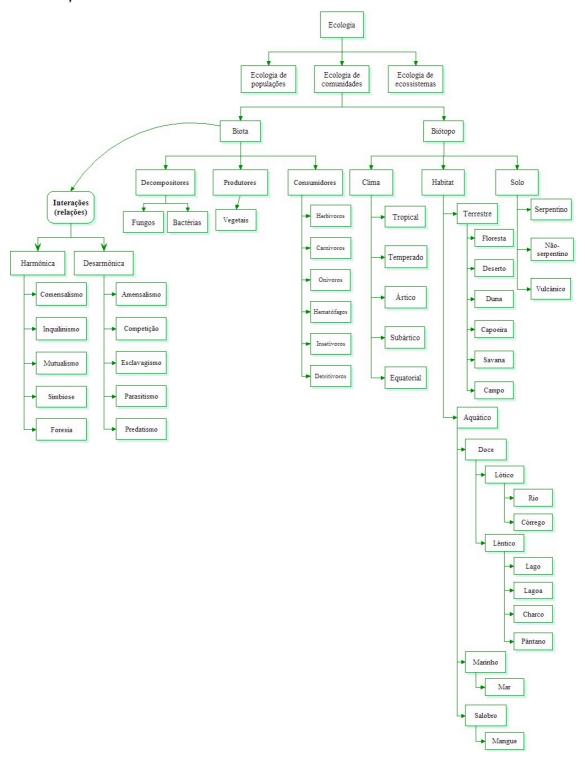
Estrutura-arbórea (2): uma ampliação da ontologia inicial

Ao contemplarmos a ontologia ampliada, a conclusão foi insatisfatória dada à fuga que isso representava em relação ao domínio da "Ecologia de Comunidades". Na busca por um novo rumo para nossa estruturação arbórea, retomamos o esquema inicial (Estrutura-arbórea (1)) e concluímos que para a categoria "Espécie", as subclasses "Herbívoros" e "Carnívoros" estavam mais de acordo com a proposta da "Ecologia de Comunidades", uma vez que todo o córpus e os textos pesquisados na Internet tratam sempre das relações dos seres vivos com a natureza e entre si, bem como de seus hábitos alimentares. A primeira importante alteração realizada foi, então, alterar o conceito "Espécie" por "Biota", uma vez que "Espécie" abarca uma significação bastante restrita, enquanto "Biota" é mais abrangente, além de expressar exatamente

aquilo que gostaríamos de descrever, ou seja, "Conjunto de todos os seres vivos (flora e fauna) de uma região". Tal proposta nos foi feita pelo consultor, e, por meio de pesquisas em glossários especializados e em dicionários, concluímos que tal categorização se aplicava melhor, de fato, ao domínio da Ecologia de Comunidades. O mesmo efetuou-se para a classe "Área": em um primeiro momento foi definida como "Superfície ocupada por uma comunidade ou táxon", ao mesmo tempo que preferimos alterá-la para "Biótopo", cuja definição é "Conjunto de circunstâncias favoráveis à vida e ao desenvolvimento de uma biocenose". Além disso, o conceito de "Biótopo" trata especificamente do tipo de "clima", do tipo de "habitat" e do tipo de "solo" de uma dada região. Sendo assim, ampliamos tal subclasse para que ela pudesse expressar a noção de biótopo. Para tal efeito, alteramos ainda a classe "Região" para "Clima", uma vez que o significado expresso por "Região" é bastante subjetivo: "Superfície ocupada por uma comunidade ou táxon semelhante ou proximamente correlacionado". Dessa forma, para a nova proposta, consideramos que a Ecologia de Comunidades trata tanto dos seres vivos quanto de seus hábitos alimentares (no caso dos "Vegetais", o critério adotado para a sua sistematização foi sua característica autotrófica, ou seja, sua capacidade de produzir seu próprio alimento) bem como de seus papéis na cadeia alimentar. Foi de nosso interesse descrever os "consumidores" de acordo com o que normalmente se convenciona para tal grupo, isto é, classificá-los como "consumidor primário, secundário e terciário" (e ainda, quaternário). Entretanto, cabe-nos uma pergunta: quem deve ser considerado um consumidor primário? De acordo com a definição dessa categoria, o consumidor primário é o "ser vivo que se alimenta de seres autótrofos, i.e., de vegetais". Porém, quando buscamos uma definição para o conceito de consumidor secundário, deparamo-nos com a seguinte significação: "ser vivo que se alimenta de um consumidor primário", ou seja, é difícil determinar em que nível da cadeia alimentar um ser vivo pode se encontrar, dado que ele pode ser considerado, em certos contextos, como qualquer um dos tipos de consumidores. Tomemos como exemplo disso, a seguinte situação: uma *capivara*, que se alimenta de vegetais (logo, consumidor primário), é devorada por uma cobra que fica sendo um consumidor secundário. Um homem pode se alimentar dessa cobra (hábito pouco disseminado entre os brasileiros, porém real em alguns países) e ser um consumidor terciário. Contudo, esse mesmo homem poderia se alimentar da capivara (uma realidade cultural mais próxima da nossa) e ser classificado como consumidor secundário. Observamos que tal distinção entre os consumidores depende da situação que está sendo observada, e tal

fato, nos fez optar pela classificação ora apresentada, pois não deixa de observar os hábitos alimentares dos consumidores e não depende de uma dada situação para classificar os seres.

Como última versão de nossa proposta de ontologia, observemos, a seguir, a sua estruturação arbórea:



Estrutura-arbórea (3): Proposta final

## 3.2. A elaboração das definições das classes e subclasses

Valendo-nos dos pressupostos teóricos de Picoche (1992) e Zavaglia (2002), para a elaboração da definição dos conceitos, procedemos com a conceituação das classes e subclasses pré-estabelecidas, processo esse que se deu cognitivamente conforme a ontologia ia sendo delineada. Para que uma classe tivesse uma subclasse, essa última deveria herdar o conceito da primeira. Dessa maneira, tais definições foram feitas com base nas informações que extraíamos do nosso córpus, além daquelas provindas de dicionários, glossários especializados, livros e consultas a textos na Internet. Observemos algumas dessas classes definidas:

1. Ecologia – Ciência através da qual estudamos como os organismos (animais, plantas e microorganismos) interagem entre si ou com o meio orgânico ou inorgânico no qual vivem. 1.2. Ecologia de comunidades – Ramo da ecologia que enfoca a associação de populações que interagem entre si. 1.2.1. Biota – Conjunto de todos os seres vivos (flora e fauna) de uma região. 1.2.1.0. Interação interespecífica - Ação exercida mutuamente entre indivíduos de espécies diferentes. 1.2.1.0.1. Harmônica – Ação exercida mutuamente entre indivíduos de espécies diferentes onde ocorre benefício para uma das espécies envolvidas ou ambas.

A seguir, vejamos como procedemos ao armazenamento em uma ferramenta computacional das informações e dos dados da nossa estruturação arbórea. Para tanto, descrevemos, primeiramente o *software* e, em seguida, a sua aplicação.

# 4. A implementação da Ontologia: a ferramenta computacional utilizada

Protégé-2000<sup>7</sup> é uma ferramenta computacional integrada, mais especificamente um editor de ontologias, usada para o desenvolvimento de sistemas baseados em conhecimento. Essa ferramenta foi desenvolvida para diferentes linguagens para a *Web* Semântica, dentre as quais *RDF*<sup>8</sup> e *RDF Schema*, que permite a estruturação de informações de um domínio específico e possibilita a comunicação, por meio de um vocabulário comum, entre agentes de *software* e páginas da *Web*. O Protégé-2000 tem por formalismo: a linguagem de representação *OKBC* (*Open Knowledge Base Connectivity*); especificação de restrições por meio de lógica de primeira ordem usando *PAL* (*Protégé Axiom Language*); flexibilidade, uma vez que permite a importação e exportação de representações em formato texto – *RDF Schema* e em bases de dados por meio de *JDBC* (*Java DataBase Connectivity*); geração de código *DAML* (*Darpa Agent Markup Language*) + *OIL* (*Ontology Inference Layer*) para a próxima versão. O Protégé-2000 tem por objetivos: (i) consentir a interoperabilidade com outros sistemas

de representação do conhecimento; (ii) ser uma ferramenta de aquisição de conhecimento que seja fácil de se configurar e manejar; (iii) ser extensível. Seu modelo de conhecimento é representado por meio de *classes* (conceitos no domínio de discurso que constituem uma hierarquia taxonômica), *instâncias* dessas classes, *slots* (que descrevem as propriedades e atributos das classes e instâncias), *facetas* (que são restrições de informações, especificando informações adicionais sobre propriedades). É baseado em *frames*; usa a arquitetura de *metaclasse*, i.e., um *template* que é usado para definir novas classes em uma ontologia (*metaclasse* é uma classe cujas *instâncias* também são *classes*) e possibilita a especificação de herança múltipla e de classes abstratas.

Isso posto, apresentamos os resultados parciais das classes e subclasses delimitadas, estruturadas e conceitualizadas amiúde, respeitando todos os campos de informação previamente propostos. Convém ressaltar que definimos letras maiúsculas para CLASSES e SUBCLASSES e letras minúsculas para os termos que serão implementados no decorrer da pesquisa. Até o presente momento, temos implementadas 64 classes e subclasses, todas elas com os campos preenchidos, a saber:

- *name*: correspondente ao nome da classe ou subclasse;
- documentation: que traz a sua definição;
- role: todas apresentam a informação "concrete";
- template slots: que atribuem as propriedades, por meio de slots, para cada classe ou subclasse;
- antônimo: quando existente, atribui o termo antônimo à entrada;
- *sinônimo*: quando existente, atribui o termo sinônimo à entrada;
- *contexto*: reporta a contextualização do córpus.

Vejamos a implementação da estruturação conceitual, feita manualmente, numa perspectiva semântica:

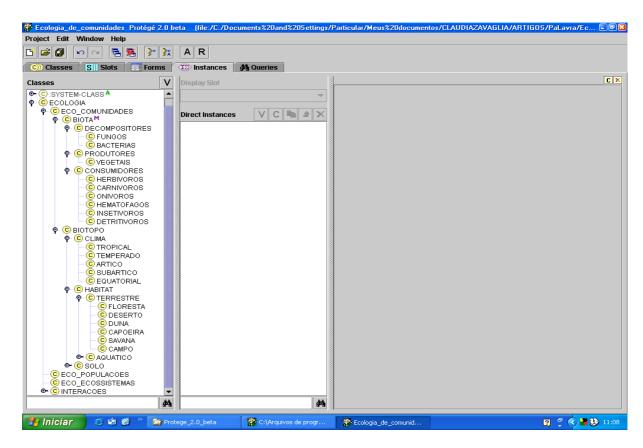


Figura (1): Implementação das Classes e Subclasses da Ecologia de Comunidades

# 5. Considerações finais

O léxico é um dos Recursos Lingüísticos primários no que diz respeito à Engenharia da Linguagem, especialmente para a Engenharia Ontológica. De fato, qualquer sistema aplicativo, para analisar ou processar uma língua natural, não pode prescindir do léxico, seja ele de domínio especializado ou não. Esse repertório lexical, por sua vez, para que seja utilizado por uma máquina, deve conter informações adequadas e codificadas para que o programa computacional ou o algoritmo possa "compreendê-las" e utilizá-las. As informações contidas em um léxico podem ser de vários níveis lingüísticos (morfológico, sintático, semântico, pragmático) e, desse modo, existem vários níveis de codificação para esses diferentes níveis, em se tratando de programas computacionais.

Como exemplificação de uma utilização concreta da Ontologia que propomos em uma máquina de busca, ao entrarmos com as palavras-chave "Comunidades e Ecologia" poderíamos recuperar o conceito expresso por "ecologia de comunidades", "ecologia" e "comunidades" que o usuário busca conhecer, caso seja a sua intenção de pesquisa, ao invés de termos como respostas páginas que somente nos informam sobre:

"Áreas de atuação (em CVs)/atividade/conhecimento"; "grade curricular"; "nomes de disciplinas/programas e linhas de pesquisa de departamentos de universidades/instituto/núcleos (graduação e pós-graduação)"; "sites de professores de ensino fundamental e médio"; "ementas de editais de concurso"; "listagem de projetos", como acontece hoje, em português, ao fazermos uma pesquisa desse tipo.

Em consonância, Melcop et al (2002) dizem que existe uma dificuldade visível e preponderante para se encontrar a informação almejada na *Web*. E acrescentam:

"A maioria dos sistemas de indexação e busca na *Web* utiliza técnicas baseadas em palavras-chave. Tais sistemas retornam uma grande quantidade de apontadores (*links*) para páginas de pouco ou nenhum interesse do usuário. A baixa precisão desses sistemas deve-se, entre outros fatores, à sua pequena capacidade de interpretar o conteúdo dos textos, e, em conseqüência, de contextualizar a seleção de documentos de acordo com o interesse do usuário". (Melcop, 2002: 1)

Nesse mesmo sentido, Rigo & Vieira (2002) apontam para a deficiência existente na Internet para busca de informações, uma vez que as ferramentas de busca empregadas aliam sua pesquisa apenas aos termos indicados pelo usuário. Essa técnica faz com que o resultado das buscas de um usuário por meio de um único termo lhe reporte, além de links que podem possuir informações acerca do conteúdo desejado, uma lista de documentos recuperados cujo termo está presente, mas não representa de forma alguma o significado da busca inicial. Isso se dá porque um mesmo termo ou unidade lexical pode possuir diversos significados em diferentes domínios ou num mesmo domínio, dada a polissemia inerente existente na maioria dos itens lexicais de uma língua natural. Além disso, esses autores afirmam que "o formato utilizado na descrição da grande maioria de documentos disponíveis na Internet (HTML) não é adequado à identificação de seu conteúdo" (Rigo & Vieira, 2002: 1). Acrescentam ainda que para sanar essas deficiências e propiciar ao usuário uma identificação correta do conteúdo de um documento aos seus interesses de busca de informações, são necessários a utilização de ontologias adequadas e o emprego de formatos de descrição baseados em XML<sup>9</sup>. A partir do momento que um termo encontra-se definido em uma ontologia, a sua associação com um outro termo utilizado para o mesmo conceito é possível, desde que ele também esteja descrito na ontologia. Ademais:

"O uso de ontologias permite também a associação de possíveis regras de inferência relacionadas a termos descritos, de modo a possibilitar também uma

outra forma de busca de informações, agora de modo automático, através de agentes de software, onde serão levados em conta possíveis relacionamentos entre as informações encontradas nos documentos pesquisados". (Rigo & Vieira, 2002: 1)

Como suporte à interoperabilidade de informações e de dados, as ontologias podem ser utilizadas como aplicação para banco de dados e recuperação de informação, já que servem de modelos conceituais globais entre entidades e relacionamentos. De fato, em motores de busca, as ontologias servem como grandes esquemas conceituais que suportam o caráter semântico das consultas.

Em comércio eletrônico, por exemplo, o conjunto de informações oferecido pela ontologia pode ser utilizado para unificar e integrar definições de produtos oferecidos pelos mais variados pontos de venda, com um formato padrão e único.

Por fim, acreditamos que para a sistematização de um conjunto de informações terminológicas de um domínio, torna-se fundamental o uso de ferramentas computacionais para a extração de termos. Para o português do Brasil, muitos projetos de construção de repertórios terminológicos ainda utilizam o critério semântico para a extração de termos, em uma abordagem manual a partir de córpus. Ainda que o critério semântico seja adequado, a extração manual é lenta, sujeita à subjetividade e à omissão de termos importantes. Nesse sentido, a extração automática de termos torna-se uma etapa essencial no processo de delimitação semântica de um domínio. Esse trabalho pretendeu revelar a importância de se utilizarem ferramentas computacionais no trabalho prático em Lingüística e suas subáreas Lexicologia, Lexicografia, Terminologia e Terminografia.

Tivemos como objetivo relatar o quão cuidadoso deve ser o trabalho do lingüista-ontólogo no processo de elaboração de estruturas ontológicas, cujo olhar deve ser extremamente atento e direcionado para o campo minado da conceitualização de categorias de mundo e de seus domínios específicos.

Intentamos demonstrar que o uso de uma ferramenta computacional para editar ontologias, como o Protégé-2000, pode e deve ser um desafio cada vez mais constante e presente nas pesquisas lingüístico-teórico-práticas desses pesquisadores, cuja aplicação é bastante variada e diversificada, podendo atender não somente aos ensejos da Lingüística Computacional e da Inteligência Artificial, mas também como ferramenta que armazena dados lingüístico-semânticos em grandes quantidades, cujas aplicações práticas vão desde uma simples inserção de dados para manipulação de conceitos, a

aplicações que podem elaborar glossários especializados mono, bi ou multilíngües, a partir de um pré-processamento de dados, a começar da maneira como foram implementadas as informações na ferramenta.

Nesse sentido, o trabalho desenvolvido justifica-se e torna-se relevante na medida que revelou facetas até então inexploradas por lingüistas em domínios específicos.

### REFERÊNCIAS BIBLIOGRÁFICAS

Almeida, G. M. de B. (2000). *Teoria Comunicativa da Terminologia(TCT)*: uma aplicação. Tese de doutorado inédita, UNESP-Araraquara.

Braga, J. L.; Torres, K. S.; Botelho, F.C. (2002). *Reengenharia e Visualização de Conceitos no WordNet*. Universidade Federal de Viçosa. Disponível em: <a href="http:///www.sbc.org.br/reic/edicoes/2000e2/cientificos/ReengenhariaEVisualizacaDeConceitosNoWordNet.pdf">http:///www.sbc.org.br/reic/edicoes/2000e2/cientificos/ReengenhariaEVisualizacaDeConceitosNoWordNet.pdf</a>>. Acesso em: 28/10/2002

Calzolari, N. et al. (1993). Enconding Lexicographic Definitions as Typed Feature Structures. In: Beckmann, F. & Heyer, G. (eds.). *Theorie und Praxis des Lexicons. Beiträge zu einem Kolloquium über Theorestiche Lexicologie und Praktische Lexicographie*. Berlim: Walter de Gruyter.

<sup>&</sup>lt;sup>1</sup> A extração dos itens lexicais especializados (que geram tanto classes, subclasses, como termos em uma ontologia) foi feita, primeiramente, de forma manual, passando em seguida, a uma extração automática dos termos, para a qual nos utilizamos das técnicas e do extrator de n-gramas elaborado por Teline et al. (2003).

<sup>&</sup>lt;sup>2</sup> Embora exista uma versão beta para o Protégé-2001, no momento deste trabalho, nos valemos da versão que se encontrava disponível para download.

<sup>&</sup>lt;sup>3</sup> Adotamos neste trabalho o termo "córpus" tanto para o singular quanto para o plural, grafado com o acento agudo na vogal tônica, em português, em detrimento do latinismo (ou anglicismo) *corpus* e *corpora*. Essa é a tendência atual da terminologia da Lingüística de Córpus no Brasil. Entretanto, em artigos e livros, encontram-se as duas opções teminológicas em vigor, de acordo com a escolha de cada autor.

<sup>&</sup>lt;sup>4</sup> De acordo com Sager (1993, *apud* Almeida, 2000), o termo terminologia possui duas acepções distintas: (i) Conjunto vocabular característico de determinada área ou subárea: **terminologia-objeto**; (ii) Conjunto de práticas e métodos utilizados na compilação, descrição e apresentação dos termos de uma determinada linguagem de especialidade e ainda o conjunto de premissas e argumentos necessários para a explicação das relações entre os conceitos e os termos, relações essas que são fundamentais para uma atividade terminológica coerente: **terminologia-instrumento**. (Sager, 1993: 22 *apud* Almeida, 2000: 36)

<sup>&</sup>lt;sup>5</sup> Elaboramos um córpus específico da Ecologia de Comunidades que conta, atualmente, com 57.501 ocorrências, armazenadas na ferramenta computacional Folio Views 4.1.

<sup>&</sup>lt;sup>6</sup> As especificações das relações na nossa estrutura ontológica basearam-se nos fundamentos teóricos do Léxico Gerativo de Pustejovsky (1995), mais especificamente da sua proposta de Estrutura *Qualia*. Por questões de espaço, tal teoria não foi pormenorizada neste artigo. Por outro lado, encontra-se detalhada em Zavaglia (2002).

<sup>&</sup>lt;sup>7</sup> http://protege.stanford.edu/

<sup>&</sup>lt;sup>8</sup> Resource Description Framework - Linguagem para decodificar conhecimentos em páginas na Web.

<sup>&</sup>lt;sup>9</sup> eXtensibile Markup Language.

- Cappelli, A. (1988). Terminologia ed Ontologia. Istituto di Linguistica Computazionale, CNR: Pisa.
- Gruber, T. R. (1993). *Toward principles for the design of ontologies used for knowledge sharing*. Presented at the Padua workshop on Formal Ontology, March 1993, to appear in an edited collection by Nicola Guarino. Padova. Disponível em: <a href="http://kslweb.stanford.edu/KSL">http://kslweb.stanford.edu/KSL</a> Abstracts/KSL-93-04.html>. Acesso em: 06/01/2002.
- Hathout, N. (1996). Pour la construction d'une base de connaissances lexicologiques à partir du Trésor de la Langue française: les maqueurs superficiels dans les définitions spécialisées. Cahier de lexicologie Revue Internationale de Lexicologie et de Lexicographie. Vol LXVIII, p. 137-173.
- Mangan, M. A. S.; Murta, L. G. P.; Souza, J. M.; Werner, C. M. L. (s.d.). *Modelos de Domínio e Ontologias*: uma comparação através de um estudo de caso prático em hidrologia. (in mimmeo)
- Mazzeranghi, D. (1991). Uma semantica intensionale per um linguaggio di rappresentazione della conescenza. *ILC KRS 3, CNR-ILC*. Pisa.
- Melcop, T. et al. (2002). Uma Ferramenta para Recuperação e Categorização de Páginas Web para Domínios Específicos. In: Workshop de ontologias para a construção de metodologias de busca na web por conteúdos educacionais XIII SBIE'2002. Anais... Unisinos. São Leopoldo, RS.
- Ortiz, A. M. (2000). *Diseño e implementación de un Lexicón Computacional para lexicografía y Traducción Automática*. Estudios de Lingüística Española. Vol. 9. Disponível em: <a href="http://elies.rediris.es/elies9/index.htm">http://elies.rediris.es/elies9/index.htm</a>>. Acesso em: 14/06/2002.
- Picoche, J. (1992). *Precis de Lexicologie Française*: l'étude et l'enseignement du vocabulaire. Paris: Nathan.
- Pustejovsky, J. (1995). The Generative Lexicon. Cambridge: The MIT Press.
- Rigo, S. & Vieira, R. (2002). Busca de informações auxiliada por ontologia. In: Workshop de ontologias para a construção de metodologias de busca na web por conteúdos educacionais XIII SBIE'2002. Anais... Unisinos. São Leopoldo, RS.
- Santos, E. T.; Barros, L. N.; Valente, V. C. P. (2001). Projetando uma Ontologia de Geometria Descritiva. 15 Simpósio nacional de geometria descritiva e desenho técnico IV International Conference on Graphics Engineering for Arts and Design. São Paulo.
- Teline, M. F.; Almeida, G. M. de B.; Aluísio, S. M. (2003). Extração manual e automática de terminologia: comparando abordagens e critérios. In: TIL 2003 Evento integrante do 16th Brazilian Symposium on Computer Graphics and Image Processing SIBGRAPI 2003, São Carlos. Proceedings of the 16th Brazilian Symposium on Computer Graphics and Image Processing. v. 1, p. 1-12.
- Tiscornia, D. (1995). *Una metodologia per la rappresentazione della conoscenza* giuridica; l'ontologia formale applicata al diritto. Articolo per conferenza di filosofia del diritto. Bologna (in mimeo).

Vasconcelos, K. F. (2003). *OntoEditor*: Um editor para manipular ontologias na Web. Dissertação de mestrado inédita, Universidade Federal de Campina Grande. Disponível em:

Zavaglia, C. (2002). *Análise da homonímia no português*: tratamento semântico com vistas a procedimentos computacionais. Tese de doutorado inédita, UNESP-Araraquara.