



**CLAUDIA DA SILVA
AMARAL SANTOS**

**TERMINOLOGIA E ONTOLOGIAS: METODOLOGIAS
PARA REPRESENTAÇÃO DO CONHECIMENTO**



Universidade de Aveiro Departamento de Línguas e Culturas

2010

**CLAUDIA DA SILVA
AMARAL SANTOS**

**TERMINOLOGIA E ONTOLOGIAS: METODOLOGIAS
PARA REPRESENTAÇÃO DO CONHECIMENTO**



Universidade de Aveiro Departamento de Línguas e Culturas
2010

**CLAUDIA DA SILVA
AMARAL SANTOS**

**TERMINOLOGIA E ONTOLOGIAS: METODOLOGIAS
PARA REPRESENTAÇÃO DO CONHECIMENTO**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Linguística, realizada sob a orientação científica da Doutora Maria Teresa Costa Gomes Roberto, Professora Auxiliar do Departamento de Línguas e Culturas da Universidade de Aveiro e sob a co-orientação da Doutora Maria Rute Vilhena Costa, Professora Auxiliar do Departamento de Linguística da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa.

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

*Ó Mãe, é o Amor que existe,
Não somos nós.
Nós somos a invenção que persiste
Em ver o Amor crescer.*

*Em nós reside a inexorável vitória
Qualquer que seja a caminhada.
Brilhante será o resultado e, em glória,
Saibamos perpetuar a estrela da madrugada.*

*Percebes porque é que sempre existimos?
Nós não somos o que está aqui fora.
Viemos de dentro, pelo Amor que sentimos,
Seremos sempre a fonte da inesgotável aurora.*

Extracto de um poema escrito para mim pelo meu filho
e pelo meu marido, a quem dedico este trabalho, com todo
o meu Amor.

o júri

presidente

Reitor da Universidade de Aveiro, que delega em
Prof. Doutor Paulo Jorge Melo Matias Faria da Vila Real
professor catedrático da Universidade de Aveiro

Prof. Doutora Maria Teresa Lino
professora catedrática da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa

Prof. Doutor António Manuel Lucas Soares
professor associado da Faculdade de Engenharia da Universidade do Porto

Prof. Doutor Manuel Célio Conceição
professor associado da Faculdade de Ciências Sociais e Humanas da Universidade do Algarve

Prof. Doutor Luís Manuel Guerreiro Alves Arroja
professor associado da Universidade de Aveiro

Prof. Doutora Rute Vilhena Costa
professor auxiliar da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa
(Co-Orientadora)

Prof. Doutora Maria Teresa Murcho Alegre
professora auxiliar da Universidade de Aveiro

Prof. Doutora Maria Teresa Costa Gomes Roberto
professora auxiliar da Universidade de Aveiro (Orientadora)

agradecimentos

Ao terminar a minha dissertação, gostaria de agradecer a todos os que contribuíram para que este momento fosse possível:

À Professora Doutora Maria Teresa Roberto, orientadora, por ter acreditado em mim e por me ter apoiado nos momentos mais difíceis, sempre com um sorriso e uma candura incomparáveis.

À Professora Doutora Maria Rute Costa, co-orientadora, pelo seu empenho na qualidade do trabalho e pela segurança que sempre demonstrou e exigiu, proporcionando-me a oportunidade de crescer científica e pessoalmente.

Ao Professor Doutor Luís Arroja do Departamento de Ambiente e Ordenamento da Universidade de Aveiro, pela sua disponibilidade e interesse pelo meu trabalho, tornando-se, juntamente com a Professora Doutora Isabel Capela, elementos fundamentais para a concretização da parte prática da dissertação.

Ao Mestre Artur Xavier Pereira da Silva, pela amizade e incansável colaboração científica na parte prática e informática.

A todos os investigadores das várias universidades que contribuíram para a construção do mapa conceptual.

Ao Professor Doutor Christophe Roche, pela gentileza e debate de ideias.

Aos colegas Ana Rita Remígio, Manuel Moreira da Silva e Alexandra Albuquerque, pelos bons momentos de investigação!

Ao Conselho Directivo do ISCA e colegas, pela compreensão e ajuda institucional.

À minha família, pilar do meu bem-estar emocional.

Ao meu filho, o meu maior tesouro.

Ao meu marido, amor da minha vida.

palavras-chave

conceito, onomasiologia, ontologias, representação do conhecimento, semasiologia, terminologia, termo, texto de especialidade

resumo

Discute-se na presente dissertação as metodologias de representação do conhecimento que podem ser utilizadas em terminologia na construção de ontologias. Através da análise de duas abordagens terminológicas – semasiologia e onomasiologia – observa-se o estatuto do texto de especialidade sob um ponto de vista teórico e prático, questionando-se a sua importância e o contributo do terminólogo e do especialista na captura de conhecimento enquanto especificação informal de uma conceptualização.

keywords

concept, onomasiological approach, ontologies, knowledge representation, semasiological approach, terminology, term, text for special purposes

abstract

The subject matter of this dissertation is the discussion of knowledge representation methodologies that can be used in terminology for the construction of ontologies. The theoretical and practical analysis of two terminological approaches – semasiological and onomasiological – will allow us to observe the role played by the text for special purposes while questioning its importance and the contribution of the terminologist and the domain specialist in the capture of knowledge as representation of an informal specification of a conceptualization.

Índice

1. Introdução.....	7
2. Classificação, organização e representação do conhecimento.....	13
2.1. Classificação, taxonomia e tesauro	13
2.2. Ontologias	22
2.2.1 Ponto de vista filosófico	22
2.2.2. Ponto de vista computacional.....	27
3. Considerações teóricas sobre o conhecimento	41
3.1 Lógica.....	41
3.2 Sobre o conhecimento	44
3.2.1 Nível do conhecimento	45
3.2.1.1 Compromisso ontológico	56
3.3 Conhecimento tácito <i>versus</i> conhecimento codificado	62
4. Considerações teóricas sobre a terminologia	71
4.1 Terminologia	72
4.1.1 Perspectiva semasiológica.....	72
4.1.2 Perspectiva onomasiológica.....	75
4.2 Conceito	84
4.2.1 Ponto de vista da terminologia	84

4.2.2 Ponto de vista da inteligência artificial	89
4.3 Relações semânticas e relações conceptuais	95
5. Metodologias de representação do conhecimento	101
5.1 Perspectiva linguística	102
5.1.1 Complexidades da língua natural	102
5.1.2 Análise de dados	104
5.1.3 Relação com o texto	113
5.2 Perspectiva extralinguística	119
5.2.1 Complexidades da linguagem formal	119
5.2.2 Análise de dados	122
5.3 Notas conclusivas	141
6. Proposta de construção de um sistema conceptual	145
6.1 Perspectiva extralinguística	147
6.1.1. Mapa conceptual	147
6.1.2 Metodologia	150
6.1.2.1 Construção do mapa conceptual	152
7. Proposta de construção de uma rede lexical	165
7.1 Perspectiva linguística	165
7.1.1 Análise do texto	165
7.1.2 Relações semânticas, marcadores linguísticos e reformulações	168

7.1.3 Metodologia	172
7.1.3.1 Construção da rede lexical	173
7.1.3.1.1 Sequências de texto	178
7.1.3.1.2 Relações semânticas	194
8. Análise comparativa entre sistema conceptual e rede lexical	199
8.1. Notas conclusivas	203
10. Conclusão	209
11. Referências bibliográficas	217
11.1 Bibliografia citada.....	217
11.2 Bibliografia consultada.....	225
Anexo 1.....	241
1. Mapa conceptual.....	241
2. Rede lexical.....	241

Índice de Tabelas

Tabela 1 – Reactores	160
Tabela 2 – Mapa conceptual total: conceitos	161
Tabela 3 – Mapa conceptual total: relações	162
Tabela 4 – Relações genéricas.....	168
Tabela 5 – Relações partitivas.....	169
Tabela 6 – Relações complexas	169
Tabela 7 – Axiomas.....	170
Tabela 8 – Marcadores linguísticos: conectores discursivos e verbos.....	171
Tabela 9 – Candidatos a termo a partir do corpus: siglas	173
Tabela 10 – Candidatos a termo a partir do corpus.....	174
Tabela 11 – Candidatos a termo multi-palavra a partir do corpus	175
Tabela 12 – Candidatos a termo a partir do corpus: frequência	176
Tabela 13 – Relações e marcadores linguísticos da rede lexical.....	189
Tabela 14 – Relações de hiperonímia/hiponímia.....	194
Tabela 15 – Relações de co-hiponímia.....	194
Tabela 16 – Relações de holonímia/meronímia.....	196
Tabela 17 – Relações de sinonímia	196

Índice de Figuras

Figura 1 – Tipos de relações para o domínio ambiental: proposta.....	153
Figura 2 – Top Level Tree	154
Figura 3 – Entrada da ETAR	155
Figura 4 – Saída da ETAR	155
Figura 5 – Tipo de Equipamento	156
Figura 6 – Processos de Tratamento	157
Figura 7 – Microrganismos.....	157
Figura 8 – Mapa conceptual total	163
Figura 9 – Sequência 1	181
Figura 10 – Sequência 2	184
Figura 11 – Sequência 3	185
Figura 12 – Sequência 4	186
Figura 13 – Sequência 5	187
Figura 14 – Sequência 6	188
Figura 15 – Mapa total da rede lexical.....	197
Figura 16 – Intersecção entre termos e conceitos: ‘Processos’, ‘Tratamentos’, ‘Tecnologias’	199
Figura 17 – Intersecção entre termos e conceitos: ‘Processos de Filtração por Membranas’	200
Figura 18 – Intersecção entre termos e conceitos: ‘Lamas’.....	200
Figura 19 – Intersecção entre termos e conceitos: ‘Reactor’	201

Figura 20 – Intersecção de conceptualizações a partir da rede lexical..... 201

1.Introdução

A reutilização e partilha do conhecimento representam uma área de pesquisa e desenvolvimento de importância crucial na sociedade tecnológica, estabelecendo uma relação directa e biunívoca com a terminologia, presente, pela sua natureza, em todas as áreas científicas.

Partindo do ponto de vista das ciências da linguagem, o presente trabalho tem por objectivo analisar a contribuição da terminologia na construção de ontologias, envolvendo, fundamentalmente, duas orientações metodológicas: i) partir do termo para o conceito, o que implica adoptar uma abordagem semasiológica, centrada nas manifestações linguísticas do conhecimento, preferencialmente através da análise de texto; ii) partir do conceito para o termo, o que nos remete para uma abordagem onomasiológica, onde a primazia do conceito é sustentada por elementos conceptuais extralinguísticos, reportando-se essencialmente ao conhecimento transmitido pelo especialista de domínio.

As ontologias associam-se a procedimentos automatizados de inferências lógicas sobre determinado conjunto de dados formalizados, expressos numa linguagem de representação. Assim, o objectivo de uma ontologia é especificar explicitamente através de axiomas uma conceptualização sobre determinado objecto. O grande interesse manifestado pelas ontologias está indubitavelmente ligado ao aparecimento da internet e à necessidade de organizar a informação contida nos milhões de documentos que, todos os dias, se anexam ao conteúdo da rede. A sociedade da informação caracteriza-se por um crescimento explosivo de mercados com ciclos produtivos muito curtos, o que exige uma grande flexibilidade e capacidade de adaptação. Para evoluirmos para uma sociedade do conhecimento, necessitamos de colaboração interdisciplinar entre comunidades no sentido de organizar esse conhecimento em conteúdos interpretáveis por máquinas. Uma das formas de potenciar a comunicação entre os agentes (pessoas, organizações e sistemas de informação) seria através das ontologias.

Enquanto estrutura conceptual, uma ontologia desenvolve nos seus vários estádios especificações informais e formais de conceptualizações. Embora concebidas para ambientes computacionais, necessitam da validação humana do seu conteúdo. Para que seja possível partilhar e reutilizar conhecimento, torna-se necessário alcançar um elevado grau de consenso, sistematização e coerência, que dependem, igualmente, da aplicação e da utilização dos dados e de pressupostos

temporais exigentes. Neste contexto, o conhecimento adquire, quase sempre, contornos funcionais, cuja eficiência da representação é proporcional ao seu grau de utilidade. Assim, uma ontologia deve ter como preocupação central a permanência e a invariabilidade dos conceitos e das suas relações, idealmente independentes de qualquer língua natural.

A presente dissertação pretende questionar a utilidade do recurso à língua natural enquanto elemento catalisador de conceptualizações informais, reequacionando o estatuto do corpus tanto na abordagem semasiológica como na abordagem onomasiológica. Se, por um lado, o linguista tende a considerar que o maior problema das ontologias é de ordem conceptual, verifica-se frequentemente que o engenheiro de ontologias entende que a questão linguística constitui o grande obstáculo à representação formal do conhecimento. Procurar-se-á observar, portanto, se, sob o ponto de vista teórico e prático, é conveniente e útil prescindir do recurso ao texto de especialidade, ou se, pelo contrário, o ambiente discursivo onde os termos, enquanto entidades lexicais, normalmente se encontram, contribui favoravelmente para a construção de ontologias.

O facto de, recentemente, se observar por parte da comunidade da inteligência artificial uma aproximação às metodologias onomasiológicas defendidas por orientações teoricamente mais clássicas da terminologia, surge num movimento aparentemente contraditório relativamente ao acesso a grandes quantidades de texto em formato electrónico, passível, por isso, de serem trabalhadas automaticamente por ferramentas de processamento de língua natural. Dada a extrema dificuldade em automatizar tarefas de inferência, a inteligência artificial opta, muitas vezes, por se distanciar da língua natural, concentrando-se mais na natureza do raciocínio do que no conteúdo das conceptualizações, tentando alcançar, através de regras rígidas de formalização, representações estáticas e invariáveis, libertas de contextualizações discursivas. A utilização de metodologias que partem do extralinguístico asseguraria, de alguma forma, a eficácia da manipulação computacional.

Na abordagem onomasiológica o conceito enquanto representação mental precede a forma linguística, sendo que as relações entre conceitos são estabelecidas em função de parâmetros extralinguísticos. Será nosso objectivo observar se a base de trabalho extralinguística constitui um elemento que garante a partilha e reutilização do conhecimento, ou se, por outro lado, poderá apresentar fracturas relativamente à relevância, amplitude e profundidade do conhecimento representado, apesar de, teoricamente, respeitar os propósitos de consenso, coerência e ausência

de ambiguidade. Propomo-nos, assim, analisar a eficácia da simplificação do objecto de estudo relativamente à complexidade do conhecimento sobre o mundo, bem como o papel do especialista enquanto repositório de conhecimento especializado organizado.

Por outro lado, o facto da abordagem semasiológica se centrar nos termos e nas relações entre os termos em ambiente discursivo, implica encarar o texto como veículo privilegiado de manifestações linguísticas que reflectem, ou podem reflectir, uma organização conceptual. Enquanto designações para os conceitos, os termos constituiriam um meio de adquirir, armazenar e representar conhecimento. No âmbito das ontologias, esta perspectiva baseia-se essencialmente na análise de corpora electrónicos, partindo da língua natural e do texto enquanto espaço de actualização permanente de conhecimento. Procuraremos, pois, observar se os elementos semânticos extraídos de texto podem ser objecto de uma transposição automática para sistemas conceptuais. Apesar de, do nosso ponto de vista, as manifestações do conhecimento se encontrarem preferencialmente em discurso verbal, é um facto que este se encontra permeado de conhecimento implícito dificilmente detectável por métodos computacionais. Simultaneamente, a construção sintáctica de um texto pode conter imprecisões e recursos estilísticos que dificultam a apreensão do significado dos termos. O texto é um espaço de actualização constante, onde se sente a forte presença do ‘não dito’, do implícito e do tácito, sendo que, frequentemente, os termos figuram no discurso, mas permanecem indefinidos, uma vez que o seu significado é implicitamente explícito para a comunidade de partilha. A nossa hipótese teórica é a de que o texto constitui ainda assim um elemento operacional com valências elásticas e potencialmente eficazes na captura de conceptualizações, procurando-se verificar se um ponto de partida semasiológico é válido na especificação de conceptualizações com objectivos de manipulação computacional de dados.

A análise desenvolvida ao longo deste trabalho tentará por isso debater, de forma sustentada, determinados pressupostos presentes tanto na perspectiva semasiológica como na perspectiva onomasiológica, defendendo, em primeiro lugar, que o conhecimento é um elemento cognitivo, sendo que não é possível extrair conhecimento de um texto, seja ele de especialidade ou não; em segundo lugar, e tendo por base a afirmação anterior, que não podemos extrair ontologias automaticamente de texto, uma vez que uma ontologia é um artefacto que obedece a um processo de construção envolvendo o consenso de uma comunidade de especialistas; em terceiro

lugar, que não existe conhecimento no texto, mas sim manifestações linguísticas de conceptualizações.

Centrando-nos simultaneamente nas metodologias que partem do termo e nas metodologias que partem do conceito, tentaremos observar, com o distanciamento empírico necessário, o contributo essencial da terminologia na construção de ontologias, não só porque termo e conceito constituem duas entidades indissociáveis, como também pelo facto de ser através deles que o conhecimento encontra um veículo de representação. Procuraremos, assim, argumentar a presença fundamental da terminologia na fase anterior à formalização do conhecimento através de uma linguagem de representação, ou seja, na fase da especificação informal da conceptualização, um pré-estágio da conceptualização formal. É nesse momento que a terminologia oferece o seu contributo enquanto ciência e conjunto de práticas, quer parta de dados linguísticos, quer de dados extralinguísticos.

Através dessa análise será possível também questionar o estatuto do texto de especialidade enquanto repositório supostamente natural de conhecimento organizado. Simultaneamente, poderemos, da mesma forma, observar as potencialidades do texto de especialidade na constituição de redes lexicais compostas por dados semânticos que se podem aproximar ou distanciar dos respectivos sistemas conceptuais do domínio. Pretende-se confirmar que a construção de uma ontologia não pode subtrair-se à questão linguística, sendo que a ontologia gravita necessariamente em torno da terminologia, capturando, através do conceptual e do lexical, as conceptualizações necessárias à partilha e reutilização de conhecimento formalizado. Apesar do conhecimento do especialista ser tendencialmente de tipo esquemático, pretendemos demonstrar que o texto de especialidade pode conter conexões semânticas que vão enriquecer o sistema conceptual. No entanto, partimos do pressuposto que o lexical não deve substituir o conceptual, uma vez que é através deste que conseguimos apreender a organização das conceptualizações

Constitui-se também como objectivo desta dissertação demonstrar que a terminologia no seu estágio actual não deve limitar-se ao texto, encarando com mais abertura o estatuto do termo enquanto unidade não necessariamente lexical. Procura-se evidenciar a necessidade de um reposicionamento dos estudos terminológicos e da função do terminólogo no contexto da construção de ontologias, bem como do seu objecto de estudo. Do mesmo modo, procura-se

demonstrar que o engenheiro de ontologias deve contar com a contribuição do terminólogo, num movimento recíproco que deverá envolver sempre, e em todos os momentos, o especialista de domínio.

A nossa postura teórica e prática apontará, assim, para a defesa de utilização tanto da metodologia onomasiológica como da metodologia semasiológica, respeitando o propósito da aplicação e o âmbito da representação, tendo sempre presente o tipo de manifestações de conhecimento que um texto pode oferecer e o tipo de conhecimento que uma ferramenta informática pode representar.

Pretende-se por isso verificar se o sucesso da captura das conceptualizações informais se encontra directamente relacionado ou não com o recurso à língua natural. A nossa hipótese teórica é a de que o sucesso ou o insucesso encontra-se antes directamente relacionado com o consenso alcançado pela comunidade de especialistas e com os limites computacionais da linguagem de representação. Consideramos por isso como ponto de partida teórico que não é necessário existir uma descontextualização linguística para a correcta apreensão da conceptualização.

Tentaremos provar que na construção de ontologias não devem existir abordagens exclusivamente extralinguísticas ou linguísticas, mas sim metodologias complementares que podem partir do linguístico ou do extralinguístico, uma vez que uma ontologia é um artefacto computacional que se constrói em colaboração interdisciplinar.

Iniciamos a dissertação com um capítulo sobre classificação, organização e representação do conhecimento, pretendendo-se esclarecer conceitos como classificação, taxonomia, tesauro e ontologia. Através da análise dos diferentes métodos de classificação, estabelecemos uma ligação entre a lógica de Aristóteles e a lógica formal contemporânea, ambas constituintes da construção de ontologias. Para que fosse possível perceber melhor a apropriação do termo 'ontologia' por parte da comunidade de inteligência artificial e a sua definição actual, foi necessário abordarmos a distinção entre duas áreas de interesse: Ontologia no sentido filosófico, e ontologia no sentido computacional.

No capítulo sobre conhecimento, começamos a nossa análise com uma breve incursão pela lógica, dissertando em seguida sobre o que se pode entender por conhecimento e representação do conhecimento, tanto do ponto de vista computacional, como do ponto de vista social. Abordamos igualmente o conceito de compromisso ontológico, condição fundamental para a partilha e reutilização de ontologias. A distinção entre conhecimento tácito e conhecimento codificado constitui também motivo de reflexão.

O capítulo seguinte tem por objectivo analisar as linhas de orientação das abordagens terminológicas utilizadas na construção de ontologias, distinguindo entre a perspectiva linguística e a perspectiva extralinguística, ou seja, entre uma abordagem semasiológica e uma abordagem onomasiológica. Analisa-se o estatuto do termo e a teoria do conceito, revisitando, nomeadamente, os princípios de Eugen Wüster, e re-equacionando as diferentes perspectivas à luz das necessidades práticas actuais dos sistemas de informação e das ontologias. Faz-se igualmente a distinção entre relações semânticas e relações conceptuais, com o objectivo de distinguir o tipo de dados que se podem extrair dos vários elementos operativos.

As metodologias de representação do conhecimento são abordadas no capítulo que se segue, tendo como principal propósito uma descrição das orientações propostas pelas abordagens semasiológica e onomasiológica na construção de ontologias, observando o tipo de análise de dados que pode ser encetado tanto pela perspectiva que parte do linguístico, como pela perspectiva que parte do extralinguístico. Debate-se a relação entre texto e ontologia e a questão da automatização e extracção automática de dados. A língua natural e as linguagens formais de representação constituem igualmente uma questão a debater. Observa-se e discute-se, ainda, uma proposta de abordagem conceptual por parte da inteligência artificial.

No último capítulo pretendeu-se observar, através de uma postura crítica, os dados que podem ser extraído das metodologias que partem do conceptual, e das metodologias que partem do linguístico. O trabalho desenvolvido situa-se na fase da especificação informal das conceptualizações e é demonstrado através da construção de um mapa conceptual e de uma rede lexical no domínio dos Tratamentos Biológicos de Águas Residuais, com as respectivas considerações finais.

2. Classificação, organização e representação do conhecimento

A capacidade simbólica, especificamente humana e que constitui a forma de apropriação do real por excelência, foi desenvolvida ao ponto de criar um universo autónomo que Popper designa por 'terceiro mundo' e que é, grosso modo, o dos produtos da mente humana. (Borges, 2002)

“No princípio era o Verbo, o Verbo estava junto de Deus e o Verbo era Deus” (Jo. 1,1); “(...) tudo se fez, por meio d’Ele e, sem Ele, nada se fez.” (Jo. 1,3). O início do Evangelho de São João, em que narra a obra salvífica do Verbo encarnado, evoca o Versículo Um do Génesis, onde se afirma “(...) no princípio, Deus criou o céu e a terra (...)” (Gen. 1,1). Esta preocupação fundamental do homem com a ordem e origem do universo encontra-se intimamente ligada à necessidade de classificar e organizar o mundo, numa busca incessante de conhecimento. O conhecimento, porém, não existe em formato utilitário. É um elemento cognitivo, algo que se encontra na mente humana. Permite, contudo, ser objecto de transmissão através da representação de conceptualizações mentais. Podemos transmitir conhecimento através de uma imagem, da língua natural, de uma linguagem computacional. Para isso, necessitamos de processos que permitam classificar, organizar e representar esse conhecimento com maior ou menor grau de complexidade.

2.1. Classificação, taxonomia e tesouro

Embora seja possível enumerar características distintivas entre os vários processos de classificação, organização e representação do conhecimento, deparamo-nos com intersecções naturais que dificultam a delimitação exacta dos seus contextos de utilização. Classificações, taxonomias, tesouros e ontologias são considerados recursos, cujo elemento mais distintivo será talvez o tipo de aplicação na qual vão ser utilizados, respondendo a diferentes necessidades de conceptualização e especificação. Como afirma Madsen e Thomsen, o próprio

Comité Técnico 37 da ISO estabeleceu em 2007 um grupo de trabalho na área das ontologias cuja primeira missão é, precisamente, “(...) to describe [the] different types of knowledge representation resources, and clarify the differences between those types, e.g. ontology, taxonomy and thesaurus.” (Madsen & Thomsen, 2008, pp. 107-108)

O conceito mais geral é o de classificação, tendo começado a ser analisado de forma sistemática por filósofos como Aristóteles. Podemos dizer que, no seguimento da terminologia aristotélica, uma classificação organiza o conhecimento em categorias ou grupos, de acordo com determinados critérios. As classificações baseiam-se essencialmente na presença ou ausência de determinada propriedade, ou na diferença específica, conceito que deriva de Aristóteles e que se refere a uma propriedade incompatível com todas as outras propriedades situadas no mesmo nível de classificação. Na abordagem aristotélica, um conceito é definido de acordo com a sua essência, e um atributo é considerado um acidente. Esses conceitos são estruturados de acordo com a diferença específica onde a propriedade essencial é usada para distinguir um conceito de todos os outros conceitos. As propriedades vão permitir estabelecer as relações entre os conceitos. A escolha de uma ou de outra propriedade determinará, por vezes de forma profunda, a selecção e disposição da realidade a classificar, levando, por consequência, à constituição de diferentes classificações. Nas palavras de Mendes, “(...) identificar as relações entre conceitos é uma forma de classificar, pois as classificações são elementos geográficos mentais.” (Mendes, 2008, p. 33)

Para fundamentar a classificação por diferença específica, Aristóteles defende no tratado *Categorias* que os conceitos quando isolados não expressam verdade ou falsidade; é somente com a combinação das ideias numa proposição que a verdade ou a falsidade são possíveis. Os elementos de tal proposição são o substantivo e o verbo. A combinação de palavras dá lugar ao discurso racional e ao pensamento, transmitindo um significado tanto nas partes como no todo¹. Para raciocinar sobre e com as categorias, Aristóteles desenvolveu um processo de lógica formal: os silogismos. A imagem da árvore, datada de III A.D. pela mão do filósofo Porfírio, e mais tarde pelo filósofo vienense Franz Brentano (1862), atesta a importância do

¹ Para Aristóteles, a contribuição da Lógica enquanto ciência e enquanto metodologia é fundamental, entendida aqui como um conjunto de processos de inferência. A formalização computacional baseia-se, em parte, nos princípios aristotélicos.

conceito de categoria introduzido por Aristóteles. O método aristotélico para definir novas categorias consistia no *genus* e na *differentia*. A *differentia*, como vimos, definia-se pelas propriedades que distinguem diferentes espécies do mesmo género. Um exemplo clássico seria: *Todos os homens são mortais: Sócrates é um homem: Sócrates é mortal*. Os silogismos eram regras para raciocinar sobre tipos e subtipos, constituindo-se como um padrão formal para representar regras de inferência.²

Uma classificação é, para Bourigault et al, “(...) la répartition systématique en classes, en catégories, d’êtres, de choses ou de notions ayant des caractères communs notamment afin d’en faciliter l’étude ; c’est aussi le résultat de cette opération.” (Bourigault, Aussenac-Gilles, &

² Com a utilização de letras (A e B), Aristóteles introduziu pela primeira vez na história a utilização de variáveis. A forma silogística da argumentação lógica foi dominante durante 2000 anos e todos os métodos baseados na Lógica, desde a árvore de Porfírio até às recentes ontologias formais, são exemplos de uma abordagem aristotélica descendente. Esta prática revelou-se fundamental para a evolução das tecnologias baseadas na engenharia do conhecimento. Embora os padrões se assemelhem a uma língua natural, estão limitados a uma sintaxe altamente estilizada e constricta, designada também por língua natural controlada. As proposições utilizadas nos silogismos por Aristóteles permitem que um raciocínio possa ser transposto para a máquina, tendo constituído, juntamente com as contribuições dos Escolásticos medievais, a base da Lógica de primeira ordem e, por exemplo, das linguagens DAML e OIL. (Beck & Pinto, 2002) Exemplo dos silogismos aristotélicos:

Tipo	Nome	Padrão
A	<i>Afirmativa Universal</i>	Todo o A é B.
I	<i>Afirmativa Particular</i>	Alguns As são B.
E	<i>Negativa Universal</i>	Nenhum A é B.
O	<i>Negativa Particular</i>	Alguns As não são B.

Charlet, 2002, p. 5) No campo das ciências da informação e documentação, e de acordo com Mendes, uma classificação é “(...) uma forma de representar o assunto dos documentos.” (Mendes, 2008, p. 13) Alguns autores colocam a questão de saber até que ponto uma classificação é natural ou arbitrária. Pombo considera que qualquer classificação implica uma arbitrariedade irreductível: “Tendo por base o estabelecimento de classes e suas respectivas fronteiras, a classificação mais não faz do que quebrar, violenta e arbitrariamente, a cadeia de imperceptíveis nuances que liga os seres entre si.” (Pombo, 1998, p. 8) A percepção das afinidades que permitam classificações ditas ‘naturais’ gerou, a seu tempo, discussões sobre os critérios a ter em conta, dando origem a metodologias diferentes³.

Uma taxonomia é um processo de classificação. Tradicionalmente ligada à biologia e à lógica, a palavra taxonomia tem a sua raiz etimológica no grego (*táxis* – ordenação; *nomos* – lei, norma, regra) e era considerada como a ciência para a classificação de organismos vivos. Das taxonomias científicas, a mais conhecida é a Taxonomia de Lineu. A origem do termo prende-se também com a biologia, onde é utilizado para definir a localização única de uma espécie dentro de uma hierarquia complexa. Em “Ten Taxonomy Myths” (Review, 2002) define-se taxonomia como um sistema para nomear e organizar as coisas em grupos que partilham características semelhantes. De acordo com Simões, o conceito de taxonomia aparece ligado no campo da informação aos sistemas de classificação. Os sistemas de classificação são estruturas de organização do conhecimento, esquemas ordenados em várias áreas, segundo determinadas regras ou normas: “Nos sistemas de classificação tradicionais este método caracteriza-se por ser um processo de comparação dicotómico entre os elementos que irão constituir o sistema de classificação, assente nas diferenças e semelhanças dos mesmos, concorrendo para que seja determinada *a priori* a posição que cada elemento irá ocupar na respectiva tabela.” (Simões, 2008, p. 71)

³ A perspectiva essencialista coloca o indivíduo como base do espaço classificatório; a perspectiva estrutural ou morfológica elege como critério de ‘naturalidade’ a existência de relações constantes entre as propriedades comuns aos elementos de uma mesma espécie; a metodologia evolutiva ou genética considera essas concretizações como ‘realizações que se sucedem efectivamente no tempo’. (Pombo, 1998)

Rees, no artigo “Clarity in the usage of the terms ontology, taxonomy and classification”, distingue taxonomia de classificação, sendo que uma taxonomia classifica os elementos de um domínio organizando-os de acordo com uma hierarquia baseada nos dados internos aos elementos contidos nessa hierarquia, ou seja, segundo as relações entre as entidades, enquanto uma classificação simples pode usar, por seu turno, dados mais arbitrários ou externos⁴. (Rees, 2003)

Dada a natureza multi-disciplinar das taxonomias, quase tudo pode ser classificado de acordo com algum esquema taxonómico. Pensa-se que a mente humana organiza naturalmente o conhecimento do mundo nesses sistemas. De facto, o termo taxonomia tem sido objecto de utilização quase universal quando nos referimos a um tipo de estrutura abstracta. Originalmente limitada a uma nomenclatura organizada dos reinos animal e vegetal, as taxonomias passaram a comportar múltiplos esquemas e aplicações para classificar objectos de conhecimento para fins científicos, técnicos e, actualmente, comerciais. Segundo Garshol, uma taxonomia pode ser entendida na área dos sistemas de informação como “(...) a subject-based classification that arranges the terms in the controlled vocabulary into a hierarchy without doing anything further, though in real life you will find the term ‘taxonomy’ applied to more complex structures as well.” (Garshol, 2004)

A primeira imagem à qual normalmente associamos o conceito de ‘classificação taxonómica’, se assim o podemos designar, é a de uma árvore. As árvores naturais são compostas por folhas, ligadas entre si por diversos ramos. As folhas são os nós da ‘árvore taxonómica’, representando as categorias com níveis maiores e menores de generalidade, e os ramos que ligam os nós representariam as relações de inclusão de categoria mais baixa numa mais alta. Insistir na estrutura em árvore é insistir, na realidade, de que a partir de qualquer nó na árvore existe pelo menos um ramo que sobe. O conhecimento deveria, assim, consistir num conjunto potencialmente infinito de árvores, com numerosas famílias, cada uma reflectindo uma visão

⁴ Como exemplo de dados internos, Rees refere o seguinte exemplo: o espinafre é um legume e nem todos os legumes são espinafres, por isso espinafre é uma subclasse de legume. A decisão de posicionar o espinafre na categoria de legume é baseada em dados inerentes às entidades, por isso seria uma parte da taxonomia (uma taxonomia com uma hierarquia de subclasses).(Rees, 2003)

específica do domínio em causa, utilizando perspectivas com níveis diferentes de granularidade.

As taxonomias têm sido muito utilizadas na área da inteligência artificial. Tal como refere Garshol,

“(…) taxonomies have recently emerged from the quiet backwaters of biology, book indexing, and library science into the corporate limelight. They are supposed to be the silver bullets that will help users find the needle in the intranet haystack, reduce ‘friction’ in electronic commerce, facilitate scientific research, and promote global collaboration. But before this can happen, practitioners need to dispel the myths and confusion, created in part by the multi-disciplinary nature of the task and the hype surrounding content management technologies.” (Garshol, 2004)

O termo tesouro, por sua vez, deriva do latim e do grego *thesaurus* que significava *tesouro*. Com o tempo passou a usar-se para designar o vocabulário de filologia, designadamente os dicionários especializados de uma determinada área de conhecimento. O primeiro emprego desta palavra data de 1531. Um tesouro constitui um sistema de indexação⁵ para representação conceptual, um vocabulário controlado, instrumento de organização, representação e recuperação do conhecimento.

O tesouro foi objecto de um crescimento exponencial resultante, primordialmente, da importância que a informação alcançou durante a Segunda Guerra Mundial. De acordo com Simões (2008), a informação passou a ser um bem de consumo. A democratização e socialização do conhecimento reclamavam pela organização e controle dos conteúdos informativos dos documentos. Foi necessário então desenvolver uma linguagem controlada para reduzir a ambiguidade semântica e sintáctica, obtendo um elevado nível de consistência

⁵Segundo Mendes, “(…) a classificação da informação é a estrutura de organização dos conceitos através de uma linguagem formal para descrever os conteúdos de documentos. Designamos por indexação o uso desta linguagem formal, isto é, a linguagem documental, para representar o conteúdo de um documento num sistema de gestão de informação com o objectivo de permitir a sua recuperação (… os instrumentos de indexação têm por base o trabalho de selecção e recolha de vocabulário que é caracterizado como um trabalho de elaboração de terminologia (… que por sua vez pode ser apresentado sob a forma de banco de dados, glossários, tesouros ou outras publicações.” (Mendes, 2008, pp. 15-16)

na representação dos assuntos e realizando pesquisas não só com termos específicos, mas também com termos mais gerais. O tesouro baseava-se numa estrutura de relações semânticas, nomeadamente hierárquicas, permitindo a pesquisa genérica e específica:

“Desejava-se uma linguagem que garantisse a consistência no input (armazenamento) e a pertinência no output (difusão e recuperação da informação). A resposta foi encontrada na linguagem natural, com uma linguagem controlada e flexível conferida pela sua estrutura combinatória que proporcionava ao utilizador a possibilidade de coordenar os conceitos. A este novo tipo de linguagem deu-se o nome de tesouro.”
(Simões, 2008, p. 47)

Existem várias definições de tesouro. A norma ISO 2788:1986 define tesouro como um vocabulário de uma linguagem de indexação controlada organizado formalmente de maneira a explicitar as relações estabelecidas *a priori* entre os conceitos. Simões cita a norma NF Z47-100-1981 onde se explicita que um tesouro é constituído por um conjunto de termos, descritores e não descritores, e de relações estabelecidas entre eles que especificam o seu campo semântico. O vocabulário que constitui um tesouro deve ser isento de ambiguidade e estruturado de forma a assegurar uma melhor definição de cada termo, bem como permitir pesquisas em vários graus de generalidade ou especificidade. Ainda em Simões, e comentando a norma BS 8723-2:2005, o objectivo de um tesouro seria orientar, tanto o indexador como o utilizador, de forma a seleccionar o termo preferido. (Simões, 2008, pp. 48-49)

O estabelecimento de relações semânticas entre os descritores confere-lhe uma estrutura semelhante a uma rede conceptual. Para Costa, um tesouro “(...) é um sistema de classificação que serve fins documentais (...)” (Costa, 1993, p. 33), fazendo uso de três tipos de relações semânticas: relações de equivalência, relações hierárquicas e relações associativas. Na perspectiva de Simões, as relações hierárquicas são o eixo estruturante do tesouro, distinguindo-o de uma lista de termos não estruturados, proporcionando forma à estrutura hierárquica que ocorre dentro de um campo conceptual. A reciprocidade entre os termos é traduzida pelos operadores TG (termo genérico) e TE (termo específico). A partir da década de sessenta, a utilização da máquina no tratamento da informação dita a preferência pelo tesouro em detrimento das classificações.

Na opinião de Garshol, as taxonomias aparecem como sendo um produto dos tesouros, por força das relações entre descritores⁶: “In a taxonomy the means for subject description consist of essentially one relationship: the broader/narrower relationship used to build the hierarchy. The set of terms being described is of course open, but the language used to describe them is closed, since it consists only of a single relationship. Thesauri extend this with the RT and UF/USE relationships, and the SN property, which allow them to better describe the terms.” (Garshol, 2004)

Embora o presente capítulo pretenda abordar a questão das ontologias de forma aprofundada, torna-se pertinente referir neste ponto que é comum encontrar na literatura da especialidade análises comparativas entre tesouro e ontologia. Para Mendes, um tesouro e uma ontologia constituem processos de formalização e representação do conhecimento, sendo que “(...) a evolução do tesouro, enquanto sistema de representação do conhecimento, reside na arquitectura da ontologia, enquanto estrutura tecnológica para o sistema conceptual e na qual o vocabulário controlado é um ponto de partida. As relações ontológicas permitem especificar o motivo de ligação entre conceitos e termos e associam a linguagem natural com a linguagem formal.” (Mendes, 2008, p. 56)

De acordo com Simões, alguns autores consideram uma ontologia um tesouro avançado, constituindo-se como mais um tipo de estrutura de organização cognitiva baseada em conceitos e nas suas relações. Esta comparação entre tesouro e ontologia prende-se, por um lado, com o facto de ambos organizarem o conhecimento para fins de recuperação, embora as ontologias se circunscrevam a ambientes informáticos, e, por outro lado, ao facto de ambos seleccionarem conceitos pertencentes a um determinado domínio estabelecendo relações entre os mesmos. A autora entende que uma ontologia é, antes de tudo, um recurso artificial,

⁶ Exemplo de relações entre descritores: BT (Broader Term) – um determinado termo é mais geral que outro; NT (Narrower Term) – um determinado termo é mais específico que outro; RT (Related Term) – dois termos encontram-se associados; UF (Use For) – um determinado termo é o termo preferencial entre um conjunto de termos sinónimos. <http://www.doc.ua.pt/> A norma ISO 2877 trata sobre tesouros monolíngues e a norma ISO 5964 centra-se nos tesouros multilíngues.

criado para organizar e recuperar conhecimento em ambientes de inteligência artificial, restrita a um domínio do saber. Todo o processo de representação pressupõe princípios e métodos de tal forma que as especificações possam ser processadas e legíveis por máquinas. E aqui encontra-se a diferença fundamental entre tesouro e ontologia: o tesouro pode aplicar-se a ambientes onde interagem os elementos homem e máquina, enquanto a ontologia, apesar de ser validada pelo homem, foi concebida para ambientes formalizados.

Para Beck e Pinto (2002) os principais pontos que distinguem um tesouro de uma ontologia prendem-se com o facto de um tesouro continuar a ser ambíguo e informal na descrição das relações entre conceitos, enquanto uma ontologia exige um grau de formalidade que não permite ambiguidades. Os dois artefactos têm como essência a organização de conhecimento numa estrutura hierárquica. Porém, um tesouro não contém definições de conceitos interpretáveis por máquinas, nem é capaz de suportar raciocínio sobre terminologia.

Comentando as diferenças entre tesouro e ontologia, Mendes afirma que

“(...) a estrutura e conteúdo de um tesouro, isto é, de um vocabulário controlado e organizado, serve a prática da indexação. Uma ontologia poderá servir para indexação se esse for o propósito da sua construção, pois uma ontologia é definida para um domínio em particular e para um determinado objectivo, no mesmo sentido, os critérios de construção da ontologia serão também os critérios de avaliação da sua qualidade. (...) estamos perante uma forma de evolução do tesouro e consideramos que a síntese de uma ontologia poderá ser uma classificação especial de um domínio que pode ser apresentada de forma dinâmica em acesso em linha. (...) Tesouro e ontologia são vocabulários controlados. Constroem-se para expor um vocabulário controlado, enquanto a terminologia organiza os termos da língua em especialidade. Assim, a ontologia e o tesouro têm no trabalho terminológico a sua base de trabalho.” (Mendes, 2008, pp. 36-39)

Segundo Simões, ontologias, taxonomias e mapas conceptuais⁷ são instrumentos que permitem uma comunicação mais rápida entre homem e máquinas, nomeadamente no acesso e recuperação de informação na internet: “Estes instrumentos baseiam-se nos princípios base dos tesouros: indexar o conteúdo dos documentos segundo estruturas que contemplem relações hierárquicas e relações associativas.” (Simões, 2008, p. 69) Apesar de existirem diferenças, apresentam-se como estruturas conceptuais utilizadas para representar informação com fins de recuperação em ambiente web e surgem como uma evolução lógica dos tesouros.

2.2. Ontologias

2.2.1 Ponto de vista filosófico

A utilização primeira do termo ontologia deriva da Metafísica e de um dos tratados de Lógica de Aristóteles intitulado *Categorias*, obedecendo a um plano essencialmente didático de iniciação filosófica⁸. A importância deste tratado, explicado mais tarde por Porfírio em *Eisagoge* (Séc. III A.D.), só é comparável, na religião, à importância da Bíblia. Trata-se de um tratado não dogmático, mas irrefutável, sobre o estudo da existência.

Ontologia, enquanto disciplina, é o estudo do ser pelo ser (Aristóteles, Meta. 1031a) e tem sido uma parte importante da metafísica desde os tempos mais antigos. A própria metafísica

⁷ Os mapas conceptuais são uma técnica de representação visual do conhecimento na qual a informação, os conceitos e as suas relações aparecem representados sob a forma de diagramas ou mapas. (Simões, 2008) Este assunto será desenvolvido no último capítulo.

⁸ Segundo Fieser e Dowden em *The Internet Encyclopedia of Philosophy*, os tratados sistemáticos de Aristóteles podem ser agrupados da seguinte forma: i) lógica - *Categorias* (classificações dos termos), *Sobre Interpretação* (proposições, verdade, modalidade), *Analítica Anterior* (lógica silogística) e *Analítica Posterior* (método científico e silogismo) - ii) física, iii) psicologia, iv) história natural, v) filosofia – com, nomeadamente, tratados sobre *Metafísica* (substância, causa, forma, potencialidade). Os escritos de Aristóteles sobre o tópico geral da lógica foram agrupados mais tarde pelos Peripatéticos sob o nome de *Organon*, ou *Instrumento*. Na perspectiva de Aristóteles, a lógica e o raciocínio eram os instrumentos básicos e preparatórios para a investigação científica. (Fieser & Dowden, 2008)

foi tradicionalmente dividida em ontologia e cosmologia. A ontologia centrava-se no estudo do ser em si mesmo, enquanto a cosmologia estudava o universo físico em geral. Implícita nesta divisão está a distinção entre metodologias. Em particular: a metodologia da cosmologia é baseada na análise de categorias como espaço, tempo, assunto e causalidade, onde o objectivo é descobrir pela observação e experiência as leis que ligam estas categorias e os seus constituintes uns com os outros, incluindo os tipos naturais de coisas (seres) na natureza. Por sua vez, a metodologia da ontologia é baseada na análise das categorias ontológicas, ou seja, as categorias do ser, sendo o objectivo descobrir as leis que ligam essas categorias e as entidades nessas categorias. O sucesso de uma categorização mede-se pelo grau de previsão e controle produzido pelas categorias. Teoricamente, boas categorizações da natureza produzem boas teorias.

Alguns autores como Fieser e Dowden (2008) afirmam, contudo, que ontologia não consiste no estudo do 'ser pelo ser'. Isso faz parte da metafísica, que significa literalmente 'o que está para além da Física', e que os estudiosos de Aristóteles utilizavam para se referirem ao que Aristóteles designava por 'primeira filosofia'. Para Aristóteles, a filosofia nasceu historicamente depois de estarem asseguradas as necessidades básicas. Cresceu a partir de um sentimento de curiosidade e deslumbramento, ao qual os mitos religiosos deram somente uma satisfação provisória. O objecto da metafísica lida com os primeiros princípios do conhecimento científico e as condições de toda a existência. A ontologia dedica-se, sim, ao estudo do ser. Esta distinção é subtil mas importante. Uma ontologia pode incluir objectos como anjos ou maçãs sem se interrogar metafisicamente sobre a existência dos anjos. Do mesmo modo, o termo ontologia é por vezes utilizado num sentido mais lato, designando o estudo do que poderia existir. Surgem também discussões sobre a confusão entre ontologia e epistemologia. Enquanto a epistemologia estuda o conhecimento e o saber, a ontologia preocupar-se-ia com o estudo da existência.

Uma outra perspectiva, como refere Gómez-Pérez et al (2004), é encarar a Ontologia, enquanto 'filosofia do Ser', preocupando-se com a essência das coisas e com a sua permanência. Os conceitos existem fora da nossa mente? Como podemos classificar as entidades que existem no mundo? O centauro, por exemplo, é metade homem, metade cavalo. Tem essência, mas não existe. Encontrar a essência através da mudança, comenta o

autor, constituía um dos grandes desafios para os antigos Gregos. Parménides assumiu a independência da essência das coisas relativamente aos nossos sentidos: Não há mudança. Uma coisa que é não pode ter começo nem fim. Este posicionamento é importante. Os nossos sentidos é que se apercebem primeiro de determinados traços, dependendo da forma como observamos. Em termos de sistemas de informação, é como se tivéssemos uma base de dados, com acessos diferentes.

Na Idade Média, a questão dos Universais foi um ponto de exploração teórica, com duas correntes: os realistas afirmavam que os universais existem, os nominalistas refutavam esta posição, advogando que os universais são só palavras para nos referirmos às coisas. Alguns teóricos afirmavam que os universais são apenas símbolos. É precisamente esta abordagem simbólica que constituiu o fundamento da física moderna, bem como o da informática. A categorização proposta por Aristóteles foi questionada por Kant, segundo o qual a essência das coisas não é só determinada pelas coisas em si mesmo, mas também pela percepção de quem as observa. Para capturar a realidade, ordenamos as nossas sensações, primeiro em espaço e tempo, e depois de acordo com as categorias. Esta nova visão das coisas pode ser aplicada às ferramentas computacionais: os objectos formalizados dependem também, e em grande escala, do desenho do próprio sistema. Aquele objecto específico tem aquelas características específicas de acordo com as decisões do criador do sistema. Ou seja, a categorização do mundo depende fortemente da pessoa que o percebe, numa tentativa de extrair a essência ao mesmo tempo que sistematizamos as características do que nos rodeia.

Na verdade, consideramos normal classificar as coisas, os entes, os factos e os acontecimentos, constituindo os suportes estáveis do nosso quadro mental. Na ausência do reconhecimento de espaços, hábitos, lugares, tempos e seres, ficaríamos irremediavelmente perdidos num labirinto. É necessário agrupar, catalogar, aproximar ou distanciar. Foucault (1966) considera as classificações anteriores às palavras e aos gestos, numa necessidade primordial de ordenar a variedade. Com o evoluir dos tempos, o ímpeto que sempre imbuíu os cientistas na luta pela categorização da natureza, implícito na visão do mundo de Aristóteles, foi acompanhando as novas metodologias, que suplantaram de forma espectacular os pressupostos aristotélicos com o advento da ciência, permitindo a categorização eficiente de

partes do mundo natural, sob as quais se ergueram teorias bem fundamentadas, e cujo sucesso pode ser mensurado em proporção ao controle sobre essas categorias.

O princípio da organização em categorias foi apresentado no séc. XVII por John Wilkins, Bispo de Chester, em *An Essay towards a Real Character and Philosophical Language* (1668). Todas as entidades da realidade deveriam ser abrangidas nestas linhas dentro de uma única árvore, extensível através de distinções cada vez mais finas. Constituindo-se como base de uma língua ideal, análoga à *characteristica universalis* concebida por Leibniz, seria possível expressar todos os conceitos por meio de composição sistemática a partir de uma lista de conceitos básicos ou simples. O objectivo de relacionar mecanicamente conceitos a primitivos foi proposto primeiro por Ramon Lull no séc. XIII. O seu sistema inspirou Leibniz que desenvolveu a sua *Universal Characteristic* em 1679⁹.

A tabela periódica dos elementos e a classificação das coisas vivas de acordo com os reinos, filos, classes, etc. é um exemplo de ontologia desenvolvida por cientistas. Existem também exemplos cujo sucesso está constantemente a ser desafiado e cujas categorias são frequentemente ajustadas. Contudo, estas ontologias oferecem aos cientistas uma grande quantidade de previsão e controle sobre os seus universos-objecto. Como afirma Koepsell, nenhuma ontologia estará jamais completa, mas podemos determinar o sucesso ou valor de uma ontologia relativamente a outra (Koepsell, 1999).

De origem grega – *ontos* para ser, e *logos* para palavra escrita ou falada e, mais tarde, para razão – o termo ontologia foi reintroduzido no Séc. XIX por filósofos alemães para distinguirem o estudo do ser do estudo dos diversos tipos de seres nas ciências naturais. Tudo o que pudesse ser dito ou *predicado* sobre alguma coisa, era considerado uma ‘categoria’, de acordo

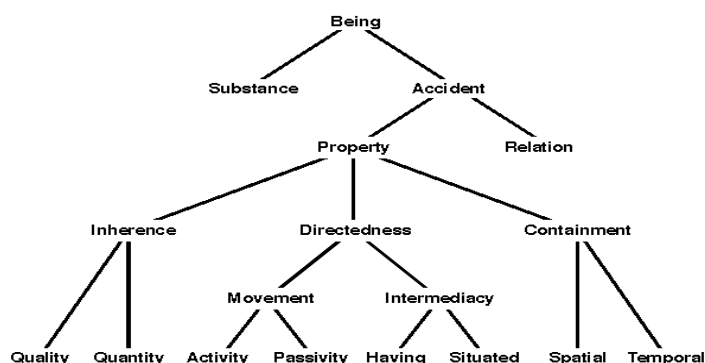
⁹ Leibniz queria construir um dicionário universal para mapear conceitos associando-os a números e um sistema de cálculo para raciocínio que automatizasse os silogismos. Para simplificar os cálculos, inventou a primeira máquina de calcular com divisão e multiplicação.

com a doutrina de Aristóteles.¹⁰ Como refere Maia, e citando Kant, “(...) ontology is the first part that actually belongs to metaphysics. The word itself comes from the Greek, and just means the science of beings, or properly according to the sense of the words, the general doctrine of being.” (Maia, 2003, p. 2) O termo é por isso frequentemente registrado em textos que discutem a existência de Deus e outros debates metafísicos.

O termo ontologia foi novamente cunhado em 1613 de forma independente por dois filósofos, Rudolf Göckel em *Lexicon philosophicum* e Jacob Lorhard (Lorhardus) em *Theatrum philosophicum*. A primeira ocorrência em inglês data de 1721, cunhagem registrada por Merriam Webster através de duas definições, e parafraçando: (1) um ramo da metafísica que se preocupa com a natureza e relações do ser; (2) uma teoria sobre a natureza do ser ou dos tipos de coisas que existem.

Uma definição mais recente encontra-se no Webster’s Third New International Dictionary (2002): “1. A science or study of being: specifically, a branch of metaphysics relating to the nature and relations of being; a particular system according to which problems of the nature of being are investigated; first philosophy. 2. A theory concerning the kinds of entities and specifically the kinds of abstract entities that are to be admitted to a language system.” A primeira entrada aponta para a definição da Filosofia, sendo que a segunda parte tem origem nos princípios de Husserl e analisa as categorias fundamentais do objecto, estado de coisas, parte, todo, as relações entre as partes e o todo e as suas leis de dependência. O segundo sentido é utilizado na Inteligência artificial e representação do conhecimento.

¹⁰ As 10 categorias básicas constam do diagrama do filósofo Franz Brentano (1862). (Sowa, 2006)



Podemos concluir, e citando Guarino, que Ontologia é “(...) that branch of philosophy which deals with the nature and the organization of reality.” (Guarino & Giaretta, 1995, p. 2) Enquanto ramo da filosofia, constitui-se como uma ciência que estuda os tipos e as estruturas dos objectos, eventos, processos e relações em todas as áreas da realidade. Esta reflexão permite-nos analisar, de seguida, uma outra concepção de ontologia, aquela que predomina actualmente nas áreas da inteligência artificial e representação do conhecimento.

2.2.2. Ponto de vista computacional

Ontologia é hoje um conceito multifacetado, diferente do que foi no passado. Por ser, na sua essência, uma área de estudo multidisciplinar imbuída de uma herança milenar oriunda da Antiguidade Clássica, deparamo-nos com uma dificuldade acrescida em definir-lhe os contornos. O grafismo do termo não é inocente. Na realidade, a concepção actual deste conceito levou a duas formas de referência: ‘Ontologia’, enquanto ciência, e ‘ontologia’, enquanto artefacto. Roche afirma que “(...) the use of the term ontology is somewhat unfortunate since it has a definite meaning in the philosophical literature which has little to do with describing the content of information repositories.” (Roche, 2003a, p. 2)

Até recentemente, as ontologias constituíam um tópico de interesse restrito a filósofos, linguistas, bibliotecários e investigadores de representação do conhecimento. No início da década de 80, porém, os especialistas de inteligência artificial perceberam a relevância da ontologia no processo de descrição do mundo para que os sistemas possam raciocinar e agir sobre ele. Ao longo dos últimos anos as ontologias transformaram-se em recursos frequentemente utilizados no campo da modelação dos sistemas de informação, sendo consideradas hoje como sistemas formais para representar conhecimento num determinado domínio, descrições, semelhantes a uma especificação formal de um programa, dos conceitos e relações que podem existir para um agente ou comunidade de agentes (humanos ou artificiais), actuando como um modelo de referência normalizado para suportar integração de informação e partilha de conhecimento. Como refere Bachimont (2006), a ontologia, na área da ciência computacional, dirige-se a todos os que necessitam de sistematizar a expressão do conhecimento que manipulam.

As ontologias ultrapassaram assim os domínios das bibliotecas, da filosofia e da investigação teórica sobre representação do conhecimento. São agora preocupações dos departamentos de marketing, dos CEOs e da indústria em geral. Desempenham um papel crucial no comércio electrónico e no apoio à interoperabilidade para facilitar a gestão e configuração do conhecimento. São vistas como vocabulários controlados integrados em catálogos, bases de dados, publicações na Web e aplicações de gestão de conhecimento. As ontologias de grandes dimensões são componentes essenciais de aplicações em linha, incluindo pesquisa, comércio e configuração.

Do ponto de vista metodológico, temos uma abordagem interdisciplinar, onde a terminologia, a linguística, a filosofia, a epistemologia e a lógica desempenham um papel fundamental na análise da estrutura de uma dada realidade a um nível elevado de generalização, bem como na formulação de um vocabulário claro e rigoroso. Na última década de séc. XX o termo ontologia tornou-se numa “buzzword” à medida que o comércio electrónico e as linguagens de metadados como a XML, entre outras, emergiram nos planos estratégicos de muitas empresas. A adicionar à partilha do conhecimento, a ontologia tornou-se uma aplicação que conquistou o seu lugar por direito. (Guarino & Welty, 2001) (McGuinness, 2003)

Uma das definições mais emblemáticas de ontologia na área da representação do conhecimento é a de Tom Gruber:

“In the context of knowledge sharing, I use the term ontology to mean a specification of a conceptualization. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or community of agents. This definition is consistent with the usage of ontology as a set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy.” (Gruber, 1993a)

Para Gruber, a palavra ‘ontologia’ tem toda uma história que gera controvérsia na área da inteligência artificial. Segundo Aristóteles, as categorias primitivas possibilitavam dar conta de tudo o que existe. A ontologia computacional vai buscar então ajuda na ontologia filosófica para fazer distinções ontológicas de forma sistemática e coerente. Em ciência computacional, ontologia é um termo técnico denotando um artefacto que é desenhado para um propósito,

permitindo modelar conhecimento sobre algum domínio, real ou imaginário. Neste contexto, não deve ser confundida com filosofia (estudo da existência) nem com epistemologia (estudo do conhecimento e do saber). Deve significar sim, a *especificação explícita de uma conceptualização*¹¹.

Segundo o autor, na década de 80 a comunidade de inteligência artificial apropriou-se do termo para se referir a uma teoria de modelação do mundo e, simultaneamente, a um componente de sistemas de conhecimento, um modelo que permitisse certo tipo de raciocínio automático. Na década de 90 foi definida como uma ‘especificação explícita de uma conceptualização’, ou seja, o conjunto de objectos, conceitos e outras entidades que presumimos existirem em determinada área, bem como as relações entre eles:

“A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold them (Genesereth & Nilsson, 1987). A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.” (Gruber, 1993b, p. 1)

O chamado ‘universo do discurso’ consiste no conjunto de objectos que podem ser representados num formalismo declarativo. Formalmente, uma ontologia é uma declaração de uma teoria lógica definindo um conjunto de primitivos de representação com os quais modelamos o domínio do conhecimento ou do discurso. Esses primitivos são classes, atributos e relações. As definições incluem informação sobre o seu significado, enquanto os axiomas formais restringem a interpretação e o uso bem formado dos termos. A ontologia fornece uma linguagem de comunicação com o agente. Enquanto especificações, as ontologias são sempre

¹¹“An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what ‘exists’ is exactly that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names are meant to denote, and formal axioms that constrain the interpretation and well-formed use of these terms.” (Gruber, 1993a)

uma mistura de partes formais e informais. As partes informais ajudam a explicar alguma coisa aos humanos, as partes formais permitem alguma análise automática: “In the context of database systems, ontology can be viewed as a level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals.” (Gruber, 2008a) Poderíamos considerar as ontologias como a materialização do nível de um conhecimento.

Os termos ‘especificação’ e ‘conceptualização’ geraram polémica¹², e uma das objecções à definição de Gruber de ontologia reside no facto de que seria demasiado alargada, abarcando artefactos tão distantes como glossários ou teorias de lógica¹³. No entanto, para Gruber, o que importa reter é que uma ontologia fornece ferramentas de representação do conhecimento, constituindo-se como um produto de engenharia definido pela aplicação e utilização. A engenharia ontológica preocupa-se em fazer escolhas de representação que capturem as distinções relevantes de um domínio no nível mais elevado de abstracção, mantendo-se o mais clara possível sobre o significado dos termos. Uma ontologia é assim uma especificação de um modelo de dados abstracto (a conceptualização do domínio), independente da sua forma particular: “The key role of ontologies with respect to database systems is to specify a data modeling representation at a level of abstraction above specific database designs (logical or physical), so that data can be exported, translated, queried, and unified across independently developed systems and services.” (Gruber, 2008a)

Uschold e Gruninger acompanham de perto as considerações de Tom Gruber:

“Ontologies are agreements about shared conceptualizations. Shared conceptualizations include conceptual frameworks for modelling domain knowledge; content-specific protocols for communication among inter-operating agents; and agreements about the

¹² Brewster e O’Hara defendem que a definição de Gruber ainda continua válida, embora com algumas restrições. O problema é que não há limites para o que podemos considerar ser uma ‘especificação’. A pesquisa elaborada pelos autores revelou mais de um milhão e seiscentos mil sítios com o termo ontologia. Apropriadamente, recordam as palavras de Simon Buckingham Shum: “If you can’t type it, it doesn’t mean anything.” (Brewster & O’Hara, 2004, p. 9)

¹³ Quando se trata de discutir as questões práticas relacionadas com as ontologias, é típico para as comunidades individuais envolvidas na pesquisa e desenvolvimento, discuti-las usando os seus próprios termos e de acordo com os enquadramentos que lhe são familiares. (Wilson, 2004) Sobre este assunto, consultar igualmente McGuinness (2003).

representation of particular domain theories. In the knowledge sharing context, ontologies are specified in the form of definitions of representational vocabulary. A very simple case would be a type hierarchy, specifying classes and their subsumption relationships. Relational databases schemata also serve as ontologies by specifying the relations that can exist in some shared database and the integrity constraints that must hold for them.” (Uschold & Gruninger, 1996, p. 6)

No seguimento dos debates sobre a utilização do termo, discutiu-se a importância da filosofia na construção e desenvolvimento de ontologias e a necessidade de uma clarificação terminológica. Em filosofia, e segundo Chandrasekaran et al, ontologia seria o estudo dos tipos de coisas que existem, aplicando-se por vezes a expressão de que as ontologias “(...) carve the world at its joints.” (Chandrasekaran, Josephson, & Benjamins, 1999, p. 20) Estes autores consideram, por exemplo, que, dadas as inconsistências no uso do termo, os teóricos utilizam por vezes a forma singular para se referirem a um conjunto específico de termos para descrever a entidade e os tipos de relações em determinado domínio, e a forma plural para falarmos sobre propostas para ontologias para diferentes domínios do conhecimento.

Guarino (1998) apresenta uma proposta diferente. Do ponto de vista metodológico, deve existir uma abordagem muito interdisciplinar, onde a filosofia e a linguística desempenham um papel fundamental na análise da estrutura de uma dada realidade a um nível elevado de generalização e na formulação de um vocabulário claro e rigoroso. Começa por isso por considerar duas interpretações para o termo ontologia, uma contável e a outra não contável: Ontologia e ontologia.¹⁴ Na frase ‘a Ontologia é uma disciplina fascinante’ obtemos uma leitura não contável do termo, referindo-se a uma disciplina filosófica. Por outro lado, na frase ‘a ontologia Cyc’, a leitura é contável, mas contabilizamos dois sentidos diferentes, conforme este seja interpretado pela comunidade filosófica ou pela comunidade dedicada à Inteligência

¹⁴ A este propósito, temos também a citação de Bachimont: “On peut caractériser une ontologie comme une structuration des concepts d’un domaine. Ces concepts sont rassemblés pour fournir les briques élémentaires et exprimer les connaissances dont on dispose dans ce domaine. Il faut bien distinguer une ontologie avec un petit « o » et l’ontologie avec un grand « O », qui est celle étudiée par la philosophie. L’ontologie philosophique signifie la « Science de l’être en tant qu’être » tandis que les ontologies, avec un petit « o » et la marque du pluriel, sont les ontologies mobilisées dans les technologies de l’information et de la communication.” (Bachimont, 2006) http://www.technolangua.net/imprimer.php3?id_article=280

artificial. No sentido filosófico, o termo ontologia é definido como um sistema de categorias que dão conta de uma certa visão do mundo, independente de uma língua particular. No entanto, a utilização mais corrente actualmente faz corresponder o termo ontologia a um artefacto de engenharia, composto por um vocabulário específico que descreve uma certa realidade, acompanhado de uma série de declarações explícitas relacionadas com o significado que se pretende transmitir desse vocabulário.

Em 1995 Guarino et al definem 'conceptualização' como "(...) an intensional¹⁵ semantic structure which encodes the implicit rules constraining the structure of a piece of reality (...)" (Guarino & Giaretta, 1995, p. 6) embora em 1998 Guarino opte por outra abordagem: o termo 'conceptualização' seria utilizado para nos referirmos à perspectiva filosófica. Recorda a origem do termo na obra de Genesereth e Nilsson de 1987¹⁶ como sendo uma estrutura $\langle D, R \rangle$, onde D é um domínio e R é um conjunto de relações relevantes em D. De acordo com Genesereth e Nilsson, uma conceptualização é o conjunto de relações extensionais que descrevem um estado particular de coisas, enquanto a noção que temos em mente é intensional, nomeadamente algo parecido com uma grelha conceptual que super-impomos aos vários estados de coisas. Guarino revê, assim, a definição de conceptualização fazendo uso do aspecto intensional, para obter uma interpretação mais satisfatória à definição de Gruber: ontologia refere-se a um artefacto constituído por um vocabulário usado para descrever uma certa realidade, acrescido de um conjunto de factos explícitos e aceites que dizem respeito ao sentido que se quer atribuir aos termos. (Guarino, 1998)

O autor tenta clarificar em 1997 de forma mais aprofundada o seu posicionamento, analisando e comparando criticamente diversas definições (Guarino, 1997, pp. 297-299):

¹⁵ Uma definição *intensional* consiste numa lista de características do conceito. Uma definição *extensional* é uma enumeração de aspectos de todas as espécies que são do mesmo nível de abstracção. Por exemplo, os planetas do sistema solar são Mercúrio, Vénus, Terra, Marte, Júpiter, Saturno, Urano, Neptuno e Plutão (*ISO-standard 704*). (M. B. Almeida & Bax, 2003)

¹⁶ *Logical Foundations of Artificial Intelligence*, Genesereth and Nilsson, 1987. "A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them." (Gruber, 1993a) <http://ksi-web.stanford.edu/people/gruber/>

1. Gruber (1995) – “An ontology is an explicit specification of a conceptualization.”
2. Wielinga & Schreiber (1993) – “A (AI-) ontology is a theory of what entities can exist in the mind of a knowledgeable agent.”
3. Alberts (1993) – “An ontology or a body of knowledge concerning a particular task or domain describes a taxonomy of concepts for that task or domain that define the semantic interpretation of the knowledge.”
4. van Heijst, Schreiber & Wielinga (1997) – “An ontology is an explicit knowledge level specification of a conceptualization (...) which may be affected by the particular domain and task it is intended for.”
5. Gruber (1994) “Ontologies are agreements about shared conceptualizations. Shared conceptualizations include conceptual frameworks for modelling domain knowledge; content-specific protocols for communication among inter-operating agents; and agreements about the representation of particular domain theories. In the knowledge sharing context, ontologies are specified in the form of definitions of representational vocabulary. A very simple case would be a type hierarchy, specifying classes and their subsumption relationships. Relational database schemata also serve as ontologies by specifying the relations that can exist in some shared database and the integrity constraints that must hold for them”, Shared Reusable Knowledge Bases Mailing List in Ushold and Gruninger (1996), “Ontologies: principles, methods and application.” *The Knowledge Engineering Review*, 11, 93-136.
6. Guarino e Giaretta (1995) (alteração da definição 1) “An ontology is an explicit, partial account of a conceptualization.”
7. Guarino (alteração da definição 6) – “An ontology is an explicit, partial account of the intended models of a logical language.”
8. Shreiber, Wielinga & Jansweijer (1995) – “An ontology is an explicit, partial specification of a conceptualization that is expressible as a meta-level viewpoint on a set of possible domain theories for the purpose of modular design, redesign and reuse of knowledge-intensive system components.”

9. Quine, W. O. (1961). *From a Logical Point of View, Nine Logico-Philosophical Essays*. Cambridge, MA: Harvard University Press. – definição clássica de ontologia: “To be is to be the value of a variable.”

De acordo com Guarino, a primeira e segunda definições apontam para o facto de uma ontologia poder ser muito mais do que uma taxonomia de conceitos, envolvendo em particular as inter-relações e restrições entre conceitos. Relativamente à definição de Wielinga & Schreiber de 1993, esta aproxima-se da noção clássica de “ontological commitment” introduzida por Quine em 1961, ou seja, “a logical theory is ontologically committed to the entities it quantifies over.” Guarino considera que esta definição não apresenta bases de sustentação, porque queremos não só dar conta do que existe, mas também dar conta da estrutura do que existe. Esta estrutura estaria de algum modo implícita na língua que utilizamos. Daí a razão pela qual o termo ‘ontologia’ é muitas vezes utilizado como sinónimo de ‘terminologia’ na comunidade de inteligência artificial. Por outro lado, a definição de van Heijst, Schreiber & Wielinga datada de 1997 clarifica a definição de Gruber, reconhecendo que as ontologias pertencem ao nível do conhecimento e que, parafraseando Guarino, *podem* depender de determinados pontos de vista. É exactamente esse grau de dependência que tem influência na reutilização e no *valor* de uma ontologia. A definição de Alberts apresenta-se mais problemática, sendo que o autor, referenciando sempre os seus comentários à apreciação crítica do artigo de van Heijst, considera que é correcta a observação de que é a interpretação semântica dos termos de um domínio que constitui uma ontologia, embora a formulação desta definição seja dúbia, uma vez que a interpretação semântica dos termos respeitante a uma tarefa ou domínio particular não diz respeito somente à taxonomia, mas envolve também as situações factuais e discursivas ligadas a esse domínio. Interpretando a definição de Gruber de 1994, considera que aqui as ontologias e as conceptualizações são positivamente encaradas como realidades distintas. Uma ontologia, neste sentido, não é uma *especificação* de uma conceptualização, mas sim um acordo, eventualmente incompleto, *sobre* uma conceptualização. De acordo com Guarino e Giarretta (1995) podemos ter diferentes graus de detalhe neste acordo, dependendo do propósito da ontologia. Em constante contraponto com o artigo de van Heijst, e no seguimento da definição de Shreiber, Wielinga & Jansweijer, o autor entende que existe compatibilidade com a definição de Gruber.

Em conclusão, Guarino considera a definição 7 como sendo “the ultimate definition” (Guarino, 1997, p. 298), a versão mais completas das definições 1 e 6. Para Guarino, “(...) the elicitation of the intrinsic structure of domain-knowledge should be the main task of ontology-building tools. The goal of the so-called ontological engineering is to develop theories, methodologies and tools suitable to elicit and organize domain knowledge in a reusable and ‘transparent’ way”. (Guarino, 1997, p. 307) É esta transparência cognitiva que confere o ‘valor acrescentado’ a uma ontologia.

No seguimento da criação da ferramenta ONTOCLEAN, Guarino e Welty (2001) afirmam que hoje o termo ontologia no sentido comercial é tido como sinónimo de ‘modelo conceptual’, quase independente dos seus antecedentes filosóficos. Acrescentam, porém, que deve existir uma diferenciação entre os dois termos: um modelo conceptual é uma implementação de uma ontologia que tem de satisfazer as contrapartidas da engenharia de uma aplicação, enquanto o desenho de uma ontologia é independente de considerações temporais, e o seu único objectivo é especificar a conceptualização do mundo que está por trás de tal aplicação¹⁷.

A perspectiva de Sowa aproxima-se de Gruber e Guarino. Distinguindo, graficamente, ‘Ontologia’ de ‘ontologia’, observa: “Ontology is the study of existence. An ontology is a system of categories for classifying and talking about the things that are assumed to exist.” (Sowa, 2006)¹⁸. O objecto de estudo de uma ontologia é o estudo das categorias de coisas que existem ou podem existir num dado domínio. O seu produto é a própria ontologia, uma lista ou catálogo de tipos de coisas a partir de um determinado ponto de vista, expresso numa determinada linguagem. A combinação dos conceitos ou predicados que constituem a

¹⁷ Guarino et al avançam num artigo datado de 2001 com questões de ordem epistemológica. Parafrazeando os autores: A ontologia é sobre o ‘mundo real’, tal como é visto por um físico? Ou devemos ter em conta a cognição e as complexas relações de dependência entre nós e o mundo? Como os sistemas de conhecimento gerem informação relevante para agentes humanos, as suas ontologias têm de ter espaço para entidades que dependem da nossa percepção e do modo como usamos a realidade para sobreviver e comunicar. (Guarino, Gangemi, Masolo, & Oltramari, 2001)

¹⁸ “The subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called an ontology, is a catalogue of the types of things that are assumed to exist in a domain of interest **D** from the perspective of a person who uses a language **L** for the purpose of talking about **D**. The types in the ontology represent the predicates, word senses, or concept and relation types of the language **L** when used to discuss topics in the domain **D**”. (Sowa, 2006)

ontologia com a lógica, resulta numa linguagem capaz de expressar relações sobre essas entidades num domínio específico.

Para Beck e Pinto (2002) as ontologias foram propostas como sendo capazes de resolver o problema de terminologia diferente para o mesmo conceito ou a utilização do mesmo termo para conceitos diferentes: “Ontologies have been proposed to solve the problems that arise from using different terminology to refer to the same concept or using the same term to refer to different concepts.”(Beck & Pinto, 2002, p. 6) Consideradas no contexto da inteligência artificial como o estudo e definição da teoria formal dos objectos, ou seja, o estudo das características básicas de toda a realidade, as ontologias devem contrastar com a sua interpretação metafísica – o estudo da essência dos seres. As dificuldades que surgem na partilha e reutilização do conhecimento devem-se, principalmente, à falta de consenso sobre o vocabulário utilizado e às diferentes interpretações em modelos de domínio: “Terms in an ontology are a representation of concepts. We should stress that in an ontology concepts are represented, not words.” (Beck & Pinto, 2002, p. 7)

Beck e Pinto, recuperando a definição de Borst de 1997 – “Ontologies are defined as a formal specification of a shared conceptualization”¹⁹, e de Gruber de 1993 – “An ontology is an explicit specification of conceptualization” – citam a definição complementar posterior de Borst de ontologia: “(...) an explicit formal specification of a shared conceptualization” (Beck & Pinto, 2002, p. 9) e, baseados em Studer et al (1998) explicam cada conceito desta definição: ‘explícito’ significa que o tipo de conceitos utilizados e as restrições ao seu uso são explicitamente definidos; ‘formal’ refere-se ao facto de que deve poder ser lida por máquinas; ‘partilhada’ reflecte a noção de que o conhecimento representado numa ontologia captura um conhecimento consensual, aceite por um grupo, e não propriedade de um indivíduo; finalmente, ‘conceptualização’ indica um modelo abstracto de um fenómeno no mundo através da identificação dos conceitos relevantes desse fenómeno.

¹⁹ W.N.Borst, *Construction of Engineering Ontologies*. Centre for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1997.

Reportando à definição de Guarino (“an ontology is an explicit, partial account of a conceptualization”), Beck e Pinto recordam que nesta definição o conceito de conceptualização não deve ser entendido da mesma forma que na definição de Nilsson e Genesereth (1987)²⁰. O propósito de uma ontologia consiste na exploração de informação específica de um domínio através da representação do significado²¹ dos termos desse domínio e utilizar essas representações de significado para organizar uma colecção com objectivo de realizar pesquisa de forma mais precisa e mais rápida:

“Ontologies attempt to exploit domain-specific information by representing the meaning of terms within a domain, and using these meaning representations to organize the collection and make search more accurate. Exactly how meaning is represented, how to organize collections around this representational framework, and how search and other inference operations are implemented are all technical issues related to ontology construction.” (Beck & Pinto, 2002, p. 1)

Assim, os autores entendem que a diferença principal entre as definições de ontologia reside na formalidade e na natureza consensual do conhecimento representado, pelo menos dentro de um determinado grupo, para que possa ser reutilizado em sistemas de conhecimento diferentes. É para isto que as ontologias são construídas. Mesmo aqui, não existe consenso quanto ao grau de formalidade que uma ontologia deve conter.

Do nosso ponto de vista, pensamos que a afirmação de Beck e Pinto (“Ontologies have been proposed to solve the problems that arise from using different terminology to refer to the same concept or using the same term to refer to different concepts.”) (Beck & Pinto, 2002, p. 6) não se traduz, na prática, por uma solução para este problema. Nem as ontologias conseguem resolver o problema da terminologia, nem a terminologia consegue solucionar as dificuldades de representação do conhecimento através das ontologias. Existe, sim, um

²⁰ Genesereth, Michael R. & Nilsson, N.J. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, Los Altos, California, 1987 in (Beck & Pinto, 2002).

²¹ Optámos pela tradução de ‘meaning’ por significado, porque pensamos que reflecte o que os autores pretendem dizer. Contudo, sentido, significado e significação são três conceitos diferentes aos quais iremos aludir no último capítulo teórico.

esforço conjunto no sentido de uma aproximação contínua a uma plataforma de entendimento que permita uma partilha e reutilização cada vez mais consensual, por um lado, e automatizada, por outro.

Para os sistemas de inteligência artificial, o que existe é o que pode ser representado numa linguagem formal. Representamos objectos, conceitos e relações de forma explícita através da ontologia. Podemos, pois, concluir que, embora uma ontologia seja um artefacto de engenharia que pode ser partilhado entre pessoas, organizações e sistemas informáticos, a representação sistemática da realidade não é garantia de amplitude de conhecimento, se o que existe é somente o que pode ser representado formalmente. Numa ontologia é suprimida a vertente cultural do termo, uma vez que estamos perante uma representação de dados descontextualizados. Isso representa um ponto de debate importante entre os defensores de uma análise terminológica atenta ao aspecto discursivo, e os proponentes de uma abordagem terminológica centrada na parte conceptual regida por princípios formais.

Até hoje, as ontologias mais completas foram elaboradas na área das ciências naturais, o que é compreensível. O mundo da realidade social é muito mais difícil de capturar. No entanto, quase todos os objectos do mundo são afectados pela realidade social. Apesar de toda a explosão de interesse nas ontologias, discute-se muito pouco sobre o âmbito actual de conhecimento que uma ontologia pode representar. A inteligência artificial considera que devemos representar o conhecimento o melhor possível. Mas será que essa concepção do conhecimento é adequada? Temos consciência que certo tipo de conhecimento não é 'representável' e a maior parte das definições de ontologia especificam que representa um entendimento comum e partilhado de estruturas conceptuais.

Entendemos que se uma ontologia tem por missão primeira ajudar a compreender o mundo através da representação de conhecimento sobre o mundo, devemos encará-la em três dimensões essenciais: conhecimento (uma conceptualização para entendermos o mundo), língua (para falarmos sobre o mundo) e lógica (para representar a manipulação do nosso entendimento sobre o mundo). A popularidade das ontologias enquanto panaceia para os problemas de organização e representação do conhecimento exige propriedades que são extremamente difíceis de reunir no mesmo contexto. Podemos normalizar linguagens de representação. Poderíamos, hipoteticamente, normalizar a língua. Não podemos normalizar o

pensamento. Por isso se afirma que uma ontologia computacional é a especificação de uma conceptualização dos objectos de um domínio segundo um determinado ponto de vista imposto pela aplicação.

3. Considerações teóricas sobre o conhecimento

3.1 Lógica

Wouldn't it be neat if we could write down everything people know in a formal language? Damn it, let's give it a shot! (...) If we want to be able to represent anything, then we get further and further from the practicalities of frame organization, and deeper and deeper into the quagmire of logic and philosophy. [McDermott 1993] in (Guarino, 1995, p. 16)

Uma ontologia pretende representar conhecimento. Falar sobre conhecimento na área da inteligência artificial implica perceber a sua transformação em dados formalizados. A Lógica, enquanto ciência constituída por um suporte teórico e um conjunto de práticas, pode ser definida de várias formas: o estudo dos processos válidos e gerais pelos quais atingimos a verdade, a ciência das leis do pensamento, ou ainda o estudo dos princípios da inferência válida.

Existe uma diversidade de áreas de estudo abrangidas por esta matéria. Desenvolvendo-se a partir de normas e raciocínios para discussão, ou seja, regras para bem argumentar, surgiu no Séc. IV a.C. com Aristóteles, um dos fundadores da lógica na filosofia ocidental. Descreveu-a enquanto dialéctica no tratado *Tópicos* e enquanto silogismo em *Primeiros Analíticos*, um tratado posterior que encara a lógica na sua essência, tal como hoje vemos a lógica matemática e simbólica contemporâneas. Foi constituída como uma ciência autónoma que se dedicava ao estudo dos actos do pensamento do ponto de vista da sua estrutura ou forma lógica (conceito, juízo, raciocínio, demonstração), daí designarmos a lógica aristotélica como lógica formal.

A lógica, enquanto sistema de regras para bem pensar e argumentar a fim de se alcançar o conhecimento verdadeiro, foi distinguida da ontologia enquanto sistema de categorias e leis

sobre o que existe²². Para uns, a ontologia seria um questionar intuitivo e informal sobre as categorias das entidades, enquanto a lógica constituiria a elaboração sistemática, formal e axiomática do material pré-digerido pela ontologia. Segundo Cocchiarella (2001), uma lógica formal rica basear-se-ia numa *lingua philosophica*, oriunda do pensamento de Descartes, que acreditava que esta se encontrava na base de todas as línguas, representando a forma da razão, e não a natureza das coisas no mundo. Essa linguagem conteria uma *mathesis universalis*, mas na sua construção teria de existir uma análise de todos os conteúdos da consciência transformados em ideias simples, os seus constituintes últimos.

Tal como tivemos oportunidade de observar no capítulo anterior, os padrões lógicos mais antigos expressos em linguagem controlada são os tipos de declaração do sistema silogístico aristotélico, onde cada regra combina uma premissa maior com uma menor para chegar a uma conclusão. Leibniz, Frege, Peirce e Husserl deram continuidade ao pensamento aristotélico. Para Leibniz (1660) existia uma linguagem universal subjacente às línguas naturais, com um sistema de símbolos que nos permitiria analisar todos os conceitos possíveis da ciência – o *ars combinatorial* – e um sistema completo de dedução para caracterizar todas as formas válidas de argumentação – o *calculus ratiocinator*. Assim, seria possível obtermos uma espécie de enciclopédia unificada da ciência, onde o carácter universal conteria uma ontologia geral e ontologias específicas para cada área científica. Em 1879 Frege aprofundou o *calculus ratiocinator* de Leibniz com *Begriffsschrift*, um sistema de anotação lógica para árvores, sendo o primeiro sistema completo de lógica de primeira ordem. Peirce desenvolveu em 1883 uma anotação algébrica para cálculos de predicado que, juntamente com a contribuição de Giuseppe Peano nos símbolos, se tornou o método de anotação mais utilizado na lógica actual. Em 1897, Peirce formulou o terceiro sistema completo, os gráficos existenciais. Outra grande referência da lógica formal foi o filósofo Edmund Husserl. O termo ‘ontologia formal’ foi usado pela primeira vez por Husserl em *Logical Investigations* (1900-1901) para designar o estudo das estruturas e relações formais – nomeadamente parte/todo – exemplificadas dentro de várias ciências materiais. Para Husserl, havia uma distinção entre lógica formal e ontologia

²² A ontologia procede a ordenações e classificações, optando por uma de duas perspectivas: aristotélica, voltando-se para a estrutura da realidade (tratados *Categorias* e *Metafísica*, Aristóteles), ou kantiana, voltando-se para a razão e o pensamento (*Crítica da Razão Pura*, Kant). (Cocchiarella, 2001)

formal. A lógica formal lida com as interconexões das verdades – relações de inferência, consistência e validade. A ontologia formal lida com as interconexões das coisas – objectos, propriedades, parte e todo.

Sowa comenta com propriedade que, para armazenarmos e partilharmos conhecimento, temos de transmiti-lo num formato acessível a humanos e máquinas. Não existe uma linguagem universal que abarque estes dois mundos. A língua natural pode potencialmente representar tudo o que diz respeito à experiência humana, denotativa ou conotativamente, incluindo a sua auto-representação através de meta-linguagens. Pelo contrário, na lógica cabe somente o que pode ser dito claramente e de forma precisa em qualquer língua natural. Parece simples. Mas não é. Chegar a um nível de precisão satisfatório na linguagem lógica é muito difícil. Porém, nenhuma outra linguagem consegue preencher os requisitos de um sistema definicional completo.

A lógica pura é ontologicamente neutra. Não faz qualquer pré-suposição sobre o que existe ou o que poderia existir em nenhum domínio e em nenhuma língua, uma linguagem plana, na qual um predicado unitário pode representar ou um conceito ou uma propriedade, enquanto os predicados binários podem representar ou atributos (relações internas) ou relações entre conceitos (relações externas). Essa neutralidade, expressa na lógica de primeira ordem²³, constitui uma vantagem, conferindo-lhe amplitude e robustez. Contudo, ontologicamente falando, pode ser contraproducente, no sentido de que os formalismos devem reflectir uma estrutura do mundo, bem como as escolhas ontológicas feitas pelo utilizador²⁴.

²³ A linguagem de primeira ordem vai captar relações entre indivíduos de um mesmo universo de discurso e a lógica de primeira ordem vai permitir alcançar particularidades de uma propriedade geral dos indivíduos de um universo de discurso, assim como derivar generalizações a partir de factos que valem para um indivíduo arbitrário do universo de discurso. Também conhecida como cálculo de predicados de primeira ordem, a lógica de primeira ordem é um sistema lógico que vai além da lógica proposicional, consistindo num conjunto de axiomas e de declarações dedutíveis a partir desses axiomas.

²⁴ Daí associarmos à lógica de primeira ordem certo tipo de estruturas de *frames* ou módulos para capturar inter-relações entre fragmentos do conhecimento que não podem ser facilmente capturados pela lógica de primeira ordem. Deste modo, os formalismos de representação do conhecimento manifestam maior eficiência computacional e uma maior facilidade de manutenção. (Guarino, 1995)

A língua natural levanta, naturalmente, imensas questões sobre a sua formalização. Mesmo controlada em termos de sintaxe e semântica, não pode ser utilizada para definição formal de conceitos. Para definirmos formalmente uma ‘conceptualização’, necessitamos de uma linguagem específica, representando conhecimento de tal forma que possa ser partilhável e reutilizável. Roche adverte, porém, que a opção pela abordagem hipotética e dedutiva do sistema formal obriga a uma aceitação das suas restrições, reflectindo-se na conceptualização do domínio. (Roche, 2007a) Considera, no entanto, que não devemos encarar o raciocínio formal e informal como incompatíveis. O primeiro deve ser visto como uma aplicação sistematizada e disciplinada das técnicas utilizadas no segundo, sendo que a base dos dois é a analogia, considerada como o processo de encontrar padrões comuns em estruturas diferentes.

3.2 Sobre o conhecimento

No tempo de Platão, comenta O’Hara (2002), surgiram as primeiras interrogações sobre o conhecimento: o que era, de onde vinha, a sua fiabilidade, e como poderia ser categorizado. Alguns opinavam que o homem nunca poderia realmente saber nada com certeza absoluta. Outros, como Platão, pensavam que a filosofia e a ciência teórica poderiam revelar verdades sobre como o universo é constituído e informação sobre como o homem deveria agir. Para este filósofo, a maioria das pessoas realmente não possuía conhecimento verdadeiro, somente crenças e opiniões muito subjectivas. Isto porque é fácil confundirmos o que vemos, com o que é verdadeiro. O conhecimento verdadeiro seria permanente. O melhor exemplo era a matemática, separada do mundo e acessível somente à mente. Hoje em dia, refere O’Hara, a maioria dos filósofos discordaria de Platão sobre a perenidade de certos factos. O conhecimento moderno parece mudar constantemente. Desconfia-se das verdades universais, separadas dos interesses e desejos do homem. Questiona-se o que é a ‘razão’ ou o que é ‘racional’. Os axiomas básicos da matemática e da lógica parecem mais fracos, sendo mais difícil saber que função a palavra ‘verdade’ tem e quão ideologicamente contaminado o nosso pensamento deve ser. O facto é que ninguém consegue dizer-nos o que é realmente o conhecimento.

Na nossa era, a inteligência artificial simbólica surgiu da tentativa de criar sistemas directamente a partir de conhecimento acerca de processos, isto é, declarativamente,

permitindo aos computadores realizar dedução automática. Esta área baseia-se fundamentalmente na separação entre conhecimento e processos dedutivos ou inferências.

3.2.1 Nível do conhecimento

Para os sistemas de informação, o conceito de conhecimento aparece ligado ao conceito de comunidade. Miguel Cornejo define comunidades como “(...) groups where knowledge and best practices in a given field are developed, nurtured and transmitted through the social interaction of its members (...) where members’ tacit knowledge is transmitted to other members, and where this know-how is occasionally codified and made explicit.” (Cornejo, 2003, p. 1)

Baseados em Newell (1980, 1982) e Cornejo (2003), observamos que na tradição da inteligência artificial, o conhecimento é definido numa forma estritamente funcional. Existem vários componentes funcionais que compõem um sistema inteligente: percepção, memória, processamento, entre outros. Um sistema inteligente está inserido num ambiente de tarefas: uma declaração de tarefa dá entrada através de um componente de percepção e é codificada numa representação inicial. Desenvolve-se assim um ciclo de actividade no qual ocorre um reconhecimento de um método a utilizar para resolver o problema. Esse método vai buscar a memória do conhecimento do mundo geral. O problema é detectado pelas estruturas de dados e processado de tal forma que fique disponível uma solução, detendo, ao mesmo tempo, o conhecimento do mundo. Ou seja: umas estruturas representam o problema, outras representam o conhecimento do mundo. Newell diz-nos que o termo ‘representação’ é usado quase tecnicamente e de forma clara em inteligência artificial e na área computacional. O termo ‘conhecimento’ é usado informalmente, apesar da sua prevalência em expressões como ‘engenharia do conhecimento’.

Cornejo clarifica que conhecimento é “(...) the information required to satisfy a need (...) knowledge is such when it enables a person to do something. (...) If it is not good for anything but the person can understand it, it will simply be called ‘information’. If the person can’t make sense of it, it will be merely considered ‘data’.” (Cornejo, 2003, p. 2) Se um sistema tem (e pode usar) uma estrutura de dados que representa alguma coisa (um objecto, um

procedimento...), podemos também, segundo Newell, dizer que o sistema em si tem conhecimento, nomeadamente o conhecimento que está inserido na representação sobre essa coisa. Até então, a 'representação' ocupava o lugar central, como o *locus* da verdadeira inteligência, tendo a lógica como suporte principal. E o autor pergunta: qual a natureza do conhecimento? De que forma está relacionado com a representação? O que tem um sistema quando tem conhecimento? O modo como a inteligência artificial utiliza o conceito de conhecimento difere de todas as outras áreas.

Newell entende que um nível consiste num meio que vai ser processado, existindo três especificações consecutivas:

- O *nível de conhecimento ou epistemológico*, que consiste numa especificação abstracta do conhecimento do domínio ou problema que se deseja modelar;
- O *nível lógico*, que converte as especificações de conhecimento em declarações de lógica formal;
- O *nível de implementação*, que codifica estas declarações em linguagem computacional.

O nível do conhecimento é um agente²⁵, um nível de descrição que se encontra acima do símbolo. Um agente processa o seu conhecimento para determinar as acções a implementar. Uma descrição do nível de conhecimento é baseada no seguinte modelo de comportamento de um agente: i) o agente possui *conhecimento*; ii) algum desse conhecimento constitui os *objectivos* dos agentes; iii) o agente tem a capacidade de desempenhar um conjunto de *acções*; iv) o agente escolhe as acções de acordo com o princípio da *racionalidade*; v) o agente vai seleccionar uma acção a desenvolver a seguir que, de acordo com o seu conhecimento, resulta no alcance de um objectivo.

Se um observador externo é capaz de atribuir a um agente algum objectivo (encontrar comida, eliminar um obstáculo) e se o mesmo observador vê que esse agente está prestes a conseguir

²⁵ Nesta área científica, falamos de agentes no sentido artificial, uma entidade autónoma que observa e age num determinado ambiente para alcançar um objectivo.

esse objectivo de forma sistemática e racional, então o observador atribui-lhe conhecimento. Usando as palavras de Newell, o conhecimento é "(...) whatever can be ascribed to an agent, such that its behavior can be computed according to the principle of rationality". (Newell, 1980, pp. 9-10)

Comentando o raciocínio de Newell, Guarino (1995) refere que o critério avaliador relevante para a base de conhecimento assim concebida não é a verdade, mas a utilidade funcional relativamente aos objectivos atribuídos ao agente. As representações existem ao nível do símbolo, sendo sistemas (estruturas de dados e processos) que concretizam um corpo de conhecimento no nível do conhecimento. Este serve como uma especificação do que uma estrutura de símbolos é capaz de fazer²⁶. Dito de outra forma: a representação consiste num sistema para dar acesso a um corpo de conhecimento. É a estrutura no nível do símbolo que concretiza o conhecimento, ou seja, reduz o conhecimento ao próximo nível inferior dos sistemas informáticos.

Assim, para Newell, o nível do conhecimento é uma aproximação. O conhecimento do mundo não pode ser capturado numa estrutura finita, pois o mundo é demasiado rico. O conhecimento é *sobre* o mundo. Na medida em que um agente pode seleccionar acções baseadas em verdades – leia-se, proposições - acerca do mundo, qualquer estrutura candidata ao conhecimento tem de conter essas verdades. Assim, o conhecimento enquanto estrutura tem de conter pelo menos tanta variedade como o conjunto de todas as verdades às quais um agente pode responder. O sistema informático tem a tarefa de gerar selecções de acções para objectivos, condicionadas por estados do mundo. O conhecimento que interessa requer gerar somente o que é relevante para a tarefa de forma inteligente. A teoria do conhecimento deve, pois, associar um conjunto de expressões lógicas ao conhecimento, precisamente por se constituir como uma aproximação, radicalmente imperfeita nuns casos, razoavelmente completa noutros²⁷. Porém, o autor prevê a existência de um segundo nível de aproximação: o nível do símbolo, que fornece um meio de ‘mecanizar’ um comportamento, sendo orientado

²⁶ Recordamos aqui a definição de Gruber de ontologia.

²⁷ Newell dá o exemplo de conhecimento enquanto uma aproximação muito incompleta (prever se uma pessoa se lembra de um número de telefone) e uma boa aproximação (prever se uma pessoa consegue encontrar o quarto na sua casa).

para o sistema. O nível do conhecimento é um nível de descrição acima do símbolo, fornecendo, por seu turno, o meio de 'racionalizar' o comportamento de um sistema do ponto de vista de um observador externo. O observador trata o sistema como uma 'caixa preta', actuando 'como se' possuísse certo conhecimento sobre o mundo e usa este conhecimento de um modo perfeitamente racional para chegar aos seus objectivos. 'Extrair' conhecimento é participar com os processos simbólicos na execução da acção, na medida em que o conhecimento levará à selecção dessas acções no nível do conhecimento. Para determinar exactamente que conhecimento se encontra na representação e caracterizá-lo, necessitamos do uso da lógica.

Tornar explícitas as noções de conhecimento e representação traria, para Newell, consequências evidentes: conhecimento e racionalidade ficariam intimamente relacionados; a existência de dois níveis (nível do conhecimento e nível do símbolo) permite desenvolver cada um de forma adequada; o conhecimento permanecerá para sempre abstracto, não é algo que seja acessível fisicamente, constituindo-se como uma aproximação com muitas falhas enquanto modelo adequado para um agente. Para se tornar viável, tem de estar ligado a uma representação de nível simbólico, sendo que a lógica será fundamentalmente uma ferramenta de análise do nível do conhecimento.

O próprio Newell faz a associação com a filosofia, comentando que a natureza da mente e a natureza do conhecimento têm sido preocupações clássicas desta disciplina. A observação de uma aproximação entre filosofia e inteligência artificial confirma as preocupações centrais com a representação do conhecimento e com o processamento de língua natural. Mas existem diferenças. A filosofia preocupa-se com a certeza: quando podemos confiar no conhecimento? A inteligência artificial é, segundo o autor, indiferente a problemas de certeza absoluta. Daí a distinção entre 'knowledge' e 'belief'. A inteligência artificial, partindo do pressuposto que todo o conhecimento contém erros, designou os sistemas de 'knowledge systems'.

Segundo Freitas (2003), a história da inteligência artificial confunde-se, de alguma forma, com o nível do conhecimento ou epistemológico, assumindo uma grande importância nos sistemas simbólicos, dado que a motivação principal do paradigma declarativo era modelar sistemas num nível mais elevado. Neste nível ainda não existe um compromisso com a implementação, estando muito próximo com o conhecimento do domínio a ser modelado. Assim, numa

primeira fase, tentava-se colocar o conhecimento do lado de fora dos sistemas, pois ainda não se vislumbrava a potencial utilização do conhecimento por entidades de software, nomeadamente por ontologias.

Musen (1992) comenta, a este propósito, que quando falamos de partilha e reutilização de conhecimento, à primeira vista há uma presunção de que o conhecimento é um bem que pode ser duplicado e andar de um lado para o outro, uma substância que pode ser adquirida a partir de especialistas humanos e transferida entre sistemas informáticos. Partilhando a opinião de Newell, Musen sublinha que o conhecimento deve ser encarado como uma abstracção que nunca poderá ser reduzida a discurso escrito. O conhecimento é o que um observador atribui a um agente inteligente (humano ou não) e que lhe permite construir o comportamento do agente como racional, no sentido em que alcança os seus objectivos. As estruturas de dados, os símbolos, representam o conhecimento. Distinguimos, portanto, entre símbolos (representações do conhecimento) e conhecimento (capacidade para um comportamento racional que essas representações podem gerar). Assim, todas as bases de conhecimento informatizadas têm significado somente quando são processadas por um intérprete, pois só por si não são suficientes para capturar conhecimento. A interpretação (por um humano ou por uma máquina) é que fornece o significado ao conhecimento. Não se trata, pois, de transferir conhecimento de um lado para o outro, mas sim de transferir bases de conhecimento. Para Musen, Newell introduziu a noção fundamental de que existe um *nível de conhecimento* no qual os informáticos podem descrever o comportamento de sistemas inteligentes que é independente das estruturas de dados usadas para codificar esses conhecimentos ao *nível do símbolo*.

A posição de Newell sobre a imperfeição do conhecimento será reforçada com Davis et al uma década depois, em 1993, num artigo intitulado "What is Knowledge Representation?" (Davis, Shrobe, & Szolovits, 1993). Davis inicia a sua reflexão com a conclusão que, embora a representação do conhecimento seja um dos conceitos mais centrais e familiares em inteligência artificial, a questão fundamental consiste em defini-lo. Vários autores escreveram sobre a variedade de representações, sobre as várias propriedades que a representação deve ter, ou ainda sobre as propriedades que são importantes para a noção de representação em

geral. Abordando directamente esta questão, enumera cinco funções que uma representação pode desempenhar.

1) Uma representação do conhecimento é, acima de tudo, um substituto para as coisas, usado para permitir a uma entidade determinar consequências pensando e não agindo, isto é, através do raciocínio sobre o mundo e não agindo sobre ele.

2) É um conjunto de compromissos ontológicos, ou seja, uma resposta à questão: em que termos devemos pensar sobre o mundo?

3) É uma teoria fragmentária de raciocínio inteligente, expressa em termos de três componentes:

i) a concepção fundamental da representação de raciocínio inteligente;

ii) o conjunto de inferências que a representação sanciona;

iii) o conjunto de inferências que recomenda.

4) É um meio para uma computação pragmática eficiente, ou seja, o ambiente computacional no qual o pensamento é concebido. Uma contribuição para esta eficiência pragmática é dada pela orientação que uma representação fornece na organização da informação, de forma a facilitar a que sejam tomadas as inferências recomendadas.

5) É uma forma de expressão humana, a língua na qual dizemos coisas acerca do mundo.

Cada uma destas funções requer um tipo diferente de representação e, por sua vez, um conjunto diferente de propriedades. Vejamos: o acto de raciocinar é um processo que se dá internamente e a maior parte das coisas sobre as quais queremos raciocinar existem somente no exterior. Isto constitui, para Davis, um facto intransponível e uma dicotomia inevitável enquanto padrão de raciocínio. A representação funciona, assim, como um substituto, algo que está 'pela coisa' e 'em vez da coisa' que existe no mundo e que substitui a interacção directa com o mundo.

Para Davis, encarar as representações como substitutos leva naturalmente a duas questões importantes. A primeira questão sobre um substituto é a sua suposta identidade: para que

serve um substituto? A segunda questão é a fidelidade: quão aproximado está um substituto da coisa real? Que atributos do original vai capturar e tornar explícito, e quais omite? A fidelidade perfeita é, na generalidade, impossível, tanto na prática como na teoria. É impossível porque qualquer coisa que não a própria coisa é necessariamente diferente da própria coisa (mesmo que seja somente na localização). Posto de outra forma: a única representação completamente fiel do objecto é o próprio objecto. Esta visão da representação serve tanto para objectos tangíveis, como para noções abstractas, processos, credos ou categorias, permitindo-lhes ser descritas dentro de uma entidade para que se possa raciocinar sobre ela.

Partindo, então, da certeza de que os substitutos imperfeitos são inevitáveis, Davis considera duas consequências: a primeira é a de que, ao descrever o mundo natural, vou necessariamente mentir, ou pelo menos omitir, a sua complexidade. A segunda é a de que todo o raciocínio bem fundamentado sobre o mundo chega invariavelmente a conclusões incorrectas, independentemente do processo de raciocínio utilizado e da representação empregue. Na verdade, uma boa parte da arte de seleccionar uma boa representação é encontrar uma que minimize, ou idealmente elimine, o erro. Torna-se necessário, por isso, avaliar os custos e os benefícios dos métodos que utilizam raciocínio. Se todas as representações são aproximações imperfeitas do real, captando umas coisas e ignorando outras, quando seleccionamos uma representação, estamos simultaneamente a tomar decisões acerca de como vemos o mundo. Dito de outra forma, seleccionar uma representação significa optar por um conjunto de compromissos ontológicos. Não se trata tanto de saber o que existe no mundo, mas como o vemos. Trata-se, na verdade, de uma espécie de lente que determina o que podemos ver, tornando nítidas umas partes e difusas outras. O efeito focagem/desfocagem não constitui algo acidental nos contornos da representação: é a sua essência. Conclui-se que, curiosamente, é útil que assim seja, porque a selecção dos compromissos dá oportunidade de focar a atenção em aspectos do mundo que acreditamos serem relevantes. Perante a complexidade do mundo natural, necessitamos de orientação para saber o que devemos ignorar e o no que devemos atentar. O que ressalta numa ontologia, a sua informação essencial, não é a linguagem de representação, mas o conjunto dos conceitos oferecidos como uma forma de pensar sobre o mundo.

No mesmo ano²⁸, Clancey defende uma visão de modelação (*modelling view*) da aquisição de conhecimento, considerando que uma base de conhecimento não é um repositório de conhecimento extraído da mente de um especialista (*transfer view*), mas o resultado de uma actividade de modelação cujo objecto é o comportamento observado de um agente inteligente embebido num ambiente externo²⁹. Para Clancey, a metáfora “knowledge acquisition bottleneck” é enganadora. Sugere que o problema de desenvolvimento de uma base de conhecimento se resume a espremer uma grande porção de conceitos e relações já formados através de um canal de comunicação estreito. Pelo contrário, a aquisição de conhecimento normalmente envolve inventar novas linguagens para modelar experiências que nunca foram articuladas. Isto porque a reinterpretação humana de regras e procedimentos é metafórica, baseada em categorizações de percepções pré-linguísticas e coordenação sensorial e motora não deliberada. A aquisição de conhecimento é um processo de desenvolvimento de modelos informáticos, que ultrapassa a mera tarefa de ‘capturar conhecimento’. Não é somente um processo de extracção de factos e regras ‘escritas’ na mente dos especialistas. Podemos representar conhecimento, mas as representações não são conhecimento em si mesmo, mas sim um mapa que descreve um território.

Comentando Clancey, Guarino entende que, como o conteúdo de uma base de conhecimento se referiria a uma realidade objectiva em vez da ‘mente’ de um agente, parece claro que – de acordo com a *modelling view* – o conhecimento estaria muito mais relacionado com a noção clássica de verdade, entendida como a correspondência para com o mundo real, e menos dependente da forma particular como um agente inteligente persegue os seus objectivos (Guarino, 1995). No entanto, as metodologias de modelação de conhecimento tenderiam a centrar-se no comportamento do agente, estando o conhecimento do domínio fortemente dependente da tarefa. Neste enquadramento, só o conhecimento relevante do domínio é levado em consideração. Mas existe um pormenor: se uma teoria de domínio não descreve, ainda que de forma parcial, uma estrutura inerente do domínio, então o que descreve? Para Guarino, descreve a mente do agente, precisamente o que a visão de modelação de Clancey

²⁸ O artigo foi publicado mais tarde em 2007, e será nessa versão que nos vamos basear.

²⁹ Uma visão semelhante é adoptada por Gruber.

pretendia evitar. A questão da modelação do conhecimento leva Guarino a concluir que, relativamente à abordagem de Clancey, uma base de conhecimento só adquirirá um valor por si se o conhecimento que contiver for realmente verdadeiro, correspondendo ao mundo para além da base de conhecimento. Assim, o estudo da ontologia, entendida como um ramo da filosofia que lida com a natureza *a priori* da realidade, pode beneficiar do processo de construção do conhecimento. Esta posição relaciona-se com a questão do interesse da inteligência artificial na natureza do raciocínio em detrimento do interesse na natureza do mundo real.

Voltando à pergunta de Newell (qual é a natureza do conhecimento?) Clancey (2007) reformula a questão: qual é a natureza da engenharia do conhecimento? E, mais uma vez, observa: a primeira preocupação da engenharia do conhecimento é modelar sistemas no mundo, e não reproduzir o modo como as pessoas pensam, sendo que as descrições ao nível do conhecimento caracterizam o comportamento humano num determinado ambiente social, e não estruturas físicas armazenadas dentro da cabeça³⁰. Dispomos hoje de sistemas especialistas de segunda geração que podem ser avaliados em termos da sua adequação ao modelo que se constrói para um determinado objectivo. Deste ponto de vista, Clancey define a engenharia do conhecimento como uma metodologia para modelar processos de forma qualitativa através de redes que descrevem relações causais, temporais e espaciais. Todos os sistemas especialistas têm dois processos explícitos: i) um modelo de um sistema no mundo (o domínio); ii) um modelo de processos de raciocínio (procedimento de inferência). As técnicas de segunda geração dos sistemas especialistas produzem uma biblioteca cada vez maior de abstrações, permitindo que sejam construídos novos programas reutilizando as representações existentes e os procedimentos de inferência. Em resumo, todas as bases de conhecimento contêm modelos de sistemas no mundo.

³⁰ O desafio da aquisição de conhecimento hoje é clarificar o que estamos a fazer (modelação informática), clarificar os problemas difíceis (a natureza do conhecimento e representações) e reformular o nosso programa de pesquisa de acordo com isso (como colaborar com os cientistas e utilizadores). (Clancey, 2007)

Um especialista humano serve como informador sobre como um dado sistema tende a comportar-se, e como pode ser desenhado ou controlado para gerar determinados comportamentos. Para Clancey, parte da confusão em relacionar bases de conhecimento a comportamento humano é que trabalhamos ao contrário dos nossos modelos: atribuímos propriedades do computador a pessoas. Explicamos o comportamento humano em termos de recuperação e interpretação de regras armazenadas, semelhantes à natureza estática das regras armazenadas na memória de um computador. Vemos o comportamento humano como sendo causado por estruturas simbólicas. Isto é verdade para o comportamento do sistema informático, mas é errado assumir que é verdadeiro para as pessoas. Não podemos confundir representações do conhecimento com as estruturas neuronais presentes no cérebro e que coordenam a nossa actividade. Por vezes esquecemos que um sistema especialista é somente um modelo, dado que as nossas representações influenciam substancialmente o modo como vemos as coisas. Quando falamos, não estamos a traduzir representações internas do que as nossas palavras significam; estamos a criar as representações na nossa actividade. Representações interpretáveis só existem fisicamente nas declarações de um observador, em desenhos, em programas informáticos, em discurso silencioso. Representar significado é um acto de percepção subsequente.

Baseado na opinião de Agre (1988), Clancey concorda que as representações, incluindo as representações do conhecimento, estão sempre abertas à interpretação; o seu significado nunca é fixo ou definido, mas sempre relativo às referências do observador no próximo acto de interpretação. A utilização humana das representações envolve uma nova conceptualização, e não a manipulação sintáctica de definições e significados, que é tudo o que os programas de computador conseguem fazer. A essa conceptualização, que nos permitem agir, chamamos conhecimento:

“A knowledge-level pattern, such as a natural language grammar, characterizes the product of interaction, how behavior routines and the world appear; it doesn't describe structures in the head. Put another way, the neural structures coordinating perception and action in people come into being in the course of interaction itself. Thus, new knowledge (capability to speak, perceive, and act) develops as new coordinations in the

course of behavior itself; people are not automatons rotely executing stored programs”.

(Clancey, 2007, p. 7)

Tal como Newell afirma, comenta Clancey, o conhecimento pode ser representado, mas nunca está ‘ali à mão’. Cada declaração feita pelo observador captura o que ele precisa dizer em algum momento, e cada uma dessas declarações é mais tarde interpretada de forma diferente. Todo o comportamento humano é imediato, sem requerer planos intermédios ou outros esquemas semânticos que modelam o que estamos prestes a dizer ou fazer. Assim, quando um observador descreve um agente inteligente, necessita fazer uma distinção entre conhecimento enquanto uma capacidade atribuída a um agente (mudando dinamicamente através da interação com o ambiente) e as representações do observador desta capacidade (estruturas perceptíveis, abertas a reinterpretação). Aparece, aqui, a famosa distinção de Ryle entre *saber como* (a capacidade de desempenhar uma acção) e *saber que* (a representação). Isto coloca a importância no entendimento das diferenças entre os sistemas especialistas actuais e a capacidade humana, bem como na exploração das utilizações de computadores além da automatização do raciocínio.

Para Clancey, se o raciocínio humano não é produzido através de representações armazenadas, como podemos construir sistemas especialistas? Como podemos adquirir conhecimento? Em primeiro lugar, temos de adoptar uma postura diferente perante os artefactos que construímos: eles são modelos, e não seres inteligentes. Segundo, uma representação não é equivalente a conhecimento: a representação do que uma pessoa sabe é só um modelo para o seu conhecimento, uma representação de uma capacidade. Não podemos inventariar conhecimento, nem reduzi-lo a um conjunto de representações. E afirma: o significado de uma representação não pode ser tornado explícito, é algo aberto. Pode ser representado, mas não definido para sempre, nem totalmente capturado por representações. De igual forma, e ainda baseado na abordagem de Lave (1988), o contexto no qual um programa é usado não pode ser tornado explícito. O contexto pode ser representado, mas o mundo não pode ser descrito de forma objectiva e exhaustiva; as circunstâncias sociais ou culturais não podem ser reduzidas a um conjunto de factos e procedimentos. É precisamente a não especificação que confere ao comportamento humano o seu carácter robusto e sempre adaptável.

3.2.1.1 Compromisso ontológico

A representação do conhecimento surge como uma forma de conceptualizar sobre o mundo. Como afirma Barros (2007), a conceptualização é um processo complexo, situado entre a apreensão da realidade e a sua interpretação. Se voltarmos atrás e recordarmos as afirmações de Gruber, uma ontologia é uma especificação de uma conceptualização envolvendo um conjunto de compromissos ontológicos. Uma conceptualização seria uma visão abstracta e simplificada do mundo que queremos representar por algum motivo, e toda a base de conhecimento ou sistema baseado em conhecimento, ou agente de conhecimento, está *comprometido* com alguma conceptualização, explícita ou implicitamente.

O conceito de ‘compromisso ontológico’³¹ é importante, uma vez que é sobre ele que vai recair a selecção das conceptualizações que queremos representar formalmente. A definição clássica de compromisso ontológico pertence ao filósofo norte-americano Willard Van Orman Quine e data de 1961. De acordo com Quine, uma teoria lógica compromete-se ontologicamente com as entidades sobre as quais raciocina e quantifica: “to be is to be the value of a variable”.

Na sua acepção clássica, uma Ontologia pretendia dar conta do que existe de forma sistemática. Na inteligência artificial, o que ‘existe’ é o que pode ser representado numa linguagem formal. Para Gruber, a ideia de compromisso ontológico é baseada na perspectiva do Nível do Conhecimento de Newell. É o nível de descrição do conhecimento de um agente que é independente da representação ao nível do símbolo usada internamente pelo agente. O conhecimento é atribuído a agentes através da observação das suas acções; um agente ‘sabe’ algo se agir como se possuísse a informação e se agir racionalmente para atingir o seu objectivo.

Ainda segundo o autor, o conjunto de objectos ou entidades que pode ser representado num formalismo declarativo é designado por universo do discurso e este conjunto de objectos ou entidades e as relações que se estabelecem entre elas estão reflectidas no vocabulário

³¹ A expressão inglesa é *ontological commitment*

representativo com o qual um programa baseado em conhecimento representa o conhecimento. Os compromissos ontológicos são acordos para usar o vocabulário partilhado de uma forma coerente e consistente. Utilizamos por isso ontologias para descrever compromissos ontológicos para um conjunto de agentes de forma a poderem comunicar sobre um domínio do discurso sem necessariamente operar uma teoria partilhada globalmente: “In essence, committing to an ontology (e.g. supporting an interface using the ontology's vocabulary) requires that statements that are asserted on inputs and outputs be logically consistent with the definitions and constraints of the ontology.” (Gruber, 2008a) Um compromisso com uma ontologia é, à partida, uma garantia de consistência, mas não de amplitude no que diz respeito às perguntas e asserções usando o vocabulário definido na ontologia.

Para Gruber, uma ontologia resulta de um contrato, uma espécie de acordo social entre pessoas que partilham algo em comum. É um artefacto de representação, ou uma ‘especificação’ distinta do mundo que tenta modelar, construída com um objectivo específico. Os profissionais que trabalham na área da inteligência artificial devem, assim, conseguir discernir entre essa especificação e o mundo. Uma ontologia é objecto de uma arquitectura, de um desenho; é um produto que, segundo Gruber, não importa muito se é verdadeiro, mas sim se é útil, se consegue satisfazer os objectivos funcionais para que foi criada.

A questão da verdade relativamente ao mundo na área das ontologias é um assunto tratado exaustivamente por Barry Smith. Segundo o autor, “(...) the philosopher-ontologist, in principle, at least, has only one goal: to establish the truth about the domain in question.” (Smith, 2003, p. 22) No mundo dos sistemas de informação, pelo contrário, uma ontologia é um artefacto de software (ou linguagem formal) que foi desenhado tendo em mente o conjunto específico de utilizações e ambientes computacionais, sendo muitas vezes encomendada por um cliente específico ou para um programa específico num contexto específico. Um dos problemas da ontologia seria então a tendência para sacrificar a adequação ontológica à eficiência do software, centrando-se na tarefa de construção da hierarquia de conceitos e prestando muito menos atenção à questão de como os conceitos representados nessas hierarquias estão de facto instanciados no mundo real. Esta abordagem aponta para uma concepção de ontologia como disciplina preocupada não com a realidade em si mas com

substitutos bem formados (ou formalizados) da realidade. Deste modo, argumenta Smith, os engenheiros de ontologias consideram que podemos representar entidades no nosso sistema somente enquanto referidas através de um vocabulário normalizado. Não podemos representar no nosso sistema entidades como elas realmente são. Dito de outra forma: a ontologia não deve lidar com a realidade, mas somente com as nossas ‘conceptualizações’, centrando-se no conhecimento e nos conceitos que utilizamos quando falamos sobre o mundo. Só que existem muitas especificações de conceptualizações, e todas podem descrever o que intuitivamente reconhecemos como sendo a mesma ontologia. Dado que a noção de conceptualização abarca não só as boas como as más conceptualizações, somente certas ontologias podem ser verdadeiras em relação a um determinado domínio da realidade. Mas como estamos a lidar com modelos ou mundos substitutos criados com objectivos práticos, estes são tratados pelo engenheiro de ontologias como verdadeiros, servindo, em primeiro lugar, os propósitos do cliente – em última análise, é o cliente quem define o seu mundo de objectos substitutos. O objectivo não é a verdade em relação a um dado domínio da realidade – o que é, de facto, muito difícil de atingir – mas somente a verdade (pelo menos) em relação a uma dada conceptualização.

Claramente, e na perspectiva de Smith, é por razões práticas impossível de incluir todos os factos pertencentes aos objectos num dado domínio de aplicação numa base de dados. Tem de ser feita uma selecção e isto tem lugar de forma pragmática. Modelos de sistemas construídos na base da assumpção do mundo fechado são, obviamente, alvos mais fáceis de um ponto de vista matemático e de programação do que os seus homólogos no mundo real. Se tentarmos construir uma ontologia a partir de uma realidade complexa de objectos constantemente a mudar, então esta assumpção não pode ser mais válida; mas sem ela, o trabalho do programador é muito mais difícil. A assumpção do mundo fechado significa não somente que só existem aquelas entidades que estão representadas no sistema, mas também que tais entidades só possuem as propriedades que estão representadas no sistema. O que isto significa é que os objectos representados no sistema não são objectos reais, são “denatured surrogates”, possuindo somente um número limitado de propriedades, sendo totalmente indeterminados relativamente a todas as propriedades e dimensões com as quais o sistema não se preocupa. Como ironiza o autor, “(...) it is as if Hamlet, whose hair (we shall suppose) is not mentioned in Shakespeare’s play, would be not merely neither bald nor non-

bald, but would somehow have no properties at all as far as hair is concerned.” (Smith, 2003, p. 73)

Para Smith, as estruturas das teorias dos conjuntos substituem a realidade. O que se apresenta problemático, é que os ‘mundos substitutos’ criados por essas conceptualizações são tratados pelo engenheiro de ontologias como verdade, ou seja, o universo do discurso será especificado pelo cliente, e, para os objectivos do engenheiro de ontologias, o cliente tem sempre razão: “It is for this reason that the ontological engineer aims not for truth, but rather, merely, for adequacy to whatever is the pertinent application domain as defined by the client.” (Smith, 2003, p. 67) A atenção está centrada na reutilização do conhecimento do domínio de aplicação de tal forma a acelerar o desenvolvimento de sistemas de software similares em cada novo contexto de aplicação. À semelhança de Guarino, defende a utilização de princípios filosóficos como orientação para os investigadores de ontologias em inteligência artificial. Os melhores candidatos para boas conceptualizações estarão sempre próximos das ciências naturais. Neste sentido, regressamos a Quine, para quem o trabalho dos engenheiros de ontologias era identificado precisamente com a tarefa de estabelecer compromissos ontológicos com os cientistas, e somente com eles. Não existe nenhuma máquina que detecte a verdade. Devemos, sim, abordar a própria realidade, aproveitando a riqueza das descrições científicas das diferentes dimensões dessa realidade³².

Sumariando o que podemos entender por compromisso ontológico, Sowa refere que o critério de Quine é

“(…) a test for determining the implicit ontology that underlies any language, natural or artificial. The philosopher Willard van Orman Quine proposed a criterion that has become famous: ‘To be is to be the value of a quantified variable.’ That criterion makes no assumptions about what actually exists in the world. Its purpose is to determine the implicit assumptions made by the people who use some language to talk about the world. (...) Quine extended the criterion to languages of any form, including natural

³² Barry Smith está profundamente envolvido no campo da biomedicina e todo o seu trabalho de investigação na área das ontologias está relacionado com este domínio de especialidade.

languages, in which the quantifiers and variables are not stated as explicitly as they are in predicate calculus (...) Quine's criterion means that the implicit ontological categories are the concept types expressed by the basic content words in the language: nouns, verbs, adjectives, and adverbs." (Sowa, 2006)

Se atentarmos nas opiniões de investigadores mais ligados à linguística, como Ahmad e Gillam (2005), vemos que um universo do discurso é entendido como um conjunto de elementos de discurso que se relacionam com credos, convenções e conhecimento partilhado por membros de uma comunidade sociolinguística, misturando-se, muitas vezes, termos e conceitos derivados da filosofia ontológica, da terminologia e da análise do discurso, numa base interdisciplinar que trabalha em conjunto na análise desses compromissos ontológicos presentes no discurso dos especialistas. E observam: Quine, precursor da ontologia tal como hoje a entendemos dentro do contexto da inteligência artificial, colocou duas questões importantes: Quais são as condições que nos levam a falar cientificamente? Como é possível o discurso científico? A resposta encontrar-se-ia no compromisso ontológico: a partir da observação dos fenómenos, os especialistas articulam-nos com outras entidades linguísticas, partilhando raízes de referência e comprometendo-se com um mesmo conjunto ou sistema de conceitos.

Podemos verificar, da análise das opiniões dos vários autores nos pontos anteriores, que conhecimento e representação do conhecimento são dois conceitos distintos. O conhecimento para um sistema de informação consiste num sistema de deduções e regras de inferência que permitam a um agente atingir determinado objectivo para satisfazer determinada necessidade. O conhecimento é objecto de computação de acordo com o princípio da racionalidade. Seleccionamos o conhecimento relevante para uma tarefa associado a uma representação simbólica, analisada através da lógica. O conhecimento é uma representação formal sobre o mundo.

Na mesma medida, uma representação para um sistema de informação é um substituto que raciocina sobre o mundo, um conjunto de compromissos ontológicos. Como afirma Clancey, as representações não são conhecimento em si mesmo, mas sim um mapa para descrever um

território, que não pretende reproduzir o que se passa na mente das pessoas: elaboramos um modelo do mundo (mundo equivale aqui a domínio) e um processo de raciocínio sobre esse mundo através de um conjunto de inferências.

Ao contrário do que era postulado pelas primeiras definições de lógica, um sistema de informação não se preocupa com a verdade, mas sim com a funcionalidade, a utilidade, a interoperabilidade, e a partilha e reutilização das representações, interessando-se pela natureza do raciocínio e não pela natureza do mundo. A verdade equivaleria a uma proposição. Mais do que isso: segundo Smith, a verdade equivaleria ao conjunto de compromissos ontológicos relativamente a um universo do discurso definido pelo cliente.

Na nossa perspectiva, uma ontologia enquanto artefacto de engenharia não lida com a realidade, mas sim com conceptualizações formalmente explícitas através de linguagens de representação. A observação de Smith relativa a boas e más conceptualizações é uma falsa questão. Uma conceptualização é boa ou má dependendo do ponto de vista do observador e dependendo do objectivo da aplicação. Aceitar uma ontologia implica aceitar os compromissos ontológicos e as conceptualizações construídas com base nesses compromissos. Por isso a noção de verdade é antecipadamente assumida como correspondendo a determinadas escolhas. É verdade em relação àquele conjunto de compromissos ontológicos, que são encarados como acordos para utilização de vocabulário partilhado de forma consensual por uma comunidade para fins de manipulação, partilha e reutilização. Uma ontologia não tem a pretensão nem de representar todo o conhecimento do mundo, nem de substituir o mundo.

Os sistemas de informação podem apresentar resultados estatisticamente relevantes, mas pouco significativos em termos de satisfação das necessidades humanas. Se encararmos uma representação como uma aproximação, não será uma aproximação aos modelos de raciocínio humano? Se partirmos do princípio que, apesar de fundamentada na lógica com regras claras e precisas, uma representação tem por objectivo subjacente ser um substituto, ainda que imperfeito, é porque tenta sempre aproximar-se dessa perfeição. O conjunto de conceitos seleccionados como forma de pensar sobre o mundo constitui, por isso, um ponto importante numa ontologia. Para seleccionarmos esses conceitos necessitamos não só da terminologia, como da epistemologia. Tal como afirma Christophe Roche, a lógica é importante, mas *a*

posteriori. Temos de sustentar as nossas escolhas e decisões sobre princípios epistemológicos, numa tentativa *a priori* de nos aproximarmos do conhecimento. (Roche, 2003, p. 4)

O conhecimento não é um bem que se possa transaccionar, é uma abstracção, algo que nunca poderá, nas palavras de Musen, ser reduzido a discurso. Pensamos que esse discurso, seja na forma de língua natural, seja numa outra linguagem simbólica, encerra uma questão subjacente importante: a forte presença de um conhecimento que não se encontra expresso - o conhecimento tácito. Nesta matéria, a terminologia e a ontologia podem ser encaradas como veículos de transmissão privilegiados desse conhecimento. Recordando a citação de Cornejo (“(...) the members’ tacit knowledge is transmitted to other members, and (...) this know-how is occasionally codified and made explicit.”), observamos que ambas são complementares na tentativa de captura de conhecimento explícito e implícito. E existem duas formas de interpretar o termo ‘explícito’: podemos entender explícito como conhecimento verbalizado e verbalizável, e explícito como conhecimento codificado numa linguagem de representação. A distância entre um e outro é grande. Lembremo-nos que para um sistema de informação só existe o que pode ser representado numa linguagem formal.

Um terminólogo deve, por isso, ter bem presente o conceito de conhecimento para os sistemas de informação. O suporte texto encerra conhecimento sob forma de manifestações linguísticas, mas o discurso sobre o mundo não se pode confundir com o mundo. Não se trata de minimizar a importância do discurso em língua natural, trata-se antes de otimizar as potencialidades que esse discurso pode trazer para a codificação de representações do conhecimento.

3.3 Conhecimento tácito *versus* conhecimento codificado

Paralelamente à definição funcional, existe uma outra perspectiva de conhecimento ligada à parte social e económica que influencia, igualmente, o desenvolvimento de aplicações para sistemas de informação.

Para determinados observadores, a engenharia do conhecimento estaria, actualmente, a mudar de rumo. O seu objectivo não se encontraria tão centralizado na captura do conhecimento, mas sim em desempenhos sociais e criação de novas capacidades através de

ferramentas. Isto tudo com uma grande envolvimento cultural, onde os modelos computacionais passam, enquanto representações, a desempenhar um papel chave na reflexão de novas formas de pensar e agir do ser humano.

Nesse sentido, verifica-se uma reflexão sobre a questão do conhecimento em termos sociais. Não só nos deparamos com comunidades de agentes que têm de interagir e comunicar de formas diferentes em relação a tarefas diversas, como a modelação do conhecimento, ou a sua representação se quisermos, tornou-se num processo muito dispendioso. Segundo alguns analistas, quanto mais reutilizável e mais partilhável, mais valioso o conhecimento³³ se poderá tornar, adquirindo um valor próprio e transformando-se num valioso bem intangível³⁴. Será assim ou precisamente o contrário? Quanto menos codificado e menos 'democratizado', mais precioso pode ser o conhecimento?

Num mundo onde os sociólogos apelidam a nossa sociedade de 'sociedade do conhecimento' e onde os 'Chief Knowledge Officers' ocupam lugares de topo nas universidades, empresas e agências governamentais, é um choque saber, diz Steve Fuller (2001), que a procura do conhecimento está a tornar-se numa espécie em extinção. A própria ideia de que o conhecimento é algo que necessita ser 'gerido', sugere que o seu crescimento não deve ser deixado ao acaso. A ideia de 'gerir conhecimento' vai contra os últimos 2500 anos de pensamento ocidental, que valorizava a procura de conhecimento *per se*, sem contabilizar custos e benefícios. O crescimento da gestão do conhecimento pode ser visto como um abrandamento no profissionalismo da procura de conhecimento. Este abrandamento é registado em dois níveis: a mentalidade da gestão do conhecimento é entendida como falta de génio científico do nosso tempo, e, num nível potencial mais profundo, diz respeito ao tipo de conhecimento que vale a pena procurar. Os gestores de conhecimento estão mais interessados em explorar o conhecimento existente de forma eficiente para capturar uma grande parte dos mercados nos quais competem. O seu interesse em produzir e distribuir novo conhecimento é válido somente até ao ponto onde os ajuda neste objectivo. As empresas ou

³³ Temos também presente, neste caso, a conceito de 'conhecimento' enquanto formalização.

³⁴ O conceito de *goodwill* contabilístico já enquadra, de alguma forma, esta percepção do conhecimento.

alugam formas de conhecimento que outras produzem mais barato, ou possuem formas de conhecimento que outros não podem produzir mais barato. A ideia do conhecimento enquanto propriedade intelectual foi finalmente levada a sério. Isto é uma proposta revolucionária, que tem sido ignorada pelos produtores de conhecimento especializado.

Na mesma linha de raciocínio, Robin Cowan e Paul David (1999) analisam a questão do conhecimento sob o ponto de vista da economia. Michael Polanyi (1958, 1967) chamou a atenção para a existência de uma 'dimensão tácita do conhecimento', uma forma ou componente do conhecimento humano distinta do conhecimento explícito, mas complementar, nos processos cognitivos conscientes. Polanyi ilustrou esta conceptualização referindo-se à percepção comum: estamos muitas vezes atentos a certos objectos sem estar a centrar a nossa atenção neles. Isto não os torna menos importantes, pois formam o contexto que torna a percepção possível, entendível e frutífera. Consequentemente, o termo conhecimento tácito³⁵ tornou-se largamente aplicado a formas de conhecimento pessoal que não é imediatamente transmissível enquanto 'informação', que por sua vez é entendida como um bem ideal que tem características económicas peculiares que a diferenciam de outras comodidades mais convencionais em termos económicos.

Os autores comentam que foi então colocado um contraste entre o 'modelo de algoritmo' da produção de conhecimento que envolve a gestação e transmissão de declarações explícitas e proposicionais, e o chamado 'modelo de aculturação' da actividade científica. Esta distinção era primeiramente evocada entre filósofos e sociólogos para desafiar a noção que a ciência, no método científico moderno, era uma fonte de declarações 'privilegiadas'. O privilégio em questão implicava que declarações científicas podiam ser limpas do contexto social no qual se tinham formado e que no qual tinham adquirido sentido, e promulgadas como pertencendo a um corpo 'codificado' de conhecimento sobre o mundo físico. Os opositores desta visão

³⁵“The term 'background knowledge' is used loosely across a range of academic disciplines without receiving a precise definition. This is significant because knowledge can only be 'background' relative to some 'foreground'. 'Background knowledge' is not to be confused with the term 'tacit knowledge' which is widely used in knowledge management, and which refers to knowledge which is 'semiconscious and unconscious knowledge held in peoples' heads and bodies.” [Leonard, D.&Sensiper. S. The role of tacit knowledge in group innovation. 1998] (Brewster, Ciravegna, & Wilks, 2003, p. 5)

apoiavam-se fortemente na importância do conhecimento tácito na conduta actual das actividades científicas. O conhecimento tácito é evocado hoje por aqueles que desafiam a abordagem ‘algorítmica’ da análise económica moderna em todos os aspectos do comportamento humano. O desafio coloca-se em relação a toda a construção de processos de decisão racionais enquanto base para a modelação e explicação do comportamento dos agentes individuais. Se aspectos cruciais do conhecimento humano não podem ser explicitamente formulados e articulados do ponto de vista linguístico, como podem servir de base a cálculos consistentes?

No novo entendimento da ciência e da tecnologia como sendo metas envolvidas intrinsecamente com o conhecimento tácito, diz-se que o entendimento essencial é a porção de conhecimento que permanece não codificado e por isso sem as propriedades de bem público que resultaria em falhas de mercado e perda de informação: “Whether a particular bit of knowledge is in *principle* articulable or necessarily tacit is not the relevant question in most behavioral situations. Rather, the question is whether the costs are sufficiently high so that the knowledge *in fact* remains tacit” (Cowan et al., 1999, p. 6). O chamado ‘sticky data’ não viaja livremente, reside na mente dos cientistas e engenheiros. Curiosamente, o conhecimento tácito começa a suplantar o ‘conhecimento codificado’, como o centro de um novo *rationale* de pesquisa dos diferentes governos que pretendem construir uma ‘competitividade’ nacional e regional através da inovação. Esta situação implica que os benefícios económicos só estão disponíveis para serem capturados localmente, sendo necessário ter actividades de pesquisa muito próximas da produção. Assim, uma noção epistemológica foi transmutada para um fenómeno de vantagem corporativa e mesmo nacional. O conceito de conhecimento codificado é entendido aqui no sentido económico e social, e não no sentido da formalização da inteligência artificial³⁶. Codificado aponta para normas ou códigos, um conhecimento

³⁶ Segundo Gualtieri e Ruffolo, existem muitos tipos de organização de conhecimento dentro e fora das empresas, distribuídos por fontes humanas e máquinas. A distinção clássica, e geralmente aceite, identifica: i) um conhecimento tácito e implícito que resulta dos processos de aprendizagem pessoal, presente dentro de cada organização em termos do conhecimento pessoal dos seus membros, e ii) um conhecimento explícito, geralmente partilhado e acessível publicamente dentro da organização através de armazenamento e processamento formais. Este conhecimento pode ser estruturado, estando disponível em bases de dados, semi-estruturado, acessível na intranet e sítios Web, ou não-estruturado, disponível em documentos de texto, procedimentos, templates, etc. (Gualtieri & Ruffolo, 2005)

registado que serve de repositório ou autoridade. Ler o código correctamente pode envolver uma aquisição prévia de conhecimento especializado (possivelmente incluindo conhecimento que não está registado sob a forma de escrita). Isto significa que o que é codificado para uma pessoa ou grupo pode ser tácito para outro, e um mistério para um terceiro, entrando aqui a importância do contexto.

Por outro lado, comentam Cowan e David, surge também o conceito de ‘livro de código’, ao qual adicionamos conhecimento. No início, não há livro de código. Assim, a actividade de codificação inicial envolve criar um ‘dicionário de especialidade’. Depois de desenvolvidos modelos e uma linguagem, podemos escrever os documentos, sendo que novos documentos vão inevitavelmente introduzir novos conceitos e nova terminologia. Muita da moderna teoria endógena baseia-se na noção de que existe um ‘stock’ mundial de conhecimento. Está implícito nesta literatura que este ‘stock’ é codificado, uma vez que parte ou todo é assumido ser livremente acedido por todos os agentes económicos no sistema que está sob análise. Mas se levarmos a sério o aspecto contextual e de codificação do conhecimento, o ‘stock’ mundial de conhecimento codificado podia ser melhor definido como a intersecção de conjuntos individuais de conhecimento codificado – a porção do que é ‘partilhado’ no sentido de ser tanto conhecido como acessível: “It is within the realm of the codifiable or articulable-yet-uncodified, that price considerations come into play in an interesting way. Within this category reside decisions about the activity of codification, its costs and benefits”. (Cowan et al., 1999, p. 13)

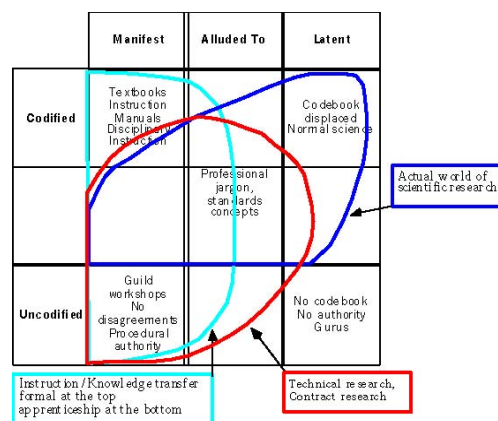
Podemos, por isso, distinguir entre conhecimento articulado (codificado) e inarticulado (não é referido explicitamente no decurso típico das actividades de conhecimento). No primeiro caso, existe um livro de código, uma vez que está implícito no facto do conhecimento ter sido codificado. No segundo, podemos considerar, por um lado, conhecimento tácito no sentido normal – não foi registado em palavras ou artefactos, não existe livro de código -, e, por outro, conhecimento registado, existindo um livro de código, embora esse livro não possa ser referido pelos membros do grupo ou não sabemos a sua localização. Surge aqui a noção de ‘displaced codebook’. Esta noção implica que um corpo codificado de conhecimento comum

está presente, mas não de forma manifesta. Os termos técnicos figuram no discurso descritivo mas permanecem indefinidos porque o seu significado é evidente para todos os interessados; as relações fundamentais entre variáveis também não são reiteradas na conversação ou nas mensagens trocadas entre os membros do grupo ou comunidade epistemológica. Identificar a zona onde o conhecimento é codificado, mas não se manifesta a existência da codificação, é extremamente importante³⁷. As línguas são dinâmicas. Consequentemente, o progresso da investigação envolve e requer a estabilização de significados. Existem custos fixos muito grandes ligados ao processo de codificação, especialmente quando o ambiente cognitivo está em mudança – situação prototípica dos discursos de especialidade. Os custos de aprendizagem e desenvolvimento de linguagens nas quais estão escritos os novos códigos decorrerão durante o período em que o ambiente de conhecimento flui, e os benefícios sentir-se-ão durante um subsequente período de estabilização e disseminação da informação.

Para Paul Duguid (2005), uma forma de tornar o conhecimento economicamente viável foi reduzi-lo a informação. De facto, uma cadeia de argumentos epistemológicos que vem desde Sócrates sugere que o conhecimento codificado, a dimensão explícita, repousa num substrato não codificável que nos diz como usar o código. O conhecimento explícito não é uma base auto-suficiente, mas uma super-estrutura dependente.

E Duguid cita Simon [1999] relativamente à desmistificação económica do conhecimento:

³⁷Os autores explicam em pormenor o gráfico que representa um novo mapeamento e topografia do conhecimento, clarificando que estão preocupados não com a taxonomia e em delinear a natureza e variedade do conhecimento humano, mas sim explicar e prever as mudanças economicamente significativas que terão lugar no carácter das actividades de conhecimento. (Cowan et al., 1999)



“All the aspects of knowledge - its creation, its storage, its retrieval, its treatment as property, its role in the functioning of societies and organization--can be (and have been) analyzed with the tools of economics. Knowledge has a price and a cost of production; there are markets for knowledge, with their supply and demand curves, and marginal rates of substitution between one form of knowledge and another. (Simon, 1999, quoted in Ancori, Bureth, & Cohendet, 2000, p 256n)” (Duguid, 2005, p. 3)

Brewster (2003), contrariando a visão menos otimista de Duguid, sublinha que o crescimento na área da gestão do conhecimento deve-se ao reconhecimento de que os activos comerciais estão no conhecimento tácito dos colaboradores e não no equipamento, tradicionalmente visto como o meio de valorizar os activos das empresas. Com a presença avassaladora da Web, torna-se crucial transformar informação excessiva em conhecimento focalizado, encarado aqui no sentido utilizado em inteligência artificial de ‘conhecimento igual a informação útil’.

Também O’Hara (2004) coloca no mesmo patamar informação e conhecimento: os recursos da Web são tecnologias para transferência de informação e conhecimento de forma eficiente, e por isso muitas das questões relacionadas com esta área têm uma base epistemológica. Por outro lado, existe uma interacção entre o epistemológico, o social e o político. Neste sentido, a forma como tratamos o conhecimento pode ter um impacto dramático nas tecnologias que produzimos. Existem muitas vantagens em codificar e ‘modelar’ o conhecimento de acordo com as nossas necessidades organizacionais. Desde já, não ficamos dependentes desse temível conhecimento tácito existente na cabeça dos colaboradores. Além disso, o conhecimento codificado tornou-se um activo. Porém, quando consideramos as consequências sociais, económicas e organizacionais da codificação, aparecem algumas desvantagens, nomeadamente tempo e dinheiro, além da colaboração de especialistas e de ferramentas informáticas.

Por outro lado, refere O’Hara, enquanto o conhecimento tácito tende a evoluir naturalmente, o conhecimento que está explicitamente escrito torna-se cristalizado. Tem de ser mantido, actualizado, podendo estar envolto em confidencialidade. Todo este processo torna-se substancialmente dispendioso, sobretudo quanto se trata de um domínio muito dinâmico. O conhecimento tácito muitas vezes inclui muito sobre a sua própria aplicação. Com o conhecimento explícito, não só temos de recuperar a informação, como o utilizador necessita

saber que informação é importante, e onde pode ser encontrada. Além disso, existem vários tipos de conhecimento, uns mais 'escorregadios' do que outros, que requerem tipos de processamento diferentes e dificuldades acrescidas de reutilização. Partindo de um dos pressupostos da nossa era de que 'conhecimento é poder', várias questões se colocam: Que conhecimento devemos representar e como devemos representá-lo? Quais são os nossos limites para essa tarefa? De quem é o conhecimento que extraímos? Estaremos a distribuir o poder de forma legítima? Como posicionar e equacionar os custos da transferência de conhecimento? Como proceder relativamente à manutenção desse conhecimento? O'Hara lembra as palavras de Steve Fuller³⁸ que visualiza o processo ideal de transferência de conhecimento como um avião que não abana, onde não se perde conteúdo através do processo de transferência, e a Web como o seu meio supercondutor. Ou seja, a manipulação do conhecimento é vista como um processo material, utilizando recursos. Mas não podemos evitar a turbulência. Na codificação do conhecimento, o que esperamos são traduções da forma para a forma, do modo representativo para o modo representativo. Cada tradução da forma para a forma pode perder conteúdo, ou envolver a geração de novo conhecimento. Acontece que este conhecimento, codificado de uma certa forma, afectará as suas propriedades. Existe, por isso, uma lacuna evidente na codificação do conhecimento tácito, conhecimento que as pessoas possuem e não está escrito.

Vemos nas várias abordagens que a dimensão cultural e social adquire um valor importante na tentativa de codificação do conhecimento tácito, encarado como um activo fundamental nas organizações e nas comunidades técnicas e científicas. As ferramentas computacionais, através das linguagens formais de representação desse conhecimento, influenciam comportamentos e formas de pensamento. Seleccionamos o tipo de conhecimento em que vale a pena investir. No entanto, a prevalência recente do 'modelo de aculturação' em relação ao 'modelo algorítmico' de que nos falam Cowan e David aponta para a dificuldade de 'limpar' o discurso científico do seu contexto, transformando-o num corpo codificado de conhecimento. Podemos

³⁸ Steve Fuller (2002) *Knowledge Management Foundations*, Butterworth-Heinemann, Woburn, MA.

reduzir certo tipo de conhecimento a fórmulas matemáticas, mas a porção de conhecimento não codificado continua a ser o ponto crucial, dificultado pelo facto de que o que é codificado para uns, pode não ser para outros.

A terminologia, independentemente da corrente teórica que a acompanha, terá sempre, a nosso ver, uma função harmonizadora. Do mesmo modo, uma ontologia constitui-se como um retrato conceptual de um domínio num dado momento, tendo como um dos objectivos principais normalizar e estabilizar os conceitos através da representação formal de conhecimento estático. Retirar totalmente a vertente cultural do termo por certas metodologias de construção de ontologias poderá tornar-se contraproducente, na medida em que se pode distanciar cada vez mais desse conhecimento tácito. Não podemos esperar extrair ontologias de discurso, seja ele de especialidade ou não, mas também não devemos esvaziar o aspecto social do conhecimento, especialmente dentro do conceito de 'comunidade de partilha'. Não é por acaso que alguns autores da área da inteligência artificial designam as ontologias como 'artefactos semânticos'.

4. Considerações teóricas sobre a terminologia

A terminologia constitui-se como um corpo teórico e um conjunto de práticas. Os especialistas recorrem à terminologia para se expressarem, os organismos internacionais de normalização apoiam-se na terminologia para sistematizar procedimentos, os países estabelecem planeamentos linguísticos utilizando terminologia, a sociedade no seu âmbito mais alargado difunde conhecimento através da terminologia. Não há conhecimento sem terminologia e não há terminologia sem conhecimento.

A terminologia atraiu o interesse de muitos investigadores de áreas e motivações diferentes: lexicógrafos, tradutores, engenheiros e cientistas viram a necessidade de melhorar a comunicação ou ter acesso a informação específica, tendo sido desenvolvidos inúmeros recursos terminológicos para satisfazer estas necessidades.

Dividimos as reflexões sobre as orientações teóricas e metodológicas da terminologia em duas grandes áreas: as que partem de uma perspectiva linguística, e as que optam por uma perspectiva conceptual. Esta última tem sido recuperada na última década pela investigação na área da inteligência artificial e pelo desenvolvimento das ontologias.

A evolução tecnológica permitiu a análise de grandes quantidades de texto em formato electrónico, mas potenciou ao mesmo tempo uma certa independência relativamente às fontes textuais tradicionais de pesquisa. Essa progressiva diversidade de metodologias assentes em processos de inferência cada vez mais automatizados vem colocar, por um lado, a questão da importância do texto de especialidade, e, por outro, um reposicionar do foco central dos estudos terminológicos: a Terminologia estuda os termos e as relações entre termos, estuda os conceitos e as relações entre os conceitos, ou estuda as relações entre termos e conceitos?

No centro deste debate, encontram-se duas grandes perspectivas: a análise semasiológica (do termo para o conceito) e a análise onomasiológica (do conceito para o termo). Abordaremos, em seguida, as suas principais linhas de orientação.

4.1 Terminologia

4.1.1 Perspectiva semasiológica

Na perspectiva semasiológica, o objecto da terminologia encontra-se essencialmente centrado no termo. Um termo pode ter várias definições. De acordo com Conceição (2005), podemos definir termo enquanto unidade de referência, unidade de denominação, unidade de representação e unidade de conhecimento. O referente é o objecto do mundo que a forma linguística denomina. Apreendemos o mundo de determinada maneira, designando-o. A designação é algo convencional, estabelecida socialmente. Os termos existem nos produtos de comunicação que são os textos. Enquanto unidade linguística, o termo é a denominação de um conceito, formulando verbalmente a percepção do mundo. Por reflectir o processo de afectação das palavras aos referentes que por sua vez reflectem a interacção entre o sujeito e o mundo, a denominação é encarada como uma reconceptualização, sendo que a relação entre a denominação e o conceito não é simétrica, uma vez que o conceito não é inteiramente denominado. A organização conceptual constrói-se através da língua, partindo-se do discurso para comunicar. Enquanto unidade de conhecimento, o termo “(...) sert de véhicule de transmission et de compréhension de ces connaissances car, en tant qu’unité linguistique, il est (tout comme la langue) fondamentalement pluriel.”(Conceição, 2005, p. 67)

Nesta perspectiva, e segundo Conceição, os termos são unidades lexicais especializadas que funcionam num sistema linguístico próprio de cada utilizador e de cada comunidade, sendo que o estudo dos termos deve ter lugar dentro das diversas situações comunicativas. Especificamente, os termos enquanto unidades de conhecimento devem ser analisados no seu ambiente discursivo:

“Les changements qui se sont produits dans les approches théoriques de la terminologie nous permettent aujourd’hui d’envisager les termes en tant que représentations de connaissances et nous conduisent à les analyser au sein de leur environnement discursive. Ils sont des manifestations de conceptualisations qui ne peuvent être décrites par le terminologue que par le travail sur les discours et leur fonctionnement. L’analyse terminologique ne s’occupe pas seulement, dans ce cas, de la recherche de la

dénomination du concept, mais aussi de l'analyse des occurrences discursifs des dénominations que aide à délimiter les concepts dénommés et comprendre les savoirs exprimés. (...) Les termes son envisagés en tant qu'unités de connaissance, accessibles à partir de l'existence discursive." (Conceição, 2005, pp. 69-70)

Esta abordagem está de acordo com a socioterminologia e com as teorias comunicativas e sociocognitivas da terminologia, defendendo que a pesquisa terminológica deve ser feita a partir de textos e de corpora textuais. Contudo, o autor alerta para o facto de que a pesquisa feita somente a partir de corpora não nega a pré-existência dos conceitos. A terminologia estuda a verbalização do conhecimento. O estudo dos termos e da sua verbalização contribui para o conhecimento do sistema linguístico em duas vertentes: a análise linguística (semântica e morfossintáctica) e a análise cognitiva (através da construção da significação no discurso). Recordando Saussure, não podemos separar a face linguística de um termo da sua face cognitiva.

Para a perspectiva semasiológica, um termo é uma unidade lexical utilizada dentro de um domínio de especialidade. A ligação estabelecida entre a unidade lexical e o domínio é central para a terminografia. Os termos podem ser simples ou complexos e pertencer a diferentes categorias gramaticais (nome, verbo, adjectivo ou advérbio). Embora sejam apreendidos em função de algo exterior à língua (o domínio), os termos materializam-se nos textos. É aí que o terminógrafo encontra os elementos que permitem proceder à sua identificação e subsequente descrição.

A abordagem semasiológica surge numa reacção contra os princípios terminológicos preconizados pela Teoria Geral da Terminologia, que, segundo Pearson, "(...) tends to study terms in isolation from text and to ignore context, even when the terms have originally been sourced in text". (Pearson, 1998, p. 2) A terminologia moderna, segundo a autora, estuda os termos em contexto, sendo que uma palavra pode adquirir o estatuto de termo em função do seu papel no texto:

"All terminologists, whether 'traditional' or 'modern', are concerned essentially with establishing knowledge structures for subject domains and they build these knowledge structures by comparing and contrasting related concepts, by examining vertical and

horizontal links between concepts. The ordering and classification of knowledge are crucial to terminological studies. In 'modern' terminology practice, the emphasis is much more on usage with the use of 'real text as primary source of data' ". (Pearson, 1998, p. 2)

Segundo Cabré, para uma concepção linguística da terminologia os objectos de conhecimento são as unidades terminológicas concebidas como combinações indivisíveis em termos de forma e conteúdo. Assim, o conteúdo está associado à forma e por isso as unidades não só designam mas também significam todas as consequências cognitivas do seu significado. Estas unidades, definidas como combinações de condições linguísticas, aparecem no discurso especializado:

"(...) terms, in their widest sense, are the units which most efficiently manipulate the knowledge of a particular subject. (...) We begin with the assumption that specialised discourse presents an organized structure of knowledge. This structure could be represented as a conceptual map formed by nodes of knowledge, which can be represented by different types of units of expression, and by relations between these nodes." (Cabré, 2003a, p. 20)

Nas palavras de Bourigault et al, "(...) these terminological units can be further analysed and organized into sophisticated networks that reflect the knowledge structure of a specialized field." (Bourigault, Jacquemin, & (eds.), 2001, p. viii)

Esta abordagem que parte do termo para o conceito constitui, do nosso ponto de vista, uma prática que não deve ser posta em causa na terminologia, tendo apresentado nos últimos anos resultados que têm sido objecto de um contínuo aperfeiçoamento, nomeadamente através das ferramentas de processamento de língua natural. Escrevemos para transmitir o nosso conhecimento. Manifestamos pensamentos em discurso.

Entendemos, contudo, que algumas das afirmações acima transcritas, nomeadamente as críticas tecidas à Teoria Geral da Terminologia, enfermam da mesma rigidez de que a acusam. Um processo comunicativo não passa somente por manifestações linguísticas. O texto constitui-se como um veículo privilegiado dessas manifestações que podem transmitir conhecimento, nomeadamente através dos termos e das relações entre termos que aí têm

lugar. No entanto, um termo pode não ser uma unidade lexical. Por outro lado, um texto de especialidade pode não representar uma estrutura organizada de conhecimento. Se assim fosse, seria relativamente fácil construir ontologias a partir de corpora textuais. Como já vimos, o papel do que se encontra ausente do texto, o conhecimento tácito e implícito, é substancial. (Costa, 2006; Costa & Silva, 2009)

Actualmente, o papel da terminologia na construção de ontologias não passa exclusivamente pelo texto. E essa tendência encontra-se, curiosamente, nos princípios metodológicos estabelecidos por Eugen Wüster.

4.1.2 Perspectiva onomasiológica

A perspectiva onomasiológica defende a primazia do conceito, de acordo com a teoria proposta por Wüster³⁹ e adoptada pela Escola de Viena, conhecida como Teoria Geral da Terminologia. Wüster tinha como objectivo eliminar o mais possível a ambiguidade através da normalização da terminologia, institucionalizando-a enquanto disciplina científica. Na tentativa de fundamentar a necessidade da terminologia normalizada, propôs elaborar princípios internacionais para o trabalho terminológico.

A terminologia enquanto disciplina autónoma surgiu entre 1930 e 1960, centrando-se na formulação de métodos terminológicos que se inclinavam para o carácter sistemático dos

³⁹ Uma das tendências mais importantes no estudo da terminologia tem as suas raízes no trabalho de Wüster (1898-1977), que consolidou a terminologia como uma disciplina independente em meados do Séc. XX. Engenheiro de profissão, com um grande interesse pela informática, um fiel do Esperanto e um grande defensor de uma comunicação profissional não ambígua, desenvolveu a teoria da terminologia na base da sua experiência terminográfica advinda da compilação de *The Machine Tool. An interlingual Dictionary of Basic Concepts* (Wüster 1968), um dicionário coligido de forma sistemática de termos normalizados feito como modelo para futuros dicionários técnicos. Este projecto foi apoiado pela Organisation for Economic Cooperations and Development (OECD) das Nações Unidas e publicado em 1968. Os princípios de Wüster foram seguidos postumamente por H. Felber, tendo como base as aulas de Wüster (1972-74) e publicados em *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie* em 1979. Wüster foi um grande impulsionador da criação do Comité Técnico 37 da ISO em 1951.

Esta matéria encontra-se amplamente debatida na literatura. Consideramos, no entanto, que, neste contexto, justifica-se um regresso às origens e uma leitura transversal, embora necessariamente breve, dos princípios e orientações metodológicas desta escola terminológica.

termos. As preocupações teóricas sobre a natureza dos termos só surgiram mais tarde. Wüster reconheceu o trabalho de Saussure, o primeiro a chamar a atenção para a sistematização das línguas, e o Círculo de Viena foi durante muitos anos a única escola a desenvolver um corpo de princípios e premissas que permitiram a compilação de dados terminológicos. A corrente positivista surgida entre as duas guerras mundiais tentava promover, nomeadamente através do Círculo de Viena, uma visão unificadora do conhecimento: o mundo é dividido em domínios estáveis, cada um equivalente a uma rede fixa de conceitos, sendo os termos os representantes linguísticos desses conceitos. (Cabré, 2000) (Bourigault et al., 2002)

Segundo Pearson (1998), a terminologia começou gradualmente a emergir enquanto disciplina linguística autónoma quando pessoas como Wüster acharam que os termos deviam ser tratados de forma diferente das palavras da língua geral em vários aspectos. Em comparação com a lexicologia onde a unidade lexical é normalmente o ponto de partida, a terminologia partiria do conceito: “(...) jede Terminologearbeit geht von den Begriffen aus (...) ” (Wüster, 1979). O conceito devia ser considerado à parte do termo. Os conceitos existem independentemente dos termos e independentemente de uma língua em particular. Os terminólogos estariam preocupados com a normalização do uso da língua, numa tentativa de desenvolver a *Soll-Norm*, ao lidar com a língua tal como deve ser usada, fixando e normalizando o significado para evitar conflitos de comunicação. Isto levou à criação de vocabulários normalizados com o objectivo de se tornarem *Ist-Norm* na comunicação técnica. Estes termos normalizados seriam utilizados para representar estruturas conceptuais que estariam na base de domínios de especialidade.

Ainda segundo Pearson, na teoria clássica da terminologia os domínios de especialidade compreendiam uma série de conceitos ou construções mentais representadas por termos, produtos de um processo mental onde objectos e fenómenos no mundo real são entendidos. Uma vez ocorrido este processo, os objectos e fenómenos ganham uma existência abstracta no nosso pensamento que, por sua vez, depende da manipulação dessas abstracções, entendidas como conjuntos de propriedades que fazem parte dos objectos, fenómenos, eventos, ou classes de objectos. Estas abstracções são conceitos. Os conceitos são entendidos como ‘puros’, usados somente por comunidades fechadas que acordaram numa série de

princípios de entendimento. A relação entre conceitos é representada por relações lógicas e ontológicas usadas para construir sistemas hierarquizados de conceitos.

Se quisermos referenciar os princípios atribuídos à Escola de Viena, podemos, por exemplo, socorrer-nos da enumeração elaborada por Temmerman (2000): i) A terminologia estuda os conceitos e depois os termos (onomasiologia). ii) Os conceitos têm contornos precisos e um lugar atribuído num sistema conceptual. iii) Os conceitos devem ser definidos de maneira tradicional. iv) O conceito e o termo têm uma relação biunívoca. v) Os termos e os conceitos estudam-se em sincronia. A relação entre termos e conceitos é acordada e normalizada.

Aprofundando um pouco mais o exposto nos parágrafos anteriores, L'Homme (2004) define do seguinte modo a abordagem onomasiológica ou conceptual e a abordagem semasiológica ou léxico-semântica⁴⁰: A abordagem conceptual ou onomasiológica consiste na teoria de que o conjunto dos termos de um domínio especializado reflecte a organização do conhecimento nesse domínio, ou seja, os termos denotam os conceitos. O conceito é uma representação mental que retém as características comuns a um conjunto de objectos. Esta representação mental é dada como se apresenta (não procuramos explicar a natureza), e precede a forma linguística, sendo que os conceitos são apreendidos por uma série de operações de classificação. Os objectos do mundo real são reunidos debaixo de uma mesma classe se partilharem características comuns⁴¹. A terminografia que partilha de uma óptica conceptual utiliza como ponto de partida o conceito e olha para o termo como a sua realização linguística. Na abordagem onomasiológica identifica-se o conceito e procuram-se as formas que o designam, sendo este descrito através de uma definição. Uma vez fixado, esse conceito engloba todas as formas linguísticas utilizadas para o nomear. As formas são reunidas e tratadas como se fossem sinónimos perfeitos.

⁴⁰Cabré classifica as duas abordagens de semiótica e linguística. Na abordagem semiótica o conceito é independente e precede a designação; além disso, a designação não é necessariamente realizada linguisticamente. Por contraste, para uma concepção linguística, o objecto de conhecimento são as unidades terminológicas concebidas como combinações indivisíveis em termos de forma e conteúdo. Assim, o conteúdo está associado à forma. (Cabré, 2003a)

⁴¹ Segue de perto os princípios aristotélicos.

Para L'Homme, a terminologia clássica pretenderia alcançar a normalização, a regulação dos termos para facilitar a comunicação. Ao normalizar, tenta eliminar as ambiguidades intervindo em fenómenos normais em língua natural como a sinonímia e a polissemia. Privilegia por isso a univocidade: uma forma corresponde a um só conceito e um conceito é expresso por uma só forma. A abordagem onomasiológica atribui grande importância às relações que os conceitos têm entre si, descritas em função de parâmetros extralinguísticos, ou seja, em função do conhecimento que os especialistas têm nesse momento sobre esse domínio. Assim, uma representação conceptual pode ser modificada para se tornar consensual ou para integrar um conceito inédito. Além disso, o mesmo conjunto de conceitos pode dar origem a mais do que um mapa conceptual em função do ponto de vista.

L'Homme nota, contudo, que a abordagem onomasiológica é adequada para descrever a óptica do especialista que vai nomear uma realidade nova, mas não reflecte realmente o trabalho do terminógrafo. Mesmo que consideremos verdadeira a delimitação precisa de um conceito enquanto guia na selecção de termos, o terminógrafo procede geralmente a uma escolha dos termos no texto. Uma vez identificados os termos, apreende o seu sentido, adoptando uma atitude inversa daquela preconizada pela terminologia clássica. Na abordagem semasiológica ou léxico-semântica definimos o sentido de uma unidade lexical ao observar o conjunto das suas interacções com outras unidades linguísticas. A observação das oposições entre essas unidades, as similitudes e a análise das suas combinatórias permite circunscrever gradualmente o seu sentido. Os sentidos são extraídos e distinguidos a partir do conhecimento que o terminógrafo tem dos termos. As distinções são confirmadas pela observação das interacções dos termos dentro dos contextos.

Analisando detalhadamente os pressupostos da escola de Viena e do chamado positivismo lógico, Cabré comenta que a teoria formulada por Wüster enfatizou a separação entre terminologia e linguística, e, mais concretamente, entre terminologia e lexicologia. Para justificar esta separação, Wüster partiu de pressupostos relativamente a dois aspectos: ao estado da língua e à evolução da língua. No primeiro grupo, estabeleceu o carácter onomasiológico da terminologia, a preexistência do conceito, a autonomia do conceito com relação à denominação semiótica, bem como o interesse exclusivo pelo léxico numa perspectiva sincrónica, em detrimento dos demais níveis linguísticos. No que diz respeito à

evolução da língua, assumiu que a terminologia devia ser objecto prioritário de prescrição e que na intervenção sobre os termos deviam prevalecer o carácter internacional e a forma escrita, sendo que as línguas seriam objecto de planificação, unificação e normalização. Seguidor das ideias do círculo de Viena e firme defensor da filosofia idealista, considerava que a língua geral não permite referir-se aos objectos de maneira unívoca nem eficiente, e por isso um registo comunicativo com um profissional devia utilizar sistemas expressivos que não dessem lugar a ambiguidades e que favorecessem a concisão – ou seja, deveria ser uma comunicação monorreferencial e monossémica. (Cabré, Feliu, & Tebé, 2001) (Cabré, 2003)

Poderíamos resumir as principais críticas tecidas por diversos autores em relação a Wüster nos seguintes pontos:

- Abordagem onomasiológica através da pré-concepção dos conceitos e não dos termos, criando mundos conceptuais fechados, precisos e estáticos a partir do conhecimento dos especialistas, onde prevalecem, portanto, os parâmetros extralinguísticos. Os conceitos existem independentemente dos termos e de uma língua particular. O conjunto de termos de um domínio – representações linguísticas dos conceitos – reflecte o conhecimento desse domínio, sendo que a terminologia estaria centrada na classificação dos conceitos na mente das pessoas. Os conceitos eram considerados como *clusters* com características internacionalmente unificadas expressas através de signos equivalentes de diferentes sistemas linguísticos e não linguísticos. A terminologia consistia simplesmente na recolha de conceitos com vista à normalização dos termos. Esta relação estável e biunívoca entre termo e conceito limitaria os problemas da variação linguística. (Desmet, 1994)(Cabré, 2000, 2003a; Maia, 2003; Pearson, 1998; Temmerman, 2000) (Kageura, Daille, Nakagawa, & Chien, 2004; L'Homme, 2004) (Smith, Ceusters, & Temmerman, 2005) (Cabré, 2006)
- Instauração de uma cisão entre terminologia e linguística, dando origem a uma imagem simplista e obsoleta, impedindo a compreensão da riqueza e complexidade do funcionamento efectivo dos textos de especialidade. As ocorrências reais dos termos nos textos não eram estudadas em profundidade, dificultando uma análise correcta da complexidade dos fenómenos terminológicos. (Rastier, 1995) (Cabré, 2000) (Kageura et al., 2004)

- Postura prescritiva, normativa e estruturalista, baseada no neo-positivismo lógico. Esta tradição visava a conformidade (e, conseqüentemente, a rigidez) comunicativa, com a institucionalização da biunivocidade e normalização do significado através de vocabulários técnicos. Os domínios de especialidade eram compostos por um conjunto de conceitos ou construções mentais representadas pelos termos, a sua forma linguística. Papel demasiadamente incisivo e limitativo desempenhado pela ISO através do TC37- criticado por alguns como confinado a uma filosofia académica e fora da realidade. (Cabré, 2000; Cabré & Estopà, 2002) (Cabré, 2003) (Candel, 2004) (Kageura et al., 2004) (Smith et al., 2005) (Smith & Ceusters, 2007)
- Visão desfasada dos fenómenos comunicativos, limitando-se ao nível dos conceitos, separando textos e termos. Espelhava uma ilusão metodológica: pretendia-se alcançar a terminologia do desejo e não da realidade. Visão estruturalista da época, demasiado restritiva e conducente a uma monorreferencialidade. A suposta universalidade do conceito conduz a uma caracterização uniforme por parte dos cientistas e técnicos, onde a única diferença seria as designações alternativas das várias línguas. Nas palavras de Rastier, uma linguística dos textos não deve ter postulados normativos, mas sim ser capaz de diferenciar os modos discursivos que caracterizam as disciplinas na sua diversidade. A actividade normativa parece legítima dentro das aplicações, mas não nas teorias. (Desmet, 1994) (Rastier, 1995) (Cabré, 2003) (Cabré, 2006)

A questão fulcral apontada à Teoria Geral da Terminologia não é a sua coerência interna, pois apresenta-se como um corpo sistematizado de princípios teóricos e metodológicos, mas sim o seu carácter reducionista que a impede de explicar a complexidade da comunicação especializada. Era, pois, uma teoria idealista, refutada pelos dados empíricos da verdadeira aplicação da terminologia. Na opinião de Cabré, esta teoria não foi contestada devido à juventude da disciplina, à falta de confrontação científica e ausência de teorias fortes no campo da terminologia.

A Teoria Geral da Terminologia tem sido objecto de análise exaustiva. As suas supostas orientações encontram-se sobejamente exploradas, analisadas, confrontadas e mensuradas, nomeadamente por contraste com outras teorias que surgiram nas últimas décadas⁴².

A relação próxima entre terminologia e ontologia é referida por Wüster. Como recorda, por exemplo, Candel num artigo intitulado “Wüster par lui-meme” (Candel, 2004), no qual cita frequentemente Wüster⁴³, este situava a terminologia em relação com a linguística, a ontologia, as ciências de informação e da documentação, e a normalização, em particular a normalização técnica. Para Candel, existe uma diferença essencial entre terminologia e linguística: é que a terminologia deve fazer empréstimos à lógica e à ontologia, e recorta-se com uma terceira ciência formal, nomeadamente a ciência da informação, tal como afirmou Wüster: “Ein wesentlicher Unterschied zwischen Terminologielehre und Sprachwissenschaft ist darin zu suchen, dass die Terminologielehre Anleihen bei der Logik und bei der Ontologie machen muss und sich mit einer dritten formalen Wissenschaft überschneidet, nämlich der Informationswissenschaft.» (Candel, 2004, p. 16)

Este posicionamento está próximo de determinadas correntes actuais da terminologia: a perspectiva onomasiológica não só se encontra intimamente ligada aos procedimentos mais recentes da inteligência artificial e da linguística computacional, como pode e deve coabitar com a perspectiva semasiológica. De facto, comenta Candel, Wüster defendia que deve existir

⁴² Nos últimos anos apareceram novas abordagens para combater as lacunas da teoria clássica descrita como se fosse uma doutrina. Podemos destacar, entre outras, a teoria comunicativa de Teresa Cabré (Teoria Comunicativa da Terminologia) (1998/99), onde se enfatiza a dimensão comunicativa da terminologia e os aspectos cognitivos e linguísticos; a socioterminologia de Gaudin (1993), que propõe que tenhamos em conta a dimensão social da terminologia; o modelo sociocognitivo de Rita Temmerman (2000) que incorpora os elementos da teoria do protótipo e leva em conta as dimensões diacrónicas e sociais dos termos; a terminologia textual de Bourigault e Slodzian (1999), onde o texto constitui o ponto de partida para descrição dos termos, e o termo apresenta-se como uma construção, ou seja, resulta da análise feita pelo terminógrafo: esta análise leva em conta o lugar ocupado pelo termo dentro de um corpus, uma validação pelos especialistas e os objectivos de uma descrição terminográfica. (L'Homme, 2004)

A partir dos anos 90, surgiram teorias alternativas em três grandes áreas: ciências cognitivas, focando o aspecto construtivo global do conhecimento e as características culturais da categorização; ciências da linguagem, centrada sobre as unidades terminológicas; ciências da comunicação, alertando para a não homogeneidade das situações de comunicação especializada. (Cabré, 2003) (Cabré, 2003a)

⁴³ As citações em alemão correspondem a textos de Wüster extraídos do artigo de Candel.

uma estabilidade entre os termos e os conceitos. Um sistema linguístico forma-se pelo facto de as relações entre as denominações e suas significações serem estáveis ou relativamente invariáveis, caso contrário não haveria compreensão. Nas palavras de Wüster: “Ein Sprachsystem entsteht dadurch, dass die Zuordnung zwischen den Benennungen und ihren Bedeutungen eine bleibende, d.h. relativ unveränderlich ist. Sonst gäbe es keine Verständigung.” (Candel, 2004, p. 17) A partir dessa relação linguística estável, devemos exigir da terminologia que seja, para utilizar uma expressão matemática, biunívoca, ou, por outras palavras, que fundamentalmente a cada conceito não seja atribuída mais do que uma significação, sendo o inverso também verdadeiro, tal como Wüster preconizava:

“Von der bleibenden sprachlichen Zuordnung ist in der Terminologie zu verlangen, dass sie, um einen Ausdruck aus der Mathematik zu gebrauchen, eineindeutig ist. D.h., dass grundsätzlich jedem Begriff nur eine einzige Benennung zugeordnet ist, und umgekehrt. (...) Es sollte also weder mehrdeutige Benennungen (Homonyme und Polyseme), noch Mehrfachbenennungen für einen Begriff (Synonyme) geben.” (Candel, 2004, p. 17)

Mas, ainda segundo Candel, a semasiologia não é ocultada. Apesar da atitude terminológica ser comumente reconhecida como sendo onomasiológica por essência, a semasiologia é tida em conta: “Zum Identifizieren und Fixieren eines Begriffes ist eine Benennung oder ein anderes Zeichen unentbehrlich. Geht man umgekehrt vom Zeichen für den Begriff aus, so wird der Begriff die Bedeutung des Zeichens oder dessen Sinn genannt.” (Candel, 2004, p. 21)

Parece-nos, pois, que o desfasamento apontado a Wüster pelos vários autores acima referenciados, com excepção de Candel, sofre de uma descontextualização temporal, social e económica. Torna-se necessário reposicionar a Teoria Geral da Terminologia na época em que surgiu, entre duas grandes guerras mundiais. Existiam objectivos eminentemente práticos, que requeriam soluções eficientes e exequíveis. Wüster era engenheiro de profissão, preocupava-se com objectos reais, coisas que pudesse visualizar. A sua preocupação com a sistematização das línguas, a eventual eliminação da ambiguidade, o acordo social entre termos e conceitos e a preocupação com o extralinguístico correspondem, em parte, aos objectivos actuais das ontologias. O uso dos termos por comunidades específicas que acordaram princípios de entendimento constitui também uma realidade. Cada vez mais temos nichos conceptuais que se cruzam na plataforma comum de conhecimento formalizado, para que possa ser partilhado

e reutilizado. No entanto, pensamos que o conhecimento que os sistemas de informação pretendem representar dificilmente será atingido se simplificarmos demasiado o objecto de estudo. Os termos e os conceitos fazem parte da complexidade comunicativa do homem e um desequilíbrio nessa análise levaria a resultados porventura altos ao nível informático e estatístico, mas baixos na profundidade das conceptualizações.

Do nosso ponto de vista, numa ontologia temos representações de conceptualizações por meio de uma estrutura que atribui um determinado lugar a determinado conceito através de uma relação acordada pelos especialistas. O texto pode fornecer dados sobre a estrutura conceptual. A orientação wüsteriana atribuía por isso grande importância às relações que os conceitos têm entre si. Já vimos que a questão das relações é fundamental na prática terminológica actual, e pensamos que é tão importante os termos como as relações entre os termos. A terminologia deveria centrar a sua atenção tanto nos termos como nas relações entre termos. Os parâmetros extralinguísticos constituíam para Wüster a base do estabelecimento das relações entre conceitos. Contudo, os parâmetros linguísticos constituem um recurso importante que devemos utilizar. Embora não fosse uma questão central para Wüster, que, ao contrário do que afirma Rastier, não estava preocupado com uma 'linguística dos textos' (Rastier, 1995), estes eram tidos em conta na procura dessa sistematização e dessa relação biunívoca entre termo e conceito. Essa mesma relação biunívoca, tão criticada pelos seus eventuais pressupostos idealistas, é crucial para as linguagens de formalização utilizadas nas ontologias. Uma abordagem onomasiológica, reflecte, assim, e sempre, um trabalho terminológico, dado que os termos estão sempre presentes, no texto ou fora dele.

Do mesmo modo, pensamos que não era objectivo de Wüster criar uma cisão entre linguística e terminologia. Wüster não estava interessado na riqueza e complexidade do funcionamento efectivo linguístico e discursivo dos textos de especialidade. Por questões metodológicas, separou os objectivos das duas ciências.

Finalmente, o eventual carácter reducionista das normas da ISO deve ser visto à luz dos objectivos da época. Devemos antes designá-lo como direccionado para propósitos muito específicos: a necessidade de uma comunicação eficiente. Hoje pretendemos exactamente a mesma coisa com a terminologia e a ontologia: uma comunicação eficiente, que seja inteligível por humanos e máquinas. Os sistemas informáticos fazem parte desse processo.

4.2 Conceito

4.2.1 Ponto de vista da terminologia

A teoria do conceito tem dificuldade em definir-se, uma vez que a terminologia interage com outras disciplinas que comportam uma longa história de preocupação com os conceitos: filosofia, epistemologia, linguística, psicologia. Esta diversidade requer uma teoria do conceito flexível por parte da terminologia, sem pretensões de universalidade. A definição e descrição de um conceito dependem, assim, de um suporte teórico variável e dos objectivos a que se propõem.

Segundo Antia,

“(…) the concept is a unit of thought, knowledge or cognition which: i) though dependent on language is independent of a given language; it can however be influenced by a variety of socio-cultural factors that correspond to linguistic boundaries; ii) is a mental representation, reduction or (re)interpretation of reality that is perceptible, imperceptible or that was previously non-existent; iii) is comprised of characteristics that are (at some point, at least) negotiated within a specialized knowledge community; iv) typically enters into some (organic or logical) relationship with other concepts; v) can exist without symbols (whether linguistic or non-linguistic), but requires symbols for purposes of communication.” (Antia, 2000, p. 82)

Um conceito é relevante enquanto unidade de pensamento quando inicia o seu ciclo de vida; no momento em que está totalmente formado e aceite pela comunidade de especialistas, transforma-se em unidade de conhecimento; a passagem a unidade de lógica dá-se através de uma linguagem de representação formalizada para fins de manipulação computacional.

A norma ISO 1087 (2000) define conceito como “(…) unité de connaissance créé par une combinaison unique de caractères.” (ISO, 2000c, p. 2) Os conceitos podem ter uma definição por compreensão ou por extensão:

“(…) an intensional definition should state the superordinate concept and list the differentiating characteristics in order to position the concept being defined in its

concept system and to delimit it from the other concepts in this system. (ISO/TC 37/SC1/CD704.2 N 133 95 EN) and (...) a definition by extension consists of the enumeration of all species, which are at the same level of abstraction, or of all individual objects belonging to the concept defined' (Felber 1984:163)". (Temmerman, 2000, p. 79)

Em terminologia, os conceitos estão relacionados através de relações lógicas ou ontológicas. Segundo Depecker,

"(...) les relations entre concepts font partie de la définition des termes: ce sont les relations qu'un terme entretient avec d'autres au plan conceptuel que permettent de le situer et de le définir dans un domaine. (...) Les relations logiques sont les relations que s'établissent entre concepts d'un point de vue formel. On peut citer comme relations logiques la relation d'identité, la relation d'implication, la relation d'inclusion, etc.) (...) Les relations ontologiques sont les relations établies para la constitution même des objets dans le monde." (Depecker, 2002, pp. 87-88)

As relações género/espécie adviriam das relações lógicas, enquanto as relações parte/todo proviriam das relações ontológicas.

Fundamentalmente, para a terminologia que parte da análise linguística, um termo é uma entidade lexical e, por isso, um signo linguístico. O signo linguístico foi definido por Ferdinand de Saussure no *Cours de Linguistique Générale*. Para Saussure, um signo linguístico une um significado a um significante, ou seja, um conceito a uma imagem acústica, sendo uma entidade estruturada de duas faces conexas.

O trabalho do terminólogo consiste, de um modo geral, em fazer corresponder a cada conceito uma ou mais designações linguísticas. A definição em língua natural determina se o conceito descrito corresponde ao objecto, entidade extralinguística. Do ponto de vista do sentido, podemos considerar duas ou mais unidades linguísticas como sinónimos se uma puder substituir a outra, mas, do ponto de vista da designação, serão consideradas equivalentes se reverterem ao mesmo conceito e, através deste, ao mesmo objecto.

Para Conceição, a perspectiva semasiológica encara a evolução do conhecimento como algo natural e desejável, por isso estuda a actualização constante do significado dos termos em

discurso. Enquanto unidade de conhecimento, um conceito não é fixo, mas sim situacional, no sentido em que a sua análise vai revelando novos traços que o modificam ao nível conceptual, alterando, simultaneamente, o termo. A análise terminológica deve dar conta desta evolução que não implica necessariamente uma alteração da denominação. O que muda são os traços conceptuais que se exprimem sob forma linguística. Como a denominação, para o autor, não provém de nenhum traço conceptual, permanece imutável. E recorre a uma citação de Diki-Kidiri para ressaltar a distinção entre conceito e significado: "(...) la distinction entre concept, entendu comme structure cognitive de catégorisation, et signifié, envisagé comme le lieu des perceptions culturelles, s'impose encore plus nettement lorsqu'on compare la dénomination d'un artefact dans plusieurs langues."⁴⁴ (Conceição, 2005, p. 56) Ainda segundo o autor, quando a pesquisa terminológica é feita a partir de dados textuais, em discurso, acedemos em primeiro lugar ao sentido (atualização em discurso) e depois à significação (sentido do signo na língua). Após uma análise exaustiva do ambiente discursivo dos termos e de consulta ao especialista de domínio, é que partimos para a análise do nível conceptual. A utilização de corpora textuais parte assim de realizações linguísticas para alcançar os correspondentes semânticos e cognitivos: "C'est dans le rapport entre le savoir linguistique et le savoir spécifique à un domaine de connaissance que réside l'aspect interdisciplinaire de la recherche en terminologie." (Conceição, 2005, p. 28)

O conceito faz, então, como afirma Cabré, a ligação entre a terminologia e a teoria do conhecimento:

"Terminology and the theory of knowledge are closely linked to one another by the notion of concept, which is the foundation of the theory of terms and the starting point

⁴⁴ Depecker comenta igualmente que não devemos reduzir o conceito ao significado. Uma das melhores provas da distinção entre conceito e significado é precisamente o facto de ser possível observar as ligações entre conceito e signo, e isso não é possível a partir do próprio signo, revelando a natureza polissémica dos signos linguísticos. (Depecker, 2002) Este comentário de Depecker aproxima-se da vertente onomasiológica, uma vez que o conceito seria independente da língua, constituindo-se como algo extralinguístico. A perspectiva semasiológica aceita também como princípio que o conceito é uma unidade de pensamento, portanto um elemento cognitivo, extralinguístico. Simplesmente, parte do semântico, da manifestação em discurso, para o conceptual. A perspectiva onomasiológica pode, num contexto computacional, prescindir da parte semântica.

for terminological research. The set of terms in a special subject field is at the same time a structure of concepts reflecting the organization of knowledge about the field in question. Terms thus become the key to the representation of specialized knowledge. An intelligent expert system, which must know these concept structures in order to process information and act accordingly, internalizes structured knowledge by means of terms. Each term is an item of specialized thought, and the relationships among terms in a single special field reflect the field's concept organization.”(Cabré, 1998 (1992), p. 167)

Comentando as posições dos autores, observamos que Conceição e Cabré não rejeitam a premissa de Wüster de que o trabalho terminológico parte do conceito enquanto unidade de pensamento. Mas a valência de unidade de conhecimento aparece mais associada ao termo, já que, sendo a pesquisa feita em ambiente discursivo, é o conjunto dos termos que reflecte o conjunto dos conceitos organizados de tal forma que expressam o conhecimento do domínio. O papel-chave é transferido para o termo, unidade que, efectivamente, aparece em discurso. Assim, são as relações entre os termos que apontam para as relações entre os conceitos. O conceito é algo que vai sendo construído através da análise terminológica, alterando os seus traços semânticos em função do discurso e do contexto de interpretação.

Se atentarmos na opinião de Conceição, vemos que, para um ponto de partida linguístico, as etapas naturais da análise terminológica são: i) termo, ii) especialista, iii) conceito. O conceito existe na língua, o significado existe no discurso. No discurso acedemos primeiro ao sentido (actualização em discurso) e depois à significação (sentido do signo na língua).

Repare-se que o extralinguístico reverte sempre ao linguístico, destino final da interpretação. O conceito, o significado, o sentido e a significação reportam todos à manifestação linguística. Pensamos pois, que, numa perspectiva linguística, a terminologia centra-se em áreas de especialidade nas quais estuda o conhecimento na sua relação com a expressão, interessando-se tradicionalmente e em particular pela parte linguística através da qual o conceito é definido ou descrito e pelas relações que se estabelecem entre esses conceitos, nomeadamente através das relações de hiperonímia/hiponímia/meronímia, dos designados “knowledge rich contexts”(Meyer, 1991) e dos contextos de reformulação. Os conceitos variam em função das conceptualizações e do ponto de vista de quem o descreve e/ou define. Enquanto designações

para os conceitos, os termos são um meio de adquirir, criar, comunicar, armazenar, representar e operacionalizar conhecimento.

Recordemos que a norma 1087-1:2000 da ISO define a terminologia como o estudo científico dos conceitos e termos utilizados nas línguas de especialidade. Uma língua de especialidade seria um sub-sistema linguístico que utiliza terminologia e outros meios linguísticos com o objectivo de alcançar uma comunicação sem ambiguidades num determinado domínio. Pensamos que uma perspectiva puramente onomasiológica, centrada somente no conceito e no extralinguístico, corre o risco de cortar a ligação dos conceitos às conceptualizações, uma vez que uma conceptualização, do ponto de vista cognitivo, está ligada ao locutor, à sua cultura e à língua natural em que se expressa, ou seja, ao discurso, objecto privilegiado da análise semasiológica. Confinar a terminologia a sub-sistemas designados de 'língua de especialidade' seria também acreditar que a actividade humana pode fragmentar-se em mundos inteiramente fechados. O discurso é, por esse motivo, um dos locais onde podemos apreender facetas importantes do conhecimento especializado.

Assim, na nossa opinião, uma terminologia na perspectiva linguística não é prioritariamente um conjunto de conceitos, mas sim um conjunto de expressões da língua natural que manifesta, através das relações que estabelece em discurso, traços semânticos que apontam para uma eventual organização conceptual. Uma língua de especialidade não se reduz, por isso, a uma terminologia: utiliza termos linguísticos e não linguísticos nos seus enunciados, mobilizando os recursos da língua natural para transmitir conhecimento especializado. Na semasiologia, a análise dos dados terminológicos baseia-se numa teoria geral da língua, embora a capacidade de comunicar numa língua de especialidade não deva limitar-se a uma competência linguística. Necessitamos tanto de conhecimento científico e técnico, como de conhecimento semiótico para interpretar um texto em língua de especialidade.

A perspectiva semasiológica parte do texto, encara os termos num âmbito alargado de comunicação, onde unidades lexicais gerais e unidades lexicais especializadas formam um todo coerente que extravasa o contexto do texto, embora a sua significação se constitua em discurso. Esta constatação não equivale a afirmar que o conjunto de candidatos a termo que podemos encontrar num texto de especialidade reflecte uma estrutura organizada de conceitos. Se fornecermos a uma pessoa que não dispõe de conhecimento do domínio esse

conjunto de termos, certamente não conseguirá elaborar um pensamento coerente e reflecti-lo em discurso, de tal forma que consiga reescrever o texto. Porém, se dermos a um especialista um discurso desconexo sob o ponto de vista semântico, mas incluindo esse conjunto de termos, consegue reorganizar o discurso de tal forma que transmita linguisticamente partes do sistema conceptual desse domínio. O conceito distingue-se, assim, do significado, do referente e da sua representação. O termo continua porém, a ser uma unidade indispensável na organização do conhecimento.

Do nosso ponto de vista, embora uma perspectiva semasiológica não seja adequada para, sozinha, formalizar conhecimento num sistema computacional, pode contribuir com informação sobre os conceitos e as suas denominações, abarcando, por isso, o linguístico e o extralinguístico, uma vez que o significado dos termos contém também uma significação ligada ao acto comunicativo e cultural da sua produção, ligada, portanto, a uma parte conceptual e extralinguística.

A terminologia que parte de uma abordagem linguística privilegia a verbalização do conhecimento. A terminologia que parte de uma abordagem extralinguística centra-se na conceptualização do conhecimento.

4.2.2 Ponto de vista da inteligência artificial

A inteligência artificial pretende representar conhecimento estático, normalizado, formalizável através de linguagens de representação. Defende, naturalmente, uma abordagem diferente relativamente ao conceito.

A adopção de uma ou de outra teoria vai condicionar todos os resultados da análise. A inteligência artificial adopta uma postura baseada, essencialmente, nos princípios aristotélicos e na organização do conhecimento a partir do conceito enquanto unidade de conhecimento extralinguístico, embora considere que uma ontologia requer uma abordagem interdisciplinar. Como afirma Roche, "(...) the ontology problem requires a multidisciplinary approach based on sound epistemological, logical and linguistic principles." (Roche, 2003, p. 6) Uma ontologia lida com questões linguísticas (usamos palavras para comunicarmos), epistemológicas (essas palavras têm um significado) e lógicas (como garantia de coerência).

É interessante constatar a deslocação de prioridades por parte dos terminólogos e linguistas, de um lado, e dos engenheiros dos sistemas de informação, do outro: enquanto os primeiros se afastaram nas últimas décadas do estruturalismo de Saussure, identificado com uma postura normativa e positivista da língua enquanto sistema, os segundos viram no triângulo semiótico uma base para a modelação e representação formal do conhecimento. Para manipularmos o conhecimento do domínio necessitamos de um meio de representação. Tendo por base o triângulo semiótico clássico, o significado do termo é o conceito a que se refere. E, nas palavras de Roche, um conceito não deve reduzir-se a uma fórmula 'bem formada'.

Podemos observar nos princípios aristotélicos as bases da teoria do conceito adoptada pelas ontologias computacionais. O princípio da *differentia* tal como aparece nos trabalhos de Aristóteles e no *Isagoge* de Porfírio vai, em algumas metodologias, constituir-se como a espinha dorsal da construção da rede conceptual da ontologia. Os cinco predicados enunciados por Porfírio - género, diferença, espécie, próprio e acidente – seguem uma abordagem mais epistemológica e lógica do que linguística. Partindo da teoria dos predicados de Aristóteles, a classificação de Porfírio é um conjunto hierárquico finito de géneros e espécies, que funciona por dicotomias sucessivas. Codificando, pela primeira vez, o esquema dicotómico que procede do geral ao particular, da maior extensão à maior compreensão, do género à *species infima*, a classificação de Porfírio iria ter uma influência substancial na posterior teoria da definição e classificação dos seres.

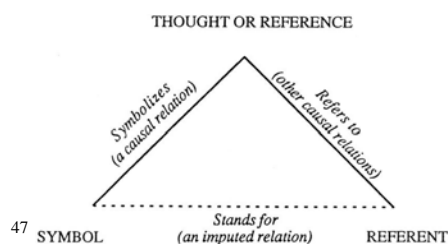
A dificuldade da classificação baseada na diferença específica - propriedade "(...) incompatível com todas as outras propriedades situadas no mesmo nível da classificação (...)" (Perelman, 1963: 232)⁴⁵ - é proporcional ao número de propriedades que podem desempenhar essa função. A realidade apresentará contornos diferentes, conforme a selecção dessa propriedade.

⁴⁵ Baseando-nos em Pombo, vemos que, num artigo intitulado "Réflexions philosophiques sur la classification" (1963), Perelman distingue dois tipos fundamentais de classificações: 1) as dicotomias baseadas na presença ou ausência de uma determinada propriedade, 2) as classificações baseadas numa propriedade qualificada como diferença específica. Ainda segundo Perelman, as dicotomias seriam as "classificações logicamente mais satisfatórias" (1963: 232) uma vez que nelas o espaço classificatório, isto é, o conjunto dos objectos a classificar, é sucessivamente dividido em dois subconjuntos simultaneamente exclusivos e exaustivos. (Pombo, 1998)

É aqui que se coloca a questão da naturalidade ou arbitrariedade da classificação, uma vez que a escolha da classificação se faz então a partir da maior ou menor atenção às afinidades que devem permitir reunir em classes as diversas realidades a classificar. Toda a classificação combina, assim, diferentes modos de aproximação à multiplicidade dos objectos a classificar: “A concept is defined by specific differentiation. It means that a concept is defined from a previously existing concept adding a new *differentia* which is then called the ‘*specific differentia*’ of the newly created concept.” (Roche, 2003a, p. 3) Um conceito pode ser definido pelo conjunto das suas diferenças, unidades elementares que não têm significado em si mesmo. Ao contrário de um atributo, a diferença não pode ser removida da definição de um objecto sem alterar a sua natureza; nem lhe pode ser atribuído um valor. Ao adicionarmos uma diferença a um conceito já existente, criamos dois conceitos novos, o primeiro ao qual a diferença pertence e o segundo que nunca poderá ter essa diferença, funcionando como uma garantia de coerência da ontologia. O significado dos conceitos estrutura-se em árvores binárias baseadas em pares de diferenças opostas: “The meaning of a term is defined by a set of *differentiae*.”⁴⁶ (Roche, 2003a, p. 4)

As teorias de Charles Peirce (1991) e o famoso triângulo de Ogden e Richards (1923)⁴⁷ contribuíram para uma investigação mais profunda do tema e adaptação à construção de ontologias. Segundo Sowa, um signo é, simultaneamente (1) uma *entidade* que representa (2) outra *entidade* a (3) um *agente*. As ontologias seriam então compostas por categorias que constituem uma rede apertada de signos, cujas conexões primárias, porém, não se encontram numa máquina, mas na mente das pessoas que os interpretam. O esforço que vem sendo

⁴⁶ Sowa define *differentia* como: “The properties, features, or attributes that distinguish a type from other types that have a common supertype. The term comes from Aristotle’s method of defining new types by stating the *genus* or supertype and stating the *differentiae* that distinguish the new type from its supertype. Aristotle’s method of definition has become the *de facto* standard for natural language dictionaries, and it is also widely used for AI knowledge bases and object-oriented programming languages” (Sowa, 2006).



desenvolvido pela inteligência artificial é tornar explícitas essas conexões mentais. A fonte última de sentido será sempre o mundo exterior e os agentes que utilizam os signos para representar entidades no mundo e as suas intenções em relação a essas entidades. Baseado no triângulo semiótico, Sowa explicita que uma metalinguagem consiste em signos que significam alguma coisa sobre outros signos, mas o que eles significam depende das relações que esses signos têm uns com os outros, com as entidades que eles representam, e com os agentes que os usam para comunicar com outros agentes.

Aristóteles distinguiu objectos, as palavras que se referem a eles, e as correspondentes experiências na psique. Frege (1879) e Charles Peirce (1991) adotaram essa distinção tripartida de Aristóteles e usaram-na nos fundamentos semânticos dos seus sistemas de lógica. Aristóteles observou que os símbolos podiam simbolizar outros símbolos, uma vez que as palavras escritas constituem os símbolos das palavras faladas. A possibilidade de conectar múltiplos triângulos de maneira diferente através da ligação dos vértices foi reconhecida por Peirce, fazendo com que possamos representar signos de signos. Para Sowa, a semiótica tem de se preocupar não só com as relações entre signos, mas com as relações destes com o mundo e com os agentes que agem sobre o mundo. O carácter evolutivo dos signos possibilita a sua interconexão.

Pensamos que a recuperação das teorias de base linguística parece apontar para uma consciência por parte da inteligência artificial da dificuldade em formalizar a língua natural, um obstáculo, porventura o mais difícil de transpor, na representação do conhecimento. Conforme tivemos oportunidade de observar, o conhecimento encontra-se na mente das pessoas. Tornar os seus traços inteligíveis para as máquinas, nomeadamente através da criação de ontologias, significa aceder às conexões mentais que relacionam os conceitos e que formam redes conceptuais potencialmente detentoras de conhecimento.

A inteligência artificial adopta as premissas aristotélicas baseadas num raciocínio lógico e a base semiótica triangular entre conceito, símbolo e referente, encarando o conceito como algo extralinguístico, embora não o confunda com o mundo que representa. Vai buscar, em primeiro lugar, o especialista, detentor, acima de tudo, do conhecimento tácito e implícito. Ao contrário do que afirmam os que partem de uma perspectiva linguística (segundo alguns autores, o conceito existe na língua e o significado no discurso), aqui o conceito não pertence a

nenhuma manifestação linguística e o que importa é a significação, e não o significado, uma vez que as relações entre conceitos não são objecto de nenhuma actualização em discurso. Pelo contrário: pretende-se capturar conhecimento estático e consensual, pois só assim poderá ser passível de manipulação computacional. Por isso, Beck e Pinto afirmam: “In pure form, an ontology is not about terms at all, only about concepts which ideally are represented in a form independent of terms in any natural language.”⁴⁸ (Beck & Pinto, 2002, p. 8)

A abordagem onomasiológica adoptada pela inteligência artificial aproxima-se mais da postura atribuída a Wüster e à designada Terminologia Clássica⁴⁹ e distancia-se da prática terminológica centrada no estudo dos termos em discurso. Se recordarmos os princípios de Wüster, uma das preocupações fundamentais era uma comunicação eficaz, mono e multilingue, daí a necessidade da relação biunívoca entre termo e conceito. Uma ontologia é, prioritariamente, um artefacto, uma construção previamente acordada entre os membros de uma comunidade para partilha e reutilização entre agentes, humanos ou não. Entende-se a preocupação dos engenheiros com a minimização de elementos que causem ‘ruído’. E a língua natural é um elemento altamente perturbador. Basta pensarmos, como diz Roche, nas figuras de retórica utilizadas num texto para alterar completamente um conceito.

O conceito é então encarado na perspectiva da inteligência artificial como um somatório de características. Ao aceitarmos as regras da teoria do conceito adoptada pela ontologia, temos de aceitar que são aquelas características, e não outras, que constituem aquele conceito num determinado momento de captura e representação daquele fragmento de conhecimento. A valência do conhecimento centra-se no conceito, e não no termo. As relações entre os conceitos reflectem a organização do conhecimento de domínio, cuja rede é elaborada com

⁴⁸ Na nossa opinião, a afirmação de Beck e Pinto é contestável. Conforme os próprios autores, contraditoriamente, afirmam, a parte conceptual é capturada através das relações entre termos: “(...) all information directly related to [this] concept is directly accessible via the term relationships. Moreover, it can use these relations to examine similar or related concepts. In this way, the user gets access to all and only the information available on a particular concept.” (Beck & Pinto, 2002, p. 5) A eliminação da importância do termo deve ser, por isso, objecto de uma reflexão mais profunda por parte da inteligência artificial.

⁴⁹ Pensamos que o termo ‘terminologia clássica’ deveria ser objecto de uma reapreciação na literatura sobre esta matéria. Se analisarmos, por exemplo, a posição de certos autores da inteligência artificial, observamos que, no seu entendimento, terminologia clássica é a postura adoptada pela semasiologia.

recurso prioritário ao especialista que trabalha em conjunto com o engenheiro de ontologias. A especificação da conceptualização de que fala Gruber é feita através de uma linguagem, e não da língua.

No nosso entendimento, o que tendencialmente se verifica é que a linguagem de representação descreve, mas não define, o conceito. Da mesma forma que determinadas correntes da terminologia reagiram a Wüster, dada a facilidade de acesso a grandes quantidades de texto em formato electrónico altamente aliciantes sob o ponto de vista de análise discursiva, a inteligência artificial reage à terminologia de base lexical devido, essencialmente, às dificuldades geradas pela língua natural. As exigências do mundo actual necessitam de resultados rápidos, reduzindo substancialmente o espectro do que é possível, necessário e conveniente formalizar. O que é importante sublinhar é que não se deve confundir representação do conhecimento com conhecimento. Dada a impossibilidade de existir uma conceptualização 'correcta' ou 'verdadeira' sob determinada realidade, reflecte-se uma visão do mundo definida por um determinado indivíduo ou grupo de indivíduos, sendo válido para um objectivo específico.

É fundamental percebermos que a terminologia no seu momento actual encontra-se profundamente ligada à inovação em todas as suas vertentes: científica e académica, social, cultural e económica. Como afirma Gresser (2008), isto implica uma cooperação apertada entre especialistas, terminólogos e engenheiros de ontologias para detectarem atempadamente a emergência de novos conceitos ou novos artefactos computacionais. A terminologia e a ontologia servem propósitos utilitários, mas o conceito de conhecimento para um terminólogo quase sempre tem contornos diferentes do conceito de conhecimento para um engenheiro do conhecimento. É essencial reduzir esse afastamento para que ambas as práticas possam colaborar eficazmente. O encontro de sinergias tenta extrair o melhor de dois mundos: por um lado, a normalização através da estabilização dos significados em representações formalizadas, por outro a captura da exuberância criativa e ininterrupta da língua e do pensamento humano.

Nas palavras de Meyer,

“(...) the major challenge of terminology is conceptual, not linguistic: terminologists are trained in linguistics and thus are properly prepared for the linguistic dimension of their task; in contrast, they are not normally domain experts, yet they require a substantial amount of expert knowledge in order to do their work. In other words, the major difficulty is pinning down the meanings of terms.” (Meyer, 1991, p. 23)

4.3 Relações semânticas e relações conceptuais

Observa-se recorrentemente na literatura uma utilização indistinta dos termos ‘relação semântica’ e ‘relação conceptual’. Do anteriormente exposto, torna-se mais claro perceber, ainda que teoricamente, de onde podemos extrair uma e outra. Para nós, uma relação semântica encontra-se ao nível do texto, e uma relação conceptual encontra-se ao nível da conceptualização extralinguística.

De acordo com Lopes e Rio-Torto,

“(...) a Semântica envolve o conhecimento intuitivo do significado das palavras de uma língua, das regras que presidem à construção das predicções e dos mecanismos que garantem a sequencialização de enunciados no plano discursivo/textual. A explicitação desse conhecimento é feita no âmbito da Semântica enquanto disciplina linguística, que assim se ocupa da descrição do significado das palavras, das frases e do texto.”(Lopes & Rio-Torto, 2007, p. 13)

Assume-se, assim, que a semântica é um ramo da linguística que se ocupa da análise do discurso em todas as suas vertentes e dimensões. Nem sempre, porém, isso é claro, independentemente da adoção de uma abordagem semasiológica ou onomasiológica.

De uma forma geral, identificam-se em discurso as relações semânticas de hiperonímia/hiponímia, holonímia/meronímia⁵⁰. Autores como Sérgio Barros incluem também

⁵⁰“Homonímia: um termo é considerado homonímico quando em várias significações designa dois ou mais conceitos sem que se possa estabelecer relações entre eles; Hiperonímia: a relação entre dois termos é classificada como hiperonímica, quando a relação hierárquica que se estabelece é de inclusão, partindo do genérico (superordenado) para o específico (subordenado), sendo que o

a sinonímia e a antonímia como pertencendo a este grupo de relações. A semântica lexical distingue estes tipos de relações semânticas que são objecto de identificação automática através da extracção computacional de termos. O autor chama a atenção para o facto de que quando nos referimos aos objectos, particularmente quando os definimos, estamos a fazer uma ‘transposição’ entre dois níveis de análise distintos ou seja, passamos de um nível concreto e ligado à realidade extralinguística para um nível de abstracção que requer o uso da língua. Quando essas propriedades são transpostas para o nível linguístico por via de uma abstracção/conceptualização, falamos de características ou traços semânticos que ajudam à elaboração de uma definição.⁵¹(Barros, 2007)

Na tentativa de distinguir estes dois planos de análise, Meyer refere que “(...) we use the term *linguistic* (in opposition to *conceptual*) in the very general sense of ‘related to the terms, as opposed to the concepts designated by the terms.” (Meyer, 1991, p. 30) Para a autora, os dados conceptuais incluem informação sobre o domínio, sinónimos e antónimos, definições, equivalentes; os dados linguísticos incluem informação gramatical, anomalias morfológicas, expressões idiomáticas. A autora separa o conhecimento léxico-semântico (linguístico) e o conhecimento enciclopédico (sobre o mundo). A terminologia é uma disciplina prática preocupada com a descrição e nomeação/designação dos conceitos em domínios de especialidade. Dado que os conceitos são o ponto de partida de todo o trabalho prático em terminologia, e dado que os conceitos são os blocos que constroem o conhecimento, a terminologia torna-se uma actividade muito intensa nesse âmbito: *descrever* conceitos envolve adquirir conhecimento sobre as suas características e *nomear/designar* os conceitos envolve

primeiro impõe sempre as suas propriedades ao segundo, criando assim entre eles, uma dependência semântica. (Costa & Silva, 2006, p. 9)

“Hiponímia – relação de inclusão unilateral; meronímia – relação de inclusão de uma parte num todo.” (Mineiro, Doria, Antunes, & Correia, 2004, p. 2)

⁵¹ De acordo com Barros, para Wüster o conceito é tomado como o ponto de partida para abordar as relações conceptuais, que lhe permite traçar uma distinção entre relações lógicas e ontológicas (Wüster 1979:40). As relações lógicas baseiam-se no grau e tipo de semelhança entre os conceitos e subdividem-se em relações de subordinação e de coordenação/diagonais - correspondem às relações genéricas da norma ISO 704 2000. As relações ontológicas derivam do contacto no espaço e no tempo entre entidades representadas por conceitos, e correspondem às relações partitivas da norma ISO 704 2000. (Barros, 2007)

fazer o *matching* das características conceptuais com as formas linguísticas (os termos). (Meyer, 1991)

As dificuldades em discernir os dois tipos de relações são também expressas por Marie Claude L'Homme. A autora entende que certas relações léxico-semânticas estão muito próximas das relações dadas pelas representações conceptuais e é muito difícil distinguir o que vem do conceito ou o que vem do termo. A terminografia dá conta das diferentes relações semânticas entre os termos, procurando também identificar as estruturas que nos dão essas relações, sendo que a recolha de termos de um domínio está estruturada e as unidades linguísticas não estão desordenadas e sem ligação. As relações que dizem respeito ao termo são as semânticas. A lista das relações conceptuais – genéricas, partitivas ou funcionais (*is_a*, *part_of*, *product_of*, etc.) - é elaborada em função dos conceitos de um domínio, sendo sua preocupação uma articulação consensual do conhecimento. (L'Homme, 2004)

Como refere Costa (2005), os termos confundem-se com unidades lexicais gerais, uma vez que são signos linguísticos idênticos do ponto de vista morfológico, sintático e lexical. Assim, as relações semânticas confundem-se por vezes com as relações conceptuais.

Para Cabré, por exemplo, as relações conceptuais são expressas por padrões linguísticos e definidas como “(...) a semantic link, that is, a particular predication among two or more specialized knowledge units” (Cabré & Feliu, 2002, p. 1), sendo que os conceitos e as relações conceptuais são elementos essenciais para a estrutura cognitiva relacionada com conhecimento especializado. Em textos de especialidade, os conceitos são expressos por termos e as relações conceptuais tornam-se as chamadas relações semânticas quando se tenta identificá-las usando diferentes expressões linguísticas. (Feliu, Vivaldi, & Cabré, 2006) Rastier, por seu lado, debruçando-se sobre este assunto, acaba por introduzir também um certo grau de ambiguidade, quando afirma que “(...) les relations sémantiques sont cruciales dans les ontologies, puisqu’elles typent les liens entre les concepts (correspondant aux entités). Le mot sémantique doit être ici entendu dans une acception logique et non linguistique : il s’agit en fait de relations conceptuelles (...) » (Rastier, 2004, p. 25)

Iniciando a nossa análise crítica pela posição dos dois últimos autores, podemos observar uma tendência de sobreposição da relação semântica com a relação conceptual, onde a manifestação linguística coincidiria com a conceptualização. No nosso ponto de vista, o discurso sobre o mundo, a representação do mundo através de uma linguagem formal e o próprio mundo são entidades perfeitamente distintas. O que é crucial para uma ontologia é a relação conceptual. Sem ela, não se consegue construir a rede de conceitos que vai representar formalmente o conhecimento. Como já referimos neste trabalho, dependendo do objectivo da ontologia, podemos ou não utilizar a análise de texto para extrair elementos que contribuam para a construção da ontologia, ou seja, podemos utilizar as relações semânticas para ajudar na representação. Esta posição não equivale a afirmar que devemos ou podemos, em determinados casos, prescindir do termo.

Para nós, a perspectiva que parte de uma abordagem linguística e semasiológica apoia-se na semântica mas também se distingue dela, não só porque se concentra no estudo dos termos e das relações entre termos – envolvendo por inerência outros elementos lexicais que contribuem para essa análise - mas também porque se preocupa com o estudo dos termos no seio da vida social, procurando compreender o mundo e encontrando as manifestações linguísticas adequadas para o representar. O facto de Meyer e L’Homme fazerem coincidir o semântico com o linguístico, através da constatação de que as relações semânticas que ligam as unidades linguísticas não se encontram ao acaso, sendo portadoras de uma organização que faz parte da sua utilização em discurso, não deve proporcionar situações em que, frequentemente, somos tentados a encarar essa estrutura discursiva como um reflexo de uma organização conceptual equivalente a conhecimento.

Na perspectiva da inteligência artificial, a relação conceptual não está dependente da relação semântica, uma vez que a relação semântica se estabelece em discurso, e, teoricamente, uma ontologia não deve depender de nenhuma língua natural para poder ser partilhada e reutilizada eficazmente pelos sistemas.

Para autores com Rastier, este é o ponto fraco de uma ontologia. Assumindo uma postura crítica relativamente à representação do conhecimento através de ontologias, considera que as relações lexicais dependem dos tipos de discurso, enquanto uma ontologia continua, em princípio, a ser a mesma, independentemente do género e do tipo de discurso. Uma ontologia

não tem, por isso, em conta, a diversidade dos discursos, já que tem como objectivo afirmar a 'unidade do Ser' de acordo com os princípios aristotélicos. Esta situação afecta a sua utilidade descritiva. Para o autor, as ontologias tentam reduzir os factores de imperfeição linguística, entre eles a polissemia e a sinonímia. Isto porque a concepção instrumental da língua prevê ligações idealmente transparentes entre língua e pensamento. As ontologias, enquanto sistemas baseados na unidade, não aceitam bem a multiplicidade. A recusa da polissemia e da sinonímia apontaria para uma suspensão do contexto. Nas suas palavras, estes artefactos traduzem uma angústia ontológica sobre a capacidade de representação da língua, bem como sobre a estabilidade do mundo. Como os sistemas informáticos são desprovidos de capacidades interpretativas, prima a representação sobre a interpretação. (Rastier, 2004)

De facto, dez anos antes Rastier (1995) já dissertava com a mesma orientação sobre o tema da imperfeição das línguas, ligado ao objectivo normativo de Wüster, ao desenvolvimento das ciências e à laicização progressiva da noção de conhecimento. A questão de saber até que ponto as línguas podem ser um instrumento fiel de um pensamento correcto - de acordo com o postulado de Condillac de que a ciência é uma língua 'bem feita' - traduz o vasto movimento de perfeição que o homem tentou desenvolver durante séculos. A difusão das disciplinas técnicas e científicas insistiam no carácter internacional das terminologias. O autor retém a passagem de Dante em *De Vulgari Eloquentia* (I, 7) onde se estabelece que só as línguas especializadas permitem à humanidade sobreviver à confusão de Babel. A redenção de Babel passaria então pelas terminologias, num aperfeiçoamento da língua que retrata o momento histórico de devastação posterior às grandes guerras mundias. Assim, uma língua perfeita seria internacional por razões míticas (redenção de Babel) e práticas (assegurar uma comunicação fácil no seio da humanidade). Os traços característicos reformulados pela terminologia de orientação onomasiológica – papel ideográfico da linguagem, permanente no tempo, invariável no espaço, sem alterações dialectais ou diacrónicas – só poderiam ser alcançados por algo artificial e por uma vontade normativa dominante.

Os argumentos de Rastier traduzem, de facto, as preocupações de uma ontologia no sentido de artefacto de engenharia para fins de representação formal do conhecimento: não está centrada na diversidade discursiva e no contexto do texto. Preocupa-se antes com a permanência e a invariabilidade dos conceitos e das suas relações ao nível extralinguístico.

Como vai formalizar os dados numa linguagem de representação formal, questões como a imperfeição das línguas e a suspensão do contexto não se colocam. Como Gruber afirma, uma ontologia é um contrato, um acordo. É, portanto, algo artificial. Assume que não faz parte dos seus objectivos representar todo o conhecimento, e mesmo os pequenos pedaços de conhecimento que tenta representar são especificações de conceptualizações que, necessariamente, reflectem uma determinada visão do mundo, embora sujeita a um consenso prévio dentro da comunidade. Mais uma vez, o conceito de conhecimento não obedece à mesma interpretação.

Na realidade, é precisamente na interpretação que, no nosso ponto de vista, reside um dos problemas maiores do conflito entre as duas abordagens. Considera-se que um texto se constitui como objecto de interpretação, enquanto uma ontologia pretende constituir-se como uma representação e não está aberta à interpretação. Será esse o verdadeiro desejo da inteligência artificial? Pensamos que não. O objectivo da inteligência artificial consiste, julgamos, em conseguir chegar à representação das conexões mentais do conhecimento, tal como um leitor consegue interpretar um texto. Se existe angústia, como afirma Rastier (2004), ela está do lado da linguagem formal e não da língua natural. O problema fundamental da semântica formal consiste na relação entre referência e inferência, uma tarefa de reconciliação difícil para a inteligência artificial.

Se dermos continuidade à definição de Semântica de Lopes e Rio-Torto, vemos que

“(...) a Semântica ocupa[-se] dos significados explícitos, convencionais e invariantes das expressões linguísticas, aqueles que permanecem estáveis independentemente das situações de uso. (...) o significado linguístico codifica informação sobre o mundo e desempenha um papel de relevo na configuração dos nossos estados mentais. (...) A semântica propõe-se tornar explícito [esse] conhecimento intuitivo intersubjectivamente partilhado.” (Lopes & Rio-Torto, 2007, p. 14)

Os objectivos das relações semânticas aproximam-se, assim, dos objectivos das relações conceptuais: a captura de algo explícito e invariável para efeitos de partilha. O ponto de partida do objecto de estudo, condicionado pelo objectivo desse estudo, vai determinar o contributo de cada uma.

5. Metodologias de representação do conhecimento

A terminologia desempenha um papel fundamental na construção de recursos de representação do conhecimento. Baseando-nos nos esclarecimentos anteriores sobre o que se pode entender por uma ontologia na área da inteligência artificial, nas várias acepções de conhecimento, nas orientações metodológicas e teóricas da terminologia e nas duas grandes abordagens relativamente à teoria do conceito, interessa agora observar, sob o ponto de vista teórico, as metodologias aplicadas pela terminologia e pela inteligência artificial na construção desses artefactos. Enquanto práticas, a terminologia e a ontologia podem servir-se de várias metodologias para a construção de recursos. Devem, por isso, estar em sintonia quanto ao ponto de partida de análise do objecto de estudo.

Na construção de ontologias, as metodologias que partem do linguístico baseiam-se essencialmente em corpora electrónicos, submetendo-os a análises linguísticas e estatísticas para tentar extrair candidatos a termo e relações entre esses termos que denunciem uma eventual estrutura conceptual que, por sua vez, seja a tradução da organização de conhecimento. A língua natural é o seu ponto de partida, principalmente o texto, encarado como espaço de actualização de sentidos e significados.

Uma outra perspectiva, adoptada por algumas abordagens da inteligência artificial, encara uma ontologia como uma construção totalmente extralinguística, tendo como papel central o especialista e o engenheiro do conhecimento que conceptualizam, a seu ver, uma realidade diferente da que se poderia eventualmente extrair a partir de elementos linguísticos presentes no texto. Nesta abordagem o texto pode ser prescindível, mas não a terminologia. A terminologia, neste ponto de vista, está para além do texto e não se deve prender ao discurso verbal sob pena de não poder ser partilhada e reutilizada. O conhecimento é uma meta a atingir através de recursos estáticos e unívocos, sem abertura à interpretação, mas sim à representação. Os sistemas de informação só conseguem funcionar desta forma. O texto de especialidade é, na maioria das vezes, um elemento perturbador e gerador de ruído.

A evolução tecnológica proporcionou duas situações antagónicas: à medida que nos entusiasmos com as quantidades infindáveis de texto em formato electrónico na internet, vamos, simultaneamente, tendo noção de que é extremamente difícil retirar desse texto aquilo que designamos por conhecimento. Existem no mercado inúmeras linguagens de representação que tentam ir além da sintaxe, numa aproximação às conceptualizações humanas. Essas aproximações serão facilitadas através da língua natural ou fora dela? Se optarmos por uma metodologia extralinguística, estaremos a desvirtuar o trabalho do terminólogo?

5.1 Perspectiva linguística

5.1.1 Complexidades da língua natural

Natural language is and always must be the primary knowledge representation device.

(Wilks, 2008, p. 3)

Uma ontologia partilhável e reutilizável tem de ser constituída na base do consenso entre os utilizadores. Mesmo que concordássemos relativamente aos conceitos do nível de conhecimento que queremos partilhar, não podemos transferir esses conceitos electronicamente a não ser através de compromissos sobre a forma como os conceitos devem ser representados ao nível do símbolo. As representações, no entanto, não estão presentes somente nas ontologias. Tudo é uma representação, começando pela própria língua, constituindo-se como uma representação do mundo, do mundo de cada um de nós, embora distinta desse mesmo mundo. O que se afirma na área das ontologias é válido, portanto, também para o sistema linguístico, incluindo o facto de as representações só terem significado quando há um acordo sobre esse mesmo significado.

A língua é naturalmente ambígua. As relações entre as palavras numa língua e as entidades a que se referem são complexas. Muitos termos podem ser usados para nos referirmos à mesma

entidade, e existem muitas entidades para as quais não temos um único termo⁵². Embora uma ontologia seja supostamente uma conceptualização partilhada, é muito difícil para os seres humanos concordarem com a categorização do mundo⁵³.

Sowa (2005) entende que o desejo natural de organizar, classificar, rotular e definir as coisas, acontecimentos e padrões para as nossas vidas quotidianas, é constantemente ultrapassado pela mudança, crescimento, inovação, progresso, evolução, diversidade e entropia inevitáveis. Estas mudanças rápidas são ainda muito mais destrutivas para as frágeis bases de dados e de conhecimento em sistemas informáticos. A informação sobre a qual as pessoas aprendem, raciocinam, agem e comunicam tem uma natureza extremamente fluída.

Por isso, de acordo com Sowa, o maior obstáculo que os sistemas de engenharia do conhecimento têm de enfrentar é a complexidade da mistura heterogénea e muitas vezes inconsistente que existe na nossa mente. Alguma dessa mistura pode ser representada de forma simbólica ou proposicional. Mesmo quando a sintaxe e a semântica de uma frase estão correctas, o ouvinte tem de reconhecer as intenções e expectativas do falante e as suas implicações em termos da situação actual. Todos os aspectos do conhecimento e experiência humanos podem ser necessários para entender uma dada frase, e tem de ser acessível a partir da memória em milésimos de segundo. Além disso, muito, senão a maioria, do conhecimento, pode ser armazenado em padrões não linguísticos de imagens derivadas de qualquer modalidade sensorial. Em poucos anos, as crianças conseguem associar padrões linguísticos com conhecimento de fundo em formas que nenhum computador consegue atingir. Por isso,

⁵² Durante os séculos XVI e XVII vários pensadores preocuparam-se com a tentativa de criar linguagens filosóficas. Estas permitiriam comunicar na perfeição ideias filosóficas e científicas e evitar o carácter vago da língua humana. Um dos expoentes destas ideias foi John Wilkins, Bispo de Chester, já referenciado neste trabalho, que em *An essay towards a Real character and Philosophical Language* publicado em 1668, percebeu que, para criar uma língua perfeita, onde cada 'palavra' se referisse a um único elemento ou ideia, teria de catalogar todo o conhecimento humano e tentou fazer isso escrevendo o seu livro. Este trabalho inspirou mais tarde Roget no famoso Thesaurus, mas foi também um antepassado da representação do conhecimento como a Cyc ou de portais da internet como o Google.

⁵³ Brewster e Wilks consideram que, apesar dos esforços contínuos para automatizar o mais possível, reduzindo o trabalho manual e mantendo as ontologias actualizadas, esta tarefa será impossível de concretizar na íntegra, tornando-se, por vezes, irrelevante. (Brewster & Wilks, 2004)

não dispomos de nenhuma teoria semântica capaz de explicar como podemos compreender a língua natural. Para o autor, a complexidade não deriva do modo como funciona o cérebro humano ou no modo como as línguas naturais designam realidades. Resulta, antes, de questões deixadas à deriva num 'repositório na penumbra' e a dificuldade em reconhecer quais delas são relevantes para a experiência. Estas exceções e casos limite resultam da natureza do mundo, e não de defeitos na língua natural. O discurso individual, embora recorra a palavras e símbolos, nunca conseguirá capturar a totalidade da complexidade de um sistema em contínuo. Kant (1800), comenta Sowa, observou que os conceitos artificiais criados pelas pessoas para um determinado fim são os únicos que podem ser definidos de forma completa. É o caso dos matemáticos.

Apesar de considerarmos a língua natural e o modo como conseguimos expressar proposições em número infinito sobre o mundo por vezes inexcrutável, a verdade é que o registo mais utilizado para a comunicação humana é o discurso verbal. A língua natural contém tudo. O ser humano é que tem extrema dificuldade em representar essa complexidade. Assim, um termo, enquanto unidade que representa conhecimento, tenta aproximar-se da organização conceptual do mundo, através da sua manifestação linguística em discurso.

5.1.2 Análise de dados

Quem parte da língua natural parte, geralmente, de pressupostos de análise lexical e semântica. Tradicionalmente, analisamos esses elementos no texto. O acesso facilitado ao texto fez com que se recorresse à semântica para tentar resolver problemas de acesso ao conceptual.

Neste momento, afirma Reiberger & Daelemans (2005), um entendimento profundo da semântica do texto ainda é impossível. No entanto, temos já à disposição ferramentas que permitem uma análise robusta com níveis elevados de *precision* e *recall*⁵⁴. Assim, a

54

$$\text{Precision} = \frac{\text{Number of Relevant Documents Identified}}{\text{Total Number of Documents Identified}}$$

aprendizagem de ontologias a partir de colecções relevantes de texto tem sido uma das estratégias mais utilizadas, combinando, geralmente, um certo nível de análise linguística com abordagens estatísticas para encontrar potenciais candidatos a conceitos, bem como as relações entre eles⁵⁵. Segundo Hans Hjelm, “(...) given a collection of texts, the ontology learning task typically consists of first recognizing the relevant objects (words or terms) in the text collection and secondly ordering these into a hierarchical inheritance structure.” (Hjelm, 2009, p. 17)

Recordando a definição de Gruber, uma ontologia é uma especificação formal explícita de uma conceptualização partilhada de um domínio de interesse, onde formal implica que a ontologia deve ser lida por máquinas, e partilhada que é aceite por um grupo ou comunidade.

Tem sido prática corrente nos últimos anos a aplicação de uma metodologia designada por aprendizagem de ontologias (“ontology learning”) que se preocupa com a aquisição de conhecimento, mais especificamente com aquisição de conhecimento a partir de texto, contendo aspectos inovadores multidisciplinares que surgem, nomeadamente, com a ajuda da lógica, da filosofia ou do processamento de imagem. Tipicamente, uma abordagem desta natureza envolve primeiro uma extracção de termos a partir de um corpus de domínio através de um processo estatístico que determina a sua relevância para o corpus. A extracção de termos implica níveis mais ou menos avançados de processamento linguístico. Estes termos são depois reunidos em *clusters* com o objectivo de identificar a taxonomia de potenciais classes. (Sintek et al., 2004) (Buitellar, Cimiano, & Magnini, 2005)

$$\text{Recall} = \frac{\text{Number of Relevant Documents identified}}{\text{Number of Relevant Documents in the Collection}}$$

In (Beck & Pinto, 2002, p. 2)

⁵⁵ Alguns exemplos recentes para aprendizagem de ontologias a partir de texto são: TextToOnto, Ontolearn, OntoLT. (Sintek, Buitelaar, & Olejnik, 2004)

Hans Hjelm (2009) sintetiza as técnicas utilizadas na aprendizagem de ontologias a partir de texto, entre elas: i) construção de *clusters* de termos baseados na análise distribucional de Zellig Harris (1954) segundo o qual “Words that occur in similar contexts tend to have similar meanings.”; ii) os pressupostos de Sinclair (1996) sobre o léxico vazio, ou seja, a definição de uma palavra num léxico deve ser vazia na sua fase inicial, sendo que a definição é construída através das observações da palavra nos diferentes contextos (textuais ou orais). O único significado (lexical) que existe é o que é dado a uma palavra através das suas co-ocorrências com outras palavras; iii) informação intra-termo; iv) análise de padrões léxico-sintáticos; v) medidas estatísticas e abordagens probabilísticas; vi) sistemas híbridos.

Autores como Maedche (2002) e Cimiano (2006) apresentam novas propostas para a aprendizagem de ontologias, incluindo análise formal de conceitos, e a combinação de diferentes fontes de conhecimento em sistemas de classificação. A aprendizagem de ontologias requer uma exploração simultânea de várias fontes de informação, onde as fraquezas de umas são remediadas pelas potencialidades de outras.

Para Hjelm, “(...) a ubiquitous raw material for ontology learning is natural language text: in the form of web documents, reports, corporate policies, documentation, periodicals, etc. If we can tap into this wealth of data and use it in the automation process, much will be won.” (Hjelm, 2009, p. 18) A ontologia é acima de tudo uma fonte de informação semântica num sistema e a técnica de aprendizagem de ontologias pretende capturar conhecimento semântico para melhorar as ferramentas de processamento de língua natural. Assumindo que as ontologias formais não fazem parte das suas preocupações, uma vez que não contêm conhecimento lexical, sendo por natureza independentes da língua, Hjelm entende que os termos são os blocos constituintes de uma ontologia. E cita o Swedish Centre for Terminology sobre o que é um termo: “We cannot tell if a word is a term just by looking at it; terms are basically just regular words. But for a word to be called a term, it should be well known – preferably acknowledged – among experts in a specialized field. Further, the experts should agree on what the term stands for.” (Terminologikum TNC in (Hjelm, 2009, p. 5)) Uma ontologia é então “(...) a knowledge-representation structure, where words, terms or concepts are defined by their mutual hierarchical relations (...) In ontologies (...) this is exactly what we find: words (or concepts) defined by their position in a hierarchy of other words and the relations that hold

between them.” (Hjelm, 2009, p. 1 e 13), tendo como objectivo encontrar relações conceptuais entre termos.

Recorrendo a Aristóteles, Lineu e Tim-Berners Lee para exemplificar metodologias de organização do mundo em estruturas hierárquicas, Hjlem refere que os sistemas computacionais têm precisamente o mesmo objectivo. Uma ontologia é um modo de representar conhecimento do mundo real, formalmente, para ser processado por humanos e máquinas. A investigação em processamento de língua natural já concluiu que tem de incorporar semântica para potenciar a reutilização. Um desses recursos é a aprendizagem de ontologias, optimizando tempo e minimizando inconsistências através de análise de grandes quantidades de texto de especialidade e da criação de ferramentas que vão treinar classificadores automatizados para reconhecer relações léxico-semânticas. Seguindo os princípios de Yorick Wilks, não podemos transmitir o significado de uma palavra sem recorrermos a outra palavra para o definir: “Meaning. . . is best thought of as other words.” (Hjelm, 2009, p. 13)

Reinberger et al (2005) exploram uma abordagem semelhante. A extracção de relações ontológicas do texto permite uma filtragem de frequência quando estão disponíveis grandes quantidades de textos. Os termos que têm relações sintácticas similares com outros termos estão semanticamente relacionados. Usando as técnicas de *clustering*, estas semelhanças semânticas ou dependências semânticas podem ser usadas para agrupar os termos em classes, fornecendo termos que sejam possíveis candidatos para estender a ontologia inicial, ou para criar uma de raiz. A afinação da técnica de *pattern-matching* pode facilitar em grande medida uma evolução na extracção não supervisionada de conhecimento ontológico.

Bourigault e Gilles (2003) defendem a aplicação do que designam por recursos termino-ontológicos a partir de texto que sejam capazes de descrever os termos e os conceitos de um domínio. Encarando a automatização com suspeita, os autores consideram que, face às necessidades e às aplicações disponíveis, os intervenientes são levadas a considerar se, por razões de pertinência e de eficácia, os recursos lexicais e/ou conceptuais que devem construir e explorar podem ou devem ser construídos a partir de fontes textuais. São por isso conduzidos naturalmente a proceder a uma aproximação interdisciplinar, onde o tratamento automático da língua pode ser o catalisador e fornecedor de métodos e instrumentos de

construção de ontologias a partir de textos. No caso da construção de uma ontologia para um sistema informático, cujos dados de entrada não serão textuais mas numéricos, a análise de texto permanece primordial, servindo de indicador da organização do sistema conceptual, uma vez que se baseia numa concepção funcional do termo e do conceito, em oposição a uma postura estática e clássica que encararia o termo como simples etiqueta do conceito.

Bourigault e Gilles entendem que esta postura é adoptada frequentemente pela inteligência artificial e por metodologias de processamento automático da língua, disciplinas influenciadas desde há longo tempo por uma semiótica do signo fundada sob a tríade termo/conceito/referente. A concepção clássica defenderia que o termo existe enquanto representante linguístico de um conceito que faz parte de um sistema conceptual único e estável que caracteriza o domínio *a priori*. Mas a constatação da variabilidade das terminologias impõe-se : num dado domínio de actividade, não existe *uma* terminologia, que representaria o saber do domínio, mas diversos recursos termino-ontológicos, bem como aplicações onde esses recursos são utilizados. Assim, podemos encarar o termo e o conceito como o resultado de um processo de análise termino-conceptual. O analista procede a um trabalho de construção de um recurso termino-ontológico para uma determinada aplicação num determinado domínio, e não a uma descoberta da terminologia de um domínio.

Para Bourigault e Gilles não se trata de preocupações com ontologias genéricas representando um conjunto máximo de conhecimento do senso comum, nem com ontologias formais no sentido de Guarino, constituindo um quadro referencial universal e formalmente válido. A solução do problema de aquisição de recursos termino-ontológicos a partir de corpus não reside unicamente no fornecimento de uma ou mais ferramentas de tratamento automático da língua. A disponibilização de tais ferramentas deve fazer-se acompanhar de uma reflexão metodológica, conducente à realização de guias e plataformas informáticas integradoras que permitam uma utilização eficaz das ferramentas. Idealmente, o o analista deve ter conhecimentos e competências técnicas na área linguística e computacional⁵⁶.

⁵⁶ A ferramenta proposta pelos autores é o TERMINAE, que se apoia na investigação da escola francesa onde convergem terminologia, linguística, engenharia do conhecimento e inteligência artificial. Baseia-se nos seguintes princípios : i) Partir de textos do domínio como fonte de conhecimento : constituem um suporte tangível, reunindo conhecimento estabilizado que serve de referência e

Pelos motivos expressos, os autores entendem que os recursos não devem ser universais, mas sim de domínio e mesmo específicos de tarefa, baseados num processo sólido de selecção de textos. A supervisão humana é, contudo, sempre necessária em todas as fases: construção do corpus, integração de critérios de decisão e diferenciação, bem como definição das propriedades formais da ontologia. Essa decisão humana resulta de um compromisso entre a informação no texto, o conhecimento especializado do domínio e o conhecimento de carácter geral. Como os conceitos numa ontologia são o resultado de um processo de construção onde somente propriedades relevantes dos objectos e ideias seleccionadas do mundo são representadas, esta construção corresponde a uma selecção de acordo com um ponto de vista do domínio de conhecimento. Assim, a ontologia é construída, tendo em mente o modo como o sistema irá utilizá-la e as necessidades dos utentes. Responder a perguntas automaticamente requer ser capaz de encontrar automaticamente as respostas em documentos, o que não é possível na maioria das vezes. A resposta pode ser fruto de negociações entre especialistas ou um compromisso entre várias fontes textuais. Assim, a ontologia é uma construção consensual sobre o domínio, adaptada à aplicação. Pela sua natureza incompleta, o texto não deve constituir-se como o único recurso. (Aussenac-Gilles, 2005)

Rokia Bendaoud et al (2007), reconhecendo que a maior dificuldade está em adquirir conhecimento (nomeadamente dos especialistas) e posteriormente manter esse conhecimento num dado domínio, propõem a construção de ontologias baseada em texto através de análise formal de relações conceptuais, semi-automaticamente. Trata-se de um método algébrico para a derivação de uma hierarquia conceptual baseado na 'metontologia'⁵⁷, uma metodologia semi-automática que constrói uma ontologia a partir de um conjunto de termos extraídos de recursos. A este propósito, Cabré et al (2005) mencionam também o que

melhora a qualidade do modelo final ; ii) Enriquecer o modelo conceptual de um componente linguístico : o acesso aos termos e aos textos que justifiquem a definição dos conceitos garante uma melhor compreensão do modelo ; iii) Utilizar as técnicas e ferramentas de TAL baseadas nos trabalhos linguísticos : permitem a exploração sistemática dos textos e os seus resultados facilitam a modelação ; iv) construir ontologias 'regionais', ou seja, consensuais dentro de um domínio e adaptadas a uma aplicação, mas não universais ; v) Aplicar os princípios sistemáticos de modelação para assegurar uma boa estruturação dos dados e facilitar a manutenção da ontologia. A ferramenta CAMÉLÉON é complementar e serve para a identificação de relações semânticas no texto. (Bourigault & Aussenac-Gilles, 2003) (Aussenac-Gilles, 2005)

⁵⁷ 'Methontology'. (Gómez-Pérez et al., 2004)

designam por 'ontologias locais'⁵⁸, onde o conhecimento é considerado como dependente do domínio e da aplicação, construídas maioritariamente a partir de textos, nomeadamente em suporte electrónico, e não tanto com a colaboração de especialistas.

Partindo dos pressupostos expostos pelos autores citados, interessa analisar as suas vantagens e fragilidades. A principal fragilidade, a nosso ver, reside na inconsistência das reflexões teóricas. Nesta dissertação, assumimos que extrair conhecimento de texto é uma tarefa extremamente difícil de concretizar. Assumimos, também, que extrair ontologias a partir de texto é algo incompatível com o próprio conceito de ontologia. O conhecimento é abstracto, cognitivo. Uma ontologia é um artefacto, uma construção, produto de um acordo consensual dos membros de uma comunidade. O texto é um recurso, fundamental em alguns casos, dispensável noutros. Texto, contudo, não é igual a terminologia.

Nesse sentido, tendo em conta a definição de ontologia para um sistema computacional e o facto de estarmos a analisar o papel da terminologia na construção desse artefacto de engenharia, faz com que tenhamos de contestar afirmações que encaram como possível, através de análise linguística e estatística, automatizar a extracção de relações ontológicas de texto, e, por extensão, a construção de ontologias directa e automaticamente a partir de texto.

Afirmações como as de Buitellar et al, segundo as quais os objectivos da aquisição de conhecimento coincidem com a aprendizagem de ontologias a partir de texto, não oferecem consistência. O que se pretenderia com a aprendizagem de ontologias a partir de texto, segundo os autores, seria adquirir conhecimento explícito implicitamente contido nos dados textuais. A extracção por métodos estatísticos de um ou vários conjuntos de termos traduziria conhecimento. Mas como podemos adquirir explicitamente o que está implícito? O especialista seria o único a ter acesso, através da leitura do texto, ao conhecimento implícito.

⁵⁸ Por oposição a 'ontologias gerais', onde a linguagem é considerada uma ferramenta comunicativa, partilhada por todos os falantes e preocupada com o conhecimento comum (Guarino 1995) *em* (Cabré et al., 2005).

Não é possível, no estágio evolutivo actual dos sistemas de informação, capturar automaticamente conhecimento implícito através de análise textual. Por outro lado, como já tivemos oportunidade de comentar, a informação semântica retirada do texto, quer seja através da análise linguística, quer seja através de métodos estatísticos e distribucionais, não equivale a conhecimento, nem reflecte de forma sólida um sistema conceptual. Daí o comentário de Rokia Bendaoud de que a dificuldade se encontra na aquisição de conhecimento por parte do especialista e na manutenção desse conhecimento. Pensamos que é uma ideia aliciante poder dispensar o especialista, mas os resultados poderão não corresponder a ontologias no verdadeiro sentido do termo para a inteligência artificial: um artefacto de engenharia para partilha e reutilização do conhecimento.

Da mesma forma, a abordagem de Hans Hjelm levanta muitas dúvidas. Começemos pela definição de ontologia: “(...) an ontology is a knowledge-representation structure, where words, terms or concepts are defined by their mutual hierarchical relations (...)” (Hjelm, 2009, p. 5), servindo para encontrar relações conceptuais entre termos. O autor coloca no mesmo nível palavras, termos e conceitos. Uma ontologia, para propósitos de manipulação formal, interessa-se pelos conceitos e relações entre conceitos. Por isso não podemos afirmar que serve para encontrar relações conceptuais entre termos. As relações entre termos são relações semânticas, que se observam em discurso. Se a aprendizagem de ontologias parte do texto, ela vai tentar encontrar relações semânticas entre termos, manifestações linguísticas de conceitos. Os conceitos não se encontram no texto. Concordamos com a afirmação de que a automatização de processos de análise de texto vai reconhecer relações léxico-semânticas, mas não conceptuais. Igual-se representação de conhecimento a uma organização da terminologia de um domínio, pois Hjelm valoriza as ‘ontologias terminológicas’⁵⁹, estruturadas como as ontologias formais, mas centrando-se na organização de termos ou grupos de termos, e não conceitos, sendo muitas vezes referidas como taxonomias. Trata-se, na nossa opinião, de uma definição pouco clara de ontologia, principalmente tratando-se de exploração de recursos relacionados com a construção de ferramentas para partilha e reutilização de conhecimento.

⁵⁹ Hjelm contrapõe as ontologias terminológicas com as ontologias formais (comuns na inteligência artificial e representação do conhecimento e que se preocupam com os conceitos e as relações entre conceitos) e com as ontologias baseadas no protótipo que aparecem como o resultado da aplicação de uma técnica de *clustering* hierárquica a dados distribucionais. (Hjelm, 2009)

Uma taxonomia não é uma ontologia, embora uma ontologia possa conter uma taxonomia. Uma ‘ontologia terminológica’ – designação com a qual não concordamos – seria, segundo o autor, uma fonte de informação semântica. Aqui estamos de acordo: obtenção de informação semântica através de texto, enquanto recurso inicial na construção de ontologias. Para nós, conceito e significado do termo são duas coisas diferentes, mas complementares, devendo ser claramente encaradas como tal.

Ainda na análise das afirmações de Hjelm, o autor entende que as ontologias formais não fazem parte das suas preocupações porque não contêm conhecimento lexical, uma vez que o conhecimento formal é independente da língua. Considerando os termos como “building blocks of the ontology”, assume uma postura marcadamente semasiológica. Quando observarmos neste capítulo a abordagem extralinguística, veremos que os argumentos de Hjelm se encontram numa posição antagónica à construção de ontologias enquanto recursos formais, idealmente independentes de qualquer língua natural. A metodologia que parte do linguístico, do texto, como a de Hjelm, enquadra-se na perfeição na visão do termo enquanto unidade de conhecimento poliédrica, cujos traços são activados em função do contexto discursivo. Mas o autor peca por, pelo menos aparentemente, desvalorizar a importância do conceito, partindo do princípio de que se pode extrapolar para a extracção de uma eventual organização conceptual a partir de dados exclusivamente semânticos. Do nosso ponto de vista, devemos encarar as metodologias de extracção de dados semânticos – para nós, é nisto que consistem as técnicas de ‘aprendizagem de ontologias’ – como um pré-estágio da ontologia. Esse pré-estágio semântico, revelando as relações entre termos, pode ser extremamente importante na construção da ontologia.

A designação de recursos termino-ontológicos por parte de Bourigault e Gilles encontra-se, a nosso ver, dentro da orientação da necessidade de práticas colaborativas na representação do conhecimento, embora possa suscitar dúvidas quanto à adequação das metodologias utilizadas para partilha e reutilização de conhecimento formalizado. Concordamos que o tratamento automático da língua pode ser o catalisador da descoberta de dados semânticos pertinentes, servindo de indicador da eventual organização do sistema conceptual, mas não de métodos de construção de ontologias a partir de texto. A abordagem ‘funcional’ do termo e do conceito reflecte, também, a metodologia de análise semasiológica, rejeitando uma concepção

estática dos conceitos como pertencentes a sistemas estáveis, caracterizados *a priori*, antes da análise discursiva. Esta aceitação, desejável até, da variabilidade das terminologias, que suplantam a simples função de etiquetas linguísticas, vai, por outro lado, dar razão à importância do especialista na construção da ontologia – aliada à necessidade de conhecimentos linguísticos e computacionais – e ao peso do extralinguístico, que, como o próprio nome indica, não provém do processamento de texto em língua natural. Dada a complexidade da representação do conhecimento, pensamos que os autores têm razão ao afirmarem que os recursos terminológicos e ontológicos não devem ser universais, mas sim de domínio ou mesmo específicos de tarefa. Discordamos com a afirmação de que esses recursos sejam baseados numa selecção rigorosa de textos. Tal como os próprios autores afirmam, a resposta automática a questões requer a capacidade de encontrar automaticamente respostas em documentos, o que não é possível na maioria das vezes.

5.1.3 Relação com o texto

A exploração de recursos em língua natural no processo de representação formal do conhecimento consiste, para os engenheiros de inteligência artificial, no seu maior desafio. Voltando a Hans Hjelm, o autor questiona: “When we learn an ontology from a text, what is it that the final model captures? Is it ‘reality’, a model of the world around us? Or maybe the representation of the world, as it appears in the writer’s head? Perhaps an amalgamation of several representations, if the text has been written by several writers?”(Hjelm, 2009, p. 24) Como vimos, as metodologias que partem da análise linguística têm como pressuposto a possibilidade de extrair conhecimento a partir de texto. Como uma ontologia pretende representar conhecimento, concluir-se-ia, por extensão, que é possível extrair ontologias a partir de texto. Tem-se verificado, por isso, uma preocupação crescente com a análise de discursos de especialidade através de corpora electrónicos, nomeadamente com a construção de recursos e ferramentas especificamente desenvolvidos para o efeito. Aparentemente pertencentes à esfera da crítica literária, questões como ‘quais os pressupostos do autor e do leitor’ começam a fazer parte dos dados a ter em conta na construção das conceptualizações, já que o sentido, baseado numa análise léxico-semântica, é construído dinamicamente em contexto.

Uma ontologia, enquanto artefacto de representação do conhecimento, é um modelo abstracto daquilo que individual ou colectivamente acreditamos ser ‘verdade’ sobre o mundo e os seus intervenientes. Uma das fontes privilegiadas de procura de conhecimento seria as manifestações discursivas do pensamento humano, ou sejam, os textos. Podemos analisar textos num computador, manipulá-los e extrair determinado tipo de informação, com a vantagem de dispormos de mais de uma geração de investigação de análise de texto e de aplicação e exploração de métodos de linguística computacional. Existem corpora específicos com grandes quantidades de texto que podem ser acedidos através da Internet, embora a sua fiabilidade possa ser questionável, uma vez que não é possível confirmar as fontes e testar a qualidade dos textos disponíveis. Além disso, a ambiguidade natural da língua, independentemente do modo de transmissão, aliada à mutação constante do conhecimento que tentamos modelar, faz com que a representação de uma conceptualização muitas vezes se torne obsoleta antes mesmo da sua concretização formal.

Para Brewster et al (2005), coloca-se aqui um dos maiores problemas para a representação do conhecimento. Dado que não conseguimos ler a mente humana, e uma vez que o texto seria uma das principais fontes de pesquisa de possível extracção de conhecimento, verificamos que uma grande parte do que precisamos identificar não se encontra explícito. Encarando o artefacto textual como algo em constante evolução, cada autor assume um conjunto partilhado de conceitos, referências ou ideias para que a comunicação se estabeleça: “In writing a text, each and every author assume a shared set of concepts, ideas and terms of reference. This is inevitable otherwise communication could not arise. A given text interacts in specific ways with this assumed ontology.” (Brewster et al., 2005, p. 2)

Segundo Brewster, esta situação tem consequências. Em primeiro lugar, o texto reforça as assumpções desse conhecimento subjacente ao dizer ao leitor qual a ontologia a utilizar para processar o texto. Em segundo lugar, o texto altera as associações e instanciações de conceitos que já estão presentes nessa ontologia. Assim, um primeiro objectivo do texto a alguns níveis é alterar a relação entre os conceitos existentes, ou alterar as instanciações desses conceitos. Uma forma do texto fornecer ‘nova’ informação ao leitor é confirmar uma relação anteriormente ausente ou apagar uma relação anteriormente assumida. Este tipo de actividade pode ser vista como tentando reestruturar a ontologia de domínio que é

claramente outra forma de manutenção de conhecimento. O autor pode propor um novo conceito ou nomear um novo procedimento, e estes actos de nomeação rotulam novos conceitos e indicam a sua relação com o resto das conceptualizações. Finalmente, um texto pode afectar a ontologia de domínio adicionando novos conceitos à ontologia de domínio já existente:

“(...) in each case, however, a text takes the background knowledge, in effect the background ontology, for granted. Texts rarely make explicit statements concerning the ontology. In Quinean terms, texts alter our knowledge of the periphery; they do not make explicit statements about the core of knowledge. This may seem strange but it is inevitable given the nature of communication. (Brewster et al., 2005, p. 2)

Segundo o autor, a engenharia do conhecimento iria necessariamente concluir que não se pode esperar extrair a ontologia subjacente de um corpus de textos específicos de um domínio porque, a cada passo, a ontologia é assumida. É necessário, por isso, encontrar técnicas para apreender as formas tácitas como o conhecimento está expresso e estabelecer métodos coerentes para sair do corpus e identificar o conhecimento que falta.

De acordo com as orientações já traçadas nesta dissertação, não podemos deixar de comentar as considerações acima expostas. Em primeiro lugar, um texto não contém uma ontologia subjacente. Isso seria contradizer os princípios de uma ontologia computacional. Mesmo que o autor se refira a uma ontologia no sentido metafórico, que é certamente o que Brewster pretende dizer, entendemos que devemos referir-nos antes a conceptualizações. O conjunto das conceptualizações pode ser representado e estabilizado através de uma ontologia. Essas conceptualizações podem manifestar-se no texto ou em outros recursos expressivos, discursivos ou não. Um texto tem, por isso, utilidade em função do que se procura e deve ser encarado como um elemento operacional.

Em segundo lugar, como já referimos, temos sérias dúvidas relativamente ao facto de conseguirmos extrair conhecimento a partir de texto. Extraí-se termos, entidades linguísticas, que são uma manifestação do conhecimento, e relações entre termos, através, nomeadamente, de reformulações. O conhecimento é demasiado complexo. A informação

contida no texto não pode, na grande maioria das vezes, ser entendida como conhecimento, uma vez que, como já vimos, o conhecimento raramente está explícito. No texto temos vocabulário e uma estrutura textual composta por morfologia e sintaxe. Essa estrutura tem um sentido em função do contexto desse determinado texto, dispondo de uma semântica. Essa semântica reflectida no nível linguístico pode ajudar na estrutura conceptual do extralinguístico, ou seja, do ontológico.

No limite, uma ontologia não equivale a conhecimento. Representa uma estrutura possível do conhecimento que, interligada com muitas outras possíveis estruturas conceptuais, pode aproximar-se do conhecimento de um domínio de especialidade. Alcançaremos algum dia o implícito, o que não é verbalizado? Dependerá, em grande parte, da evolução das linguagens formais. A complexidade do conhecimento será sempre uma constante, embora possamos evoluir na forma da sua transmissão. É o conteúdo da representação, ou seja, o conjunto de conceitos escolhidos e a sua inter-relação, que fornece uma perspectiva particular do mundo. O problema não está, assim, na língua, está no sujeito que a utiliza, que vai forçosamente seleccionar um vocabulário e estruturar o discurso de forma diferente.

Quando um autor cria um texto assume um certo número de pressupostos. Existe uma assumpção linguística ligada à língua utilizada e uma assumpção cognitiva ligada à capacidade da audiência de compreender o vocabulário e os termos utilizados. De facto, um autor assume que o público a quem se dirige partilha do mesmo tipo de conceptualizações. Um texto constitui, por isso, um acto de manutenção do conhecimento. Torna-se extremamente complexo construir processos computacionais que consigam capturar a essência do que não está no texto. A situação mais expectável para encontrar manifestações linguísticas de conhecimento base (ou seja, a especificação explícita de relações entre termos) seria o texto científico ou académico, tentativa prototípica de alterar as conceptualizações estabelecidas pela comunidade. Acontece que um texto de especialidade não transmite necessariamente conhecimento organizado para o leitor. Além disso, há no texto de especialidade uma forte componente de conhecimento implícito (Costa & Silva, 2009) sendo que o especialista muitas vezes não explicita as relações semânticas no texto, dificultando a organização conceptual do domínio. Poderíamos esperar que quando um conceito ou a sua lexicalização correspondente é introduzida pela primeira vez, surgissem definições no texto, ou quando um conceito fosse

emprestado de outra disciplina, existisse uma definição do termo, mas tal não acontece com a frequência desejada. Apesar disso, o texto de especialidade representa uma construção dinâmica através da contínua actualização das conceptualizações, sendo auto-regulado pela comunidade científica, valorizando o papel fundamental do especialista. Pensamos, contudo, que o texto enquanto elemento operacional, permanece um elo importante na ligação entre a língua e o pensamento. Apesar de o texto ser por natureza incompleto, essa condição não é necessariamente má para a construção de ontologias computacionais.

Rastier, crítico contundente de uma abordagem exclusivamente extralinguística, chama a atenção para o perigo de reduzirmos a semântica do texto a fórmulas lógicas. Ainda que todas as proposições que compõem uma ontologia sejam declarativas, no sentido de serem verdadeiras ou falsas, os textos, para uma ontologia, não se privam de serem contraditórios, uma vez que no discurso, mesmo de especialidade, existem vários pontos de vista que não são descrições equivalentes da mesma coisa⁶⁰. Um signo isolado das suas condições de interpretação, ou seja, do seu contexto, resume-se a um artefacto normativo da prática ontológica. Só que o sentido de uma palavra não lhe é imanente, é antes o produto de uma interpretação : « Si les textes techniques et scientifiques utilisent préférentiellement certaines lexicalisations synthétiques, la correspondance entre les unités microsémantiques (dont les termes comme sémèmes construits) et les unités macrosémantiques (comme les acteurs ou les fonctions dialectiques) n'est jamais univoque : la différence des paliers de complexité interdit de considérer que le texte se résume aux mots qui le composent. » (Rastier, 1995, p. 16) E questiona : A terminologia tem por objecto os sistemas conceptuais ou as estruturas lexicais? Será a terminologia forçada a optar entre estudar uma hierarquia presente numa ontologia, ou estudar a interacção dos significados no texto?

Pensamos que não. A terminologia não tem de optar entre o estudo dos termos e relações entre termos ou o estudo dos conceitos e relações entre conceitos. Deve, sim, a nosso ver,

⁶⁰ O autor utiliza neste artigo a expressão '*purificação da língua*', defendendo a ideia já referenciada de que as ontologias contemporâneas foram criadas pelo positivismo lógico do Círculo de Viena, cujo objectivo era purgar as línguas dos seus defeitos, inscrevendo-se na continuidade das pesquisas sobre as línguas perfeitas, base das linguagens formais. (Rastier, 2004)

distinguir claramente as suas práticas e objectivos de acordo com o propósito da aplicação dos dados. A pesquisa e observação dos termos nos textos podem combinar as abordagens semasiológica e onomasiológica. A abordagem onomasiológica sozinha pode não ser linguisticamente suficiente. Para um engenheiro do conhecimento, talvez sim. Para um linguista, dificilmente o será. Nesse sentido, pensamos que seria aconselhável que as metodologias para a construção de ontologias que se baseiam no processamento de texto tivessem em consideração uma aproximação conceptual no início da análise de dados, observando o modo como o especialista gera e gere o seu discurso em contexto.

Rogers (2004), comentando a terminologia sociocognitiva proposta por Temmerman, refere que a abordagem onomasiológica de identificação de conceitos, resultado de uma classificação mental e procedimentos de ordenação, resulta na prática também de uma análise de textos. A abordagem linguística tem tido dificuldade em distinguir a análise conceptual da análise semântica, e, numa tentativa de tornar explícitas para a máquina determinadas suposições, tem vindo a desenvolver, por exemplo, trabalhos sobre o reconhecimento de “knowledge hot-spots” através de pistas linguísticas baseadas em padrões lexicais. A autora considera que a maior parte dos investigadores nesta matéria, incluindo os defensores da abordagem onomasiológica, concordam que os textos constituem uma parte importante da análise terminológica e ontológica, tanto para fins teóricos como práticos.

A questão que se coloca relativamente às ontologias é se estes sistemas constituem um primeiro passo para algo muito mais abrangente ou nunca irão passar de domínios muito limitados. A utilidade das ontologias é questionada, muitas vezes, pela baixa complexidade e resultados pouco expressivos que apresenta, comparativamente ao esforço empregue na sua construção e implementação. No entanto, sabemos que a procura da perfeição faz parte da nossa natureza e nos últimos cinquenta anos assistimos a uma evolução tecnológica maior do que em toda a história da humanidade. Permanece, por isso, um desafio, conseguirmos formalizar conceptualizações através da análise de texto.

5.2 Perspectiva extralinguística

5.2.1 Complexidades da linguagem formal

A grande dificuldade da inteligência artificial é lidar com a língua natural. Para tentar solucionar esse problema tem de optar, basicamente, entre dois caminhos: ou aceita essa dificuldade, valorizando a riqueza do discurso verbal, e tenta otimizar ferramentas que consigam aproximar-se da semântica do texto e que, por conseguinte, transmitam, de alguma forma, a rede de conceptualizações, ou evita o confronto com a natureza fluida do discurso verbal e concentra-se directamente nos conceitos organizados pelo especialista de domínio, trabalhando, essencialmente, com a terminologia fora do texto de especialidade.

Para a inteligência artificial, o que existe é o que pode ser verbalizado. Não podemos modelar e formalizar o que não podemos ‘dizer’, e ‘dizer’ é diferente de ‘pensar’. Sabemos que a língua natural é um meio redundante, que permite o uso de mais do que um sinal para referir o mesmo objecto no mundo. Um termo, que pode ser uma frase nominal simples ou complexa, pode referir-se a mais do que um conceito, ou vice-versa. Ou seja, a ontologia de conceitos que representa o mundo por um lado e os termos na língua natural por outro, não têm uma relação de um para um, mas sim de muitos para muitos: “(...) the ontology of concepts that represent the world on the one hand and the terms in the language on the other do not have a one-to-one relationship, but are rather distributed many to many.” (Buitelaar, 2000, p. 2)

O conceito de discurso é, neste contexto, perfeitamente distinto da linguística, como esclarece Gruber:

“For AI systems, what ‘exists’ is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse (...) In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms.” (Gruber, 1993a, p. 1)

As ontologias descrevem, através de uma linguagem de representação, compromissos ontológicos para um determinado conjunto de agentes de tal forma que possam comunicar sobre um domínio do universo do discurso em ambiente de partilha. O agente ‘sabe’ se agir *como se fosse* possuidor da informação no sentido de alcançar determinado objectivo, numa base de pergunta-resposta.

O texto, pelo contrário, é espaço de polissemia, imprecisão, variação e ambiguidade (Costa, 2006). Só que grande parte dos conceitos e das suas relações são unicamente intelectualizáveis e verbalizáveis por via de enunciados linguísticos, não sendo possível, muitas vezes, formalizá-los. As representações formais das conceptualizações requerem, tanto na abordagem de base linguística como extralinguística, o apoio dos especialistas de domínio para construir e gerir gradualmente versões cada vez mais complicadas das relações conceptuais convergentes e divergentes. Como sublinha Kocourek (1991), os textos vão fazer surgir outro aspecto fundamental dos termos: a sua brevidade. O avanço do saber – a por vezes a sua regressão – reflecte-se na terminologia. Os termos são concretizados nos textos, e o sistema terminológico adapta-se, modifica-se e refaz-se constantemente.

Recordando os comentários de Brewster, um texto é um acto de manutenção do conhecimento, no sentido em que reforça as assumpções de base, altera ligações e associações e adiciona novos conceitos. Isto significa que o conhecimento de base é raramente expresso de uma forma interpretável para a máquina:

“An inherent assumption of all the authors in this field is that ‘concepts’ are the key building blocks⁶¹, and we manipulate concepts with words. All ontologies I have encountered use words to represent the concepts and to mediate or provide a correspondence with the external world. Consequently a large range of items in the world or experiences which do not lend themselves readily to verbal expression cannot be modelled. We could describe this as the ‘Ontological Whorf-Sapir Hypothesis’ i.e. that

⁶¹ Repare-se no contraste com a posição de Hans Hjelm quando defende a captura de ontologias através de processamento de texto: “(...) terms are the building blocks of the ontology.” (Hjelm, 2009, p. 27)

that which cannot be captured by words cannot be represented in an ontology.”(Brewster et al., 2005, p. 4)

Tendo em conta a citação acima transcrita, compete-nos reforçar que, do nosso ponto de vista, uma ontologia não lida com palavras. Lida com conceitos e relações entre conceitos. As manifestações linguísticas dos conceitos são os termos que podemos encontrar no discurso verbal ou não. Um termo não é necessariamente uma entidade lexical e a expressão verbal não é o único meio de transmissão de conhecimento. Não podemos por isso reduzir a procura de conhecimento ao texto. Não esqueçamos que uma ontologia, encarada como artefacto de engenharia com o propósito de partilha e reutilização de conhecimento, tem sentido fora da língua, não tem sentido fora de uma linguagem de representação.

A transformação de conhecimento dinâmico e prático em conhecimento armazenável e estático tem sido uma tarefa central na gestão do conhecimento. Um dos maiores problemas da inteligência artificial está relacionado com o facto de lidar com informação estruturada de forma heterogénea. As ontologias tentariam mitigar este efeito. Tal como os seres humanos, aplicam critérios de selecção e técnicas de reorganização dos dados oferecidos pela realidade adaptando o conhecimento a novas situações, sendo que os domínios a formalizar devem ser objecto de uma estrutura conceptual relacional. A representação formal e simbólica do conhecimento repousa sobre as linguagens formais compostas por regras e padrões associados a fórmulas. Uma linguagem formal deve ter uma semântica formal para precisar que significações associar às fórmulas em função das significações dos símbolos que contêm e da maneira como são associados.

É por isso necessário entender que ao aceitarmos uma ontologia, aceitamos o conjunto de compromissos ontológicos consensualmente seleccionados pela comunidade, sem pretensões de universalidade, mas com resultados eventualmente eficazes na representação formal de fragmentos de conhecimento que possam ser partilhados e reutilizados entre sistemas e entre pessoas. Sem esta condição, não temos ontologia. Temos outra coisa, porventura útil, mas, ainda assim, diferente.

As ontologias devem co-evoluir com as suas comunidades de uso e a interpretação humana do contexto linguístico e extralinguístico no uso e desambiguação de uma ontologia desempenha

um papel importante, melhorando a qualidade e legitimidade de colaborações orientadas para objectivos nas comunidades. É no contexto social que a conceptualização faz sentido, é aí que tem lugar a sua génese, desenvolvimento e extinção. Posso ‘conhecer’ um texto do ponto de vista linguístico e discursivo sem ter qualquer conhecimento sobre o domínio. Por isso as estruturas linguísticas não encerram, necessariamente, estruturas conceptuais.

A terminologia constitui, neste processo, uma peça chave, não só porque termo e conceito são, na nossa perspectiva, duas entidades indissociáveis, mas também porque o conhecimento necessita de um veículo de representação. A terminologia desempenha esse papel, intervindo essencialmente nas primeiras etapas de construção das ontologias através da ‘especificação informal da conceptualização’.

5.2.2 Análise de dados

Enquanto a perspectiva linguística parte do discurso verbal na forma de texto, a perspectiva extralinguística parte do especialista enquanto depositário de conhecimento especializado, partilhando de uma visão muito particular da terminologia e da língua natural, que optamos por explorar aqui através das ideias expressas por Christophe Roche⁶², investigador na área da inteligência artificial e ontologias, cuja abordagem teórica e metodológica vamos passar a expor.

Para Roche, o conhecimento é como uma esfera: delimita a estrutura do que sabemos e do que não sabemos. Baseado na lógica de Aristóteles apresentada em *Organon* e nas teorias de Kant, entende que existem vários tipos de conhecimento: conhecimento terminológico e ontológico, que é necessariamente estático (*Eu tenho um livro*); conhecimento sobre o estado do mundo, que é mutável e não pode ser representado pelas ontologias, mas pelas relações com os objectos (*O meu livro está na mesa*); conhecimento inferencial (*Se tenho o meu computador ligado, posso aceder à internet*). Criticam-se as ontologias por modelarem

⁶² A análise da perspectiva de Christophe Roche centra-se nos artigos que a seguir se enumeram. Como vamos observar o ponto de vista do autor de uma forma global, referenciaremos apenas o artigo no caso de uma citação. Artigos por ordem cronológica: (Roche, 2003a); (Roche, 2003); (Roche, 2005); (Roche, 2006); (Roche, 2006a); (Tricot & Roche, 2006); (Roche, 2007); (Roche, 2007a, 2008, 2009)

conhecimento estático, mas uma ontologia não tem por objectivo representar todo o conhecimento, somente aquele conhecimento que permite descrever os objectos do mundo. A partir do momento em que se verifica uma ruptura epistemológica ou a entrada de novos conceitos, há uma mudança na teoria, por isso não é contraditório afirmarmos que temos certo tipo de conhecimento estático num dado momento.

Devemos modelar o mundo que queremos conceptualizar em função de princípios epistemológicos que são subentendidos, mas também através de representações que sejam reutilizáveis. O princípio de que a organização do saber está ligada à língua e de que, por consequência, conceptualizamos em função das bases epistemológicas da língua, está errado. O conhecimento não se confunde com o discurso nem com o objecto, sendo que a lógica ontológica tem de ser percebida no sentido matemático. Para um engenheiro, o conhecimento científico é um conjunto de signos que interpretamos e conhecemos as regras de manipulação sem ambiguidade, cujo sentido é claramente definido, sem interpretação.

Segundo Roche, a ontologia provém do extralinguístico, por isso o conhecimento que expressa deve ser independente, estável, idealmente invariável, preocupando-se com conceptualizações baseadas em linguagens formais, e não em línguas naturais. Mesmo em ambiente controlado, a língua natural não revela a natureza ontológica das coisas, sendo que a variabilidade do significado não permite delimitar a estabilidade dos conceitos, não sendo por isso adequada para representações formais nem para transmissão de conhecimento: “Natural language is not a suitable language for specifying conceptualization: it is not its aim; as well as the main goal of writing text is not to define ontology.” (Roche, 2007a, p. 6) Mesmo aplicando as técnicas de análise da semântica distribucional, vamos observar coisas interessantes, mas não extralinguísticas. Se utilizássemos a língua natural no contexto das ontologias, as críticas à perspectiva semasiológica da terminologia aplicar-se-iam igualmente ao resultado de uma tal abordagem. O texto contém conhecimento linguístico, onde utilizamos palavras para falarmos sobre o mundo, sendo a semântica uma teoria que estuda as relações entre os signos. Esse conhecimento necessita de um agente cognitivo para o interpretar e uma ontologia não necessita do agente cognitivo.

Na perspectiva de Roche, a terminologia enquanto ciência e disciplina autónoma mobiliza para os seus objectivos a linguística, a lógica e a epistemologia, com uma forte intersecção com as

línguas de especialidade. Constitui-se como uma ‘prática teorizada’, a ‘ciência dos termos’ (Roche, 2005), partilhada com a linguística – de quem se distingue pela modelação científica das palavras – e com a epistemologia – com quem partilha a procura de uma compreensão do mundo.⁶³ As ciências cognitivas, designadas por Roche de ‘ciências moles’, não derivam da terminologia e são válidas unicamente para domínios não científicos, onde questões relacionadas com a verdade, por exemplo, podem ser colocadas. Conceitos claramente definidos, com base aristotélica, são aplicados nas ‘ciências duras’. Roche admite que existem diferentes concepções de terminologia de acordo com o domínio de aplicação, como a abordagem linguística, embora considere que, mesmo que o termo seja uma entidade lexical e que denote um conceito, não temos conceitos no texto, e todas as abordagens da terminologia se baseiam em conceitos.

Assumindo uma postura próxima das teorias de Wüster, pretende mostrar a importância da lógica e da epistemologia na terminologia. A terminologia de base wüsteriana e a engenharia do conhecimento partilham o princípio da procura de conhecimento estabilizado, e, sendo a terminologia oriunda dos domínios científicos, partilha igualmente com a ontologia o postulado de Condillac de que uma ciência é acima de tudo uma ‘língua bem feita’. O estudo das práticas discursivas da língua ligado a uma técnica ou a uma ciência através de documentos científicos não pertence, em bom rigor, à terminologia, mas sim à linguística. As críticas à abordagem wüsteriana foram feitas do ponto de vista linguístico, e são de alguma forma justas, porque a língua natural não tem vocação para ser normalizada. Mas sob o ponto de vista científico, a definição de termo deve basear-se num sistema lógico. Confunde-se o discurso com a conceptualização, a palavra com o conceito⁶⁴. Não devemos confundir a

⁶³ “En toute rigueur, une terminologie n’est pas une langue. Elle s’intéresse à des notions (concepts) et aux expressions les nommant (approche onomasiologique) ou à des expressions métiers et aux connaissances qu’elles désignent (approche sémasiologique). (...) La terminologie ne s’intéresse aux mots que dans la mesure où ils désignent des notions (ou concepts) (...).” (Roche, 2005, pp. 4-5)

⁶⁴ “Le concept structure la réalité de manière stable, indépendamment de la langue – il y a ici bi-univocité entre le concept et sa dénomination. Il ne doit pas être confondu avec le mot – un terme s’emploie comme un mot en texte – dont le signifié découpe de manière contingente une réalité construite en discours. (...) Le concept a disparu au profit du mot qui en parle – et non du signe qui le

terminologia com a sua manifestação linguística – a verbalização de uma prática através de uma linguagem de especialidade. A conceptualização e a representação dos objectos do mundo para fins de manipulação devem ser as suas actividades centrais.

Na sociedade numérica em que vivemos, Roche considera fundamental uma operacionalização da terminologia, reactualizando o primado do conceito e do extralinguístico⁶⁵ para numerosas aplicações. A designação de ‘ontoterminologia’ traduz essa necessidade, preservando ao mesmo tempo a sua dimensão sociolinguística: “Nous introduisons le néologisme ontoterminologie pour désigner cette approche qui place l’ontologie au centre de la terminologie.” (Roche, 2007, p. 8) A determinação do sentido dos termos requer o conhecimento preliminar do sistema conceptual para que consigamos alcançar a sua significação. Neste contexto, a ontologia, enquanto método de modelação e resultante da sua aplicação a um domínio, constitui uma das vias mais prometedoras para a construção e representação do sistema conceptual das terminologias⁶⁶. A terminologia não é somente uma ciência dos signos, é uma ciência das coisas que não deve ser reduzida à análise linguística do discurso científico e técnico, à pesquisa do sentido dos termos em discurso ou a uma lexicografia de especialidade. Se a terminologia não se alterar, teórica e metodologicamente, corre o risco de desaparecer enquanto disciplina autónoma, ou ser absorvida por uma lexicografia de especialidade ou pela engenharia do conhecimento. Na primeira, estudaria os fenómenos linguísticos, na segunda, a problemática da representação computacional do conhecimento. Roche está consciente da importância da terminologia, por isso afirma que

définirait. (...) la solution réside davantage dans la prise en compte d’une organisation extralinguistique que dans les termes employés”(Roche, 2009, p. 2)

⁶⁵ “Pouvons-nous dire pour autant que la terminologie relève uniquement de la linguistique? L’extralinguistique, que de facto s’exclut de la linguistique, et dont relève la conceptualisation du monde, intéresse au premier chef la terminologie.” (Roche, 2005, p. 2)

⁶⁶ “La conceptualisation du monde et sa représentation nous amènent à la notion d’ontologie. Cette dernière constitue aujourd’hui une des voies les plus prometteuses quant à la modélisation et à la représentation formelle du système notionnel des terminologies.”(Roche, 2005, p. 6)

“(…) la terminologie, en tant que discipline scientifique, est indispensable si on considère que son objet premier est de comprendre le monde, décrire les objets qui le peuplent et trouver les mots justes pour en parler. Bien que la terminologie classique s’inscrive dans cette finalité, force est de reconnaître qu’elle ne donne pas entièrement satisfaction tant d’un point de vue logique (en proposant des définitions consistantes) que computationnel (à travers une représentation du système notionnel qui puisse donner lieu à un calcul informatique).”(Roche, 2009, p. 2)

Ao contrário dos princípios filosóficos que norteavam a ontologia metafísica, hoje o interesse pelas ontologias não está no estudo do ser enquanto ser, mas sim na descrição de um domínio cuja conceptualização é consensual e coerente. Neste processo, o problema da língua natural constitui o maior obstáculo a enfrentar na construção de ontologias⁶⁷. Uma ontologia constrói a conceptualização do mundo em bases lógicas e empíricas, onde os objectos são descritos tal como são apreendidos, somente através das qualidades que a experiência lhes atribui. Os aspectos descritivos e de raciocínio desses domínios teriam por isso primazia sobre os discursos aos quais podem dar lugar. Contudo, Roche considera que o grande problema das ontologias permanece o mesmo de há vinte anos atrás: representamos conhecimento sem necessariamente compreender o conhecimento. Para Roche, o linguista e o filósofo, ao contrário do informático, ainda não sabem como lidar com as ontologias. Torna-se importante que a terminologia e as ferramentas de processamento de língua natural revejam e redefinam as suas metodologias e o seu objecto de estudo. (Roche, 2007a)

Relativamente ao termo, entende que a sua principal função é denominar o mundo, e não designar o mundo. A partir do momento que o utilizamos em discurso, passamos a ter uma construção de sentido restrita. A conceptualização só é clara se for descontextualizada. Para Roche “terme d’usage” é a expressão utilizada em texto técnico ou científico, que pode não coincidir com o “terme normé”, utilizado ao nível conceptual e ontológico pelo especialista. Este representa verdadeiramente a conceptualização e pode estar totalmente ausente do

⁶⁷ “Pour qui a assisté et participé à des réunions de conception de systèmes d’information, on peut être étonné (...) de ce souci constant de vouloir sortir de la langue naturelle et de ses ambiguïtés.” (Roche, 2007, p. 4)

discurso verbal. Dado que cada especialista se expressa verbalmente de uma determinada forma, temos de sair da língua natural para preservar a conceptualização, propondo termos normalizados que podem não ser utilizados em discurso, mas sobre os quais se acorda a significação do sentido. O “terme d’usage” apoia-se no “terme normé”, otimizando o conhecimento quando é transmitido em situação de comunicação. Um termo num texto torna-se uma palavra. Por isso, um termo só existe verdadeiramente fora do texto. Assim, podemos definir termos em terminologia que nunca são usados em texto, e, por outro lado, a terminologia pode conter termos que não são palavras. Um termo deve ser unívoco, e, em discurso, perderia essa propriedade.

Para Roche, sendo o conceito um elemento de pensamento, uma entidade cognitiva⁶⁸, não o podemos alcançar através do texto. No texto não temos conceitos nem conceptualizações. Temos uma estrutura lexical com relações linguísticas. Por isso, para operacionalizarmos a terminologia necessitamos colocar no centro das nossas preocupações o especialista e o engenheiro do conhecimento. A língua natural não é um bom veículo de transmissão de conhecimento de domínio, dado que a comunidade de especialistas não está sempre de acordo sobre a conceptualização do mundo ou sobre o modo como exprime o seu saber. O especialista tenta capitalizar conhecimento através da terminologia, porque é algo que percebe, mas o desacordo tende a diminuir quando recorrem a esquemas e não a texto. O objectivo das ontologias é atingir uma descrição estável do mundo sobre o qual se possa acordar. Esse conhecimento, frequentemente tácito, corresponde a uma cristalização num determinado momento de um saber comum e partilhado. Explicitar esta conceptualização comum, mas implícita, é um problema difícil que não pode ser resolvido sem a ajuda dos especialistas de domínio.

Para Roche, se considerarmos a terminologia como uma ciência independente da linguística, e uma ontologia como uma especificação formal uma conceptualização baseada numa realidade extralinguística, a extracção de ontologias a partir de texto é uma prática que não conduz a resultados susceptíveis de serem partilhados e reutilizados. Baseado na dualidade entre

⁶⁸ Roche diferencia a sua definição de conceito como entidade cognitiva da definição de Temmerman.

conhecimento linguístico e extralinguístico, o autor desmonta a teoria de que a estrutura lexical do texto coincida com a estrutura conceptual do domínio para provar que não é possível, ou pelo menos é muito pouco profícua, a tarefa de extracção de ontologias através de texto: “Linguistics is mainly interested in the relationships between *signifier* and *signified* when ontology is mainly interested in the relationships between concept and object.” (Roche, 2007a, p. 6) Tal como Brewster, Roche insiste na natureza incompleta do texto, categorizando-o como um documento pertencente à ‘língua em acção’. O texto apresenta-se como um ‘retrato linguístico’ de um determinado domínio, num determinado contexto e num determinado momento, e não, recorrendo a Gruber, como uma especificação da conceptualização de conhecimento invariável. O texto científico e técnico não deve ser considerado como repositório de conhecimento extraível para ontologias: “Even if we can extract useful information from texts about the domain conceptualization, we have also to bear in mind that the conceptual structure extracted from texts (from the lexical structure) is by definition corpus-dependent and cannot be generally considered as a domain ontology.” (Roche, 2006, p. 6)

A rede de palavras ligadas por relações linguísticas não encaixa necessariamente com a conceptualização de um domínio. Uma ontologia construída a partir de texto, se isso fosse possível, corresponderia a uma abordagem semasiológica, enquanto uma ontologia construída a partir do especialista participaria de uma abordagem onomasiológica. O texto pode ser objecto de observação e análise. Mas, o que fazem os extractores com os milhares de expressões que detectam? Pouco, pensa Roche. No domínio científico manipulamos conceitos e por vezes em número muito reduzido. Para a semântica referencial, um termo aponta para outro termo, mas isso não é suficiente, porque é construída no discurso. O termo aponta antes para um objecto, algo que está fora do discurso. A semântica distribucional é uma ferramenta interessante, mas não traduz nada em concreto do ponto de vista conceptual. Obtemos descrições de palavras, não conceitos. São *clusters* linguísticos, dependentes do corpus, e o conceito é extralinguístico. Por isso afirma que “(...) ontological knowledge, which is necessary for text understanding, is not in general embedded into documents.” (Roche, 2007a, p. 1)

Roche não rejeita totalmente o tratamento do texto. Além da identificação de conhecimento semântico, pode fornecer um certo conhecimento conceptual. A extracção de candidatos a

termo e de relações semânticas do corpus através de análise automática de texto é hoje um domínio de pesquisa activo. Mas “(...) there is no direct relationship between ontology and text.” (Roche, 2006a) Para Roche, o que temos de reter é que as estruturas lexicais que podemos extrair dos textos não se sobrepõem à conceptualização do mundo: dizer não é conceber. Uma estrutura lexical, num determinado contexto, reflecte o uso linguístico particular da conceptualização do domínio. Esquecer isto é conduzir a sistemas não reutilizáveis dependentes de um corpus que não pode ser qualificado de sistema conceptual nem de terminologia.

A abordagem da inteligência artificial através da perspectiva de Roche conduz, naturalmente, a uma reacção da nossa parte. Embora reconheçamos que o autor se destaca clara e positivamente nas suas preocupações com a contribuição da terminologia e da linguística na área da representação do conhecimento, importa clarificar algumas ideias centrais desta aproximação teórica.

A nossa primeira questão é saber se podemos considerar uma ontologia como um artefacto de representação do conhecimento. A ideia central de Roche sobre esta matéria é a consciência da separação entre o mundo e o que podemos dizer sobre o mundo. Nesse ponto, estamos de acordo. Mas para a inteligência artificial, o conhecimento representado pela ontologia é estático, de carácter descritivo, sem interpretação. E aqui várias questões se levantam: Existe conhecimento sem interpretação? O conhecimento não provém da cognição? Não é, por essência, algo dinâmico? Na nossa opinião, uma ontologia pode não conter, sob o ponto de vista restrito da lógica, uma interpretação, mas, dado que é um objecto concebido pelo homem, manifesta de alguma forma uma visão do mundo, interpretando-o. Apesar de ser partilhada, comum, sujeita às regras de uma linguagem formal e somente a elas, temos um tipo de conhecimento construído por um conjunto de criadores/observadores/interpretantes com uma vertente comunicativa. Por outro lado, o conceito de conhecimento é algo em constante mutação. Se considerarmos que a ontologia representa conhecimento estático, a questão da brevidade e da utilidade de uma ontologia torna-se uma questão importante. Teríamos ainda de esclarecer o conceito de ‘agente cognitivo’, mais de acordo com as propostas de Newell, para que não existam dúvidas sobre o tipo de cognição em causa.

Pensamos, por isso, que uma ontologia é antes uma ‘manifestação do conhecimento’. Na nossa perspectiva, o que muda é a conceptualização e, nesse sentido, uma ontologia não trabalha sobre a essência dos objectos, mas sobre conceptualizações dos objectos.

Quanto à terminologia, Roche assume que a sua preocupação fundamental é o conceito enquanto entidade extralinguística, e não o termo. O conceito é definido por um conjunto de características numa linguagem formalizada com o objectivo de partilha e reutilização de conhecimento. A partir daqui, constrói os seus argumentos no sentido de justificar a terminologia como uma ciência que não pertence à linguística, embora recorra a ela para o estudo e análise dos termos científicos. Como não define o conceito, somente o descreve como somatório de características, pode dispensar o termo enquanto entidade lexical. Ao afirmar em (Roche, 2007a) que uma ontologia é uma descrição de conceitos, não uma definição de conceitos, permanece fiel ao princípio da diferença específica. Mas descrever um conceito faz-nos possuidores da sua significação? Segundo o autor, uma ontologia dá conta de uma significação. O significado e o sentido encontram-se no texto, não na ontologia. Se uma ontologia é algo formal, o que existe antes desse estágio formal? Conceptualizações? E o que são conceptualizações? Descrições em língua natural das características dos conceitos?

Entendemos que existem várias práticas terminológicas e cada uma vai activar determinadas abordagens, tendo em vista o objectivo da aplicação. Não partilharmos por isso da restrição de Roche da aplicação, e da própria existência da terminología, a ‘domínios científicos’, exclusivos das ‘ciências duras’. Parece-nos claro que todos os domínios desenvolvem a sua própria terminologia e que as ‘ciências moles’ são igualmente científicas. A dificuldade está na formalização desse conhecimento. As conclusões de Roche estão muito ligadas às dificuldades sentidas pelo seu papel de engenheiro de ontologias numa área científica muito específica, em ambiente altamente tecnológico, com objectos concretos, uma ciência ‘pura e dura’, propícia a conceptualizações através de esquemas, utilizando termos que muitas vezes não são unidades lexicais. É menos difícil, diríamos, representar formalmente este universo do discurso, e mais fácil argumentar que a terminologia exerce a sua função fora do discurso verbal. Descrever a natureza do conhecimento independentemente da sua expressão linguística é um objectivo válido, embora muito difícil de alcançar. É necessário, para isso, que todos os intervenientes estejam conscientes do tipo de conhecimento que uma ferramenta informática pode fornecer

e do tipo de conhecimento que uma análise terminológica de texto pode dar. A semasiologia e a onomasiologia têm pontos de partida diferentes, activam propriedades e valências diferentes, apresentam resultados diferentes, mas convergem para um mesmo objectivo⁶⁹.

Concordamos com a opinião de que os conceitos não se encontram no texto. São entidades conceptuais, extralinguísticas. Mas esse facto não impede que possamos encontrar manifestações linguísticas que apontam para uma organização conceptual. A terminologia, inserida nas ciências da linguagem, também tem em conta a parte conceptual. O termo é algo pluridimensional. Tem uma dimensão linguística e simultaneamente uma dimensão cognitiva, dependendo do sistema onde é utilizado. Pensamos que o terminólogo, quando consulta o especialista – de resto, uma contribuição fundamental para o seu trabalho – não se limita a questionar a parte linguística, bem pelo contrário. Concordamos com o argumento de que não devemos reduzir a terminologia à linguística, mas achamos falso conceber a terminologia como uma disciplina totalmente externa à linguística.

Roche afirma que “(...) si l’utilisation d’un langage formel permet de s’abstraire des problèmes d’interprétation et d’ambiguïté que pose la langue d’usage et d’assurer certaines propriétés comme la cohérence et l’opérationnalisation des systèmes notionnels, elle ne permet pas de résoudre tous les problèmes et en particulier celui de la construction des systèmes notionnels.” (Roche, 2007, p. 7) A linguagem formal não consegue, por si só, construir sistemas conceptuais. Precisa de outros contributos. Além disso, na nossa perspectiva, há necessidade de uniformizar a terminologia utilizada pelos terminólogos e pelos engenheiros informáticos. Não podemos confundir lexicografia com terminografia, por um lado, e, por outro, o conceito de termo é oriundo das ciências da linguagem, e não deve ser utilizado em áreas importantes como a inteligência artificial sem um debate teórico alargado.

Do mesmo modo, tendo em conta a perspectiva de Roche do que é uma ontologia⁷⁰, pensamos que, se a ontologia procura as palavras justas para compreender o mundo, recorre,

⁶⁹ Tal como Roche afirma, “(...) doit-on nécessairement décréter qu’une discipline a la préséance sur une autre? Qui, du langage ou de la pensée ou du réel, précède et conditionne l’autre? Nous sommes davantage en présence d’appréhensions différentes mais complémentaires et indissociables du monde correspondant à des pratiques différentes.” (Roche, 2005, p. 2)

necessariamente, ao linguístico, por isso não deve apresentar princípios orientadores exclusivamente extralinguísticos. Se a popularidade da ontologia se deve em grande parte à capacidade que teria para capturar um entendimento partilhado dos termos⁷¹, reforça a nossa opinião de que a inteligência artificial cedo se apercebeu que a representação do conhecimento passaria por resolver o problema da língua natural, sendo, por consequência, também uma questão linguística. Os engenheiros teriam de se basear em princípios linguísticos e terminológicos para conseguirem chegar a uma semântica precisa. As ciências da linguagem encaram este processo como um todo, encaram a complexidade e a ambiguidade como fenómenos naturais e estudam-nos em todas as suas vertentes.

A este propósito, consideramos interessante realçar uma perspectiva adoptada por alguns investigadores da inteligência artificial que propõem uma abordagem mista, no sentido em que leva em consideração não só a participação fundamental da comunidade de especialistas enquanto fornecedores de conceptualizações, mas também dá liberdade e até aconselha essa comunidade a recorrer à língua natural nos seus vários registos. Tendo como base a abordagem de Cruse⁷² da semântica lexical, a teoria de Conceptual Blending⁷³ proposta por

⁷⁰ Nas definições de ontologia apresentadas nos vários artigos observa-se uma evolução relativamente ao propósito de uma ontologia. Começa por relacioná-la com a língua natural, constituindo-se também como um conjunto de termos e definições, passando a assumir-se mais tarde como uma descrição extralinguística de conceitos e relações entre conceitos. Distancia-se também cada vez mais da palavra 'termo' e passa a utilizar somente 'conceito'.

⁷¹ "Our main objective is to define the meaning of terms (words) which refer to conceptual knowledge (...) It implies that we rely on the classical semiotic triangle (...) where the meaning of a term is the concept which it refers to." (Roche, 2003a, p. 2)

⁷² Cruse propôs uma teoria que designou por Teoria da Construção Dinâmica de Sentido (*Dynamic Construction of Meaning* in Cruse, D. Alan (2000) "Aspects of the micro-structure of word meanings," in Y. Ravin & C. Leacock, eds. *Polysemy: Theoretical and Computational Approaches*, Oxford University Press, Oxford, pp. 30-51. Teoria desenvolvida em Croft, William, & D. Alan Cruse (2004) *Cognitive linguistics*, Cambridge University Press, Cambridge, MA.), onde a suposição fundamental é que o aspecto mais estável de uma palavra é o seu signo falado ou escrito; o significado é instável e dinâmico pois é construído em cada contexto onde é utilizado. Cruse cunhou o termo *microsentido* para cada variação subtil de significado à medida que a palavra é usada em jogos de palavras diferentes. Ou seja, as palavras têm um número infinito de microsentidos dinâmicos que se alteram e que dependem do contexto.

Soares e Pereira (2008) relativamente à construção daquilo que designam por ‘artefactos semânticos’, correspondem a uma ‘especificação informal da conceptualização’⁷⁴. Soares e Pereira defendem “a socio-cognitive turn” (2008, p. 2) na engenharia do conhecimento, capaz de gerar ferramentas que consigam lidar com contextos complexos, não estruturados e não localizados, característicos, de resto, da maioria do conteúdo de partilha de conhecimento em ambiente colaborativo: “According to this view, meaning is constructed at the conceptual level: meaning construction is equated with conceptualization, a dynamic process where linguistic units serve as prompts for an array of conceptual operations and the recruitment of background knowledge.” (Soares & Pereira, 2008, p. 3) Com base nestes princípios teóricos, concebem plataformas de integração conceptual em redes colaborativas através de ferramentas informáticas onde cada organização desenvolve conceptualizações que necessitam de um entendimento comum. Essas conceptualizações situam-se na fase inicial da construção da ontologia, ou seja, na sua fase informal, sem intervenção ainda de uma linguagem de representação, socorrendo-se das mais variadas fontes de análise, nomeadamente textuais, sendo alvo de constante negociação através da colaboração directa dos especialistas de domínio que expressam esse consenso graficamente através de um mapa

⁷³ Na Teoria da *Conceptual Blending* a construção de sentido é algo contínuo com forte presença humana, diferente de construções altamente formalizadas. Através dessa teoria, estabelece uma ligação entre a semântica cognitiva e a representação do conhecimento, considerando os quatro princípios da abordagem da semântica cognitiva: “1/ the conceptual structure is embodied (...) 2/ the semantic structure is the conceptual structure (...) 3/ meaning representation is encyclopedic (...) 4/ meaning construction is conceptualization.” (Soares & Pereira, 2008, p. 3)

O primeiro pressuposto implica que a semântica cognitiva explore a relação entre o homem e o mundo exterior, construindo a sua teoria em sintonia com essa relação; o segundo entende que a estrutura semântica (“the meanings conventionally associated with words and other linguistic units”) (Soares & Pereira, 2008, p. 3) sejam equivalente à estrutura conceptual; quanto ao terceiro, entende que as palavras são pontos de acesso a repositórios mais vastos de conhecimento, ultrapassando o conhecimento dicionarístico, e que apontam para conceitos ou domínios conceptuais; finalmente, o quarto pressuposto considera que a língua em si não codifica sentido, a conceptualização é que equivale à construção de sentido. As palavras são um dos pontos de partida da construção de sentido, ou seja, da conceptualização.

⁷⁴ Expressão que contrasta propositadamente com a definição de ontologia de Gruber (“an explicit formal specification of a conceptualization”) e que já utilizámos nesta dissertação.

conceptual. O objectivo é uma “co-construction of semantic resources.” (Soares & Pereira, 2008, p. 9)

A nova geração de ferramentas proposta por Soares e Pereira pode indicar, a nosso ver, uma tomada de consciência de que: primeiro, as ferramentas propostas não conseguiram dar resposta às necessidades da representação do conhecimento; segundo, especialista e texto são complementares, sendo contraproducente separar os dois; terceiro, devemos optar por extrair de todos os recursos as suas mais-valias, minimizando as dificuldades e otimizando os contributos. Nesta abordagem, existe um revisitar de princípios semasiológicos que são transpostos para a própria ontologia. Em ambiente de partilha, um recurso semântico é um processo contínuo, a tal ‘construção dinâmica de significados’ que a semasiologia defende, e não somente um conjunto de conceitos formalizados e estáticos. Poderíamos colocar a questão de saber se, tendo em conta a postura de Roche sobre a necessidade de formalizar de forma unívoca e perfeitamente explícita os conceitos e as suas relações, até que ponto a adopção de pressupostos da semântica cognitiva iria permitir sustentar esse objectivo. Soares e Pereira, porém, são claros nesta matéria: a teoria da Conceptual Blending aplica-se à ‘especificação informal da conceptualização’, sendo necessário contar sempre com a intervenção humana. Torna-se necessário, por isso, esclarecer que não devemos confundir recursos semânticos com ontologias, nem ter como pressuposto que esses recursos semânticos partem do texto.

Partilhamos assim da opinião de Soares e Pereira de que as palavras são pontos de acesso a repositórios mais vastos de conhecimento, e por isso um bom ponto de partida para determinado tipo de construção de conceptualizações. Recordamos que, do nosso ponto de vista, os conceitos não se encontram no texto. Contudo, defendemos também que os termos e as relações semânticas entre termos podem desvendar aproximações, maiores ou menores, a uma organização conceptual do domínio. Finalmente, a designação de ‘recurso semântico’, pelas razões expostas neste trabalho, estaria ligada à manifestação linguística do conhecimento. Tendo em conta a exposição de Soares e Pereira, parece-nos que a análise de dados não é baseada somente em texto, mas também, e em grande parte, no debate entre especialistas que, como os próprios autores referem, resulta na elaboração de mapas

conceptuais. Em abono de uma comunicação clara entre linguistas, terminólogos, especialistas e informáticos dentro do contexto da construção de ontologias, propomos a designação de ‘recurso informal’, como contraponto ao momento em que a utilização de uma linguagem formal altera, de forma decisiva, o formato destas contribuições.

Concluimos, por isso, que é o processo de construção de ontologias que gravita em torno da terminologia, e não o contrário. É a terminologia que constrói o sistema conceptual, porque o especialista pode não recorrer ao texto, mas recorre sempre à terminologia no acto de comunicação. O mesmo se verifica, de resto, com o engenheiro de ontologias, embora dificilmente o admita.

Quanto às observações de Roche relativamente ao termo, concordamos que este, como já referimos várias vezes, não é necessariamente uma entidade lexical, e Roche tem razão ao chamar a atenção para esse facto. No contexto das ontologias, pensamos também que, se aceitarmos uma conceptualização como algo extralinguístico, a descontextualização do discurso verbal surge como algo natural, embora não deva ser uma condição, mas sim uma metodologia, que poderá ser utilizada de acordo com o objectivo da aplicação. Não podemos deixar de contestar, porém, algumas afirmações proferidas. As várias definições que Roche apresenta tentam sustentar a tese de que a terminologia e a linguística não têm relação directa. Apesar de encararmos um texto como um acto de manutenção do conhecimento, rico em recursos sintácticos conotativos, pensamos que o termo em discurso pode ser monossémico – note-se que monossémico é diferente de unívoco⁷⁵. Um texto de especialidade pode conter inúmeras definições explícitas dos termos do ponto de vista discursivo. Lembremo-nos, porém, que uma conceptualização é forçosamente diferente de um ser humano para outro. Para alcançar a univocidade necessária a uma representação formal, foi necessário existir debate, consenso, partilha. Os caminhos que se percorreram no extralinguístico, são também percorridos no linguístico.

⁷⁵ “Um *termo* é considerado monossémico quando tem uma só significação e designa um só *conceito*. A *monossemia* corresponde a um ideal de comunicação em situação de especialidade. No entanto, é um fenómeno raro em língua, mais frequente em discurso. A inclusão do *conceito* designado por um *termo* num *sistema conceptual* para um dado *domínio* aumenta as probabilidades da relação biunívoca efectiva entre um *termo* e um *conceito*.” (Costa & Silva, 2006, p. 8)

É verdade que temos palavras no texto. Mas determinadas palavras assumem o estatuto de termo num determinado contexto. O termo existe no texto e fora do texto, sendo que é em discurso que a terminologia se actualiza. Há que distinguir entre léxico, léxico de especialidade e terminologia. O léxico pertence ao nível da língua, o vocabulário pertence ao nível do discurso. Devemos, portanto, distinguir uma entidade linguística que designa um conceito, e uma entidade linguística que ocorre em discurso de especialidade mas que não designa um conceito. Daí que rejeitemos a afirmação de que um termo no texto se transforme numa palavra e que só exista enquanto termo fora do texto.

Do mesmo modo, ao contrário do ponto de vista da inteligência artificial, muitas vezes a conceptualização só é clara quando é verbalizada, provindo também, senão na maioria das vezes, da língua natural. Nas palavras de Roberto, "(...) the basic component of communicative ability is linguistic competence (...)." (Roberto, 1999, p. 85) A ciência serve-se da língua natural como veículo privilegiado na transmissão do conhecimento, caso contrário este encontrar-se-ia encerrado dentro do nosso cérebro. A crítica que Roche desfere à língua natural recai também sobre a ontologia: do mesmo modo que restringe as funções da língua natural à comunicação, limita a ontologia a um artefacto lógico entre sistemas informáticos.

Relativamente ao conceito e ao tipo de relações que podemos encontrar no texto, a nossa posição encontra-se já explanada nesta dissertação, sendo que concordamos com a afirmação do autor sobre o facto de que no texto encontramos informação lexical. Mas não só. Como o próprio autor afirma, "(...) a concept is not a well formed formula: 'Being is. Being is not true or false'." (Roche, 2003a, p. 3) Está aqui patente a necessidade da epistemologia para dar corpo a essa representação formal, caso contrário teríamos fórmulas que, embora ligadas por relações, constituiriam um conjunto de dados sem uma estrutura verdadeiramente coerente. Esses princípios epistemológicos estão não só no engenheiro do conhecimento, como também no especialista, no terminólogo, e em todos os intervenientes da construção de uma ontologia. Se um conceito não é uma fórmula, um somatório de atributos, não deve ser tratado pela engenharia do conhecimento como tal.

Roche baseia-se também no triângulo semiótico e no estruturalismo para, juntamente com a base aristotélica, conseguir provar a utilidade de uma representação formal do conceito. Essa situação pode ser uma realidade, pois quando formalizamos uma ontologia temos uma

linguagem simbólica e lógica. Mas a ontologia é construída para partilha e reutilização. Se não temos a intervenção da língua natural em nenhum estágio da sua construção, só servirá propósitos de partilha entre sistemas, só conseguirá ser lida por máquinas, ou por comunidades ligadas a símbolos que não linguísticos, como a matemática ou a física⁷⁶. Sabemos que o signo linguístico é algo convencional e arbitrário. Mas, se o termo enquanto unidade lexical corresponde a um conceito também arbitrário e convencional, levantam-se questões em termos da natureza do conhecimento que transmitem. Em última análise, uma ontologia não seria constituída por conceitos, mas sim por conjuntos de diferenças. Na realidade, o que preocupa o engenheiro de ontologias são as diferenças, pois é sobre elas que o sistema faz inferências. Porém, ao seleccionarmos essas diferenças, estamos a assumir um compromisso ontológico. Para isso os engenheiros que trabalham com ontologias necessitam também reflectir teoricamente sobre os princípios linguísticos e terminológicos, em conjunto com as ciências da linguagem. Na realidade, a terminologia sempre utilizou o conceito na óptica que Roche defende para a inteligência artificial. Para o terminólogo, o conceito está inserido num sistema que se manifesta através do termo. O terminólogo estuda, tradicionalmente, a manifestação linguística do sistema. O facto é que a terminologia determinou muito antes da ontologia da inteligência artificial as bases da representação do conhecimento. Fazia parte dos objectivos de Wüster, só que na sua época não se aplicava ainda a formalização, nem existia internet.

Como Roche defende a integração dos princípios epistemológicos, é natural que se aproxime de Aristóteles, ligado à filosofia, a uma percepção do mundo que é um pouco estranha aos informáticos puros. Pensamos que é aí que Roche tem a sensibilidade para estar mais próximo da terminologia, uma vez que esta não se limita a análise de dados assépticos: estuda toda uma envolvente para se aproximar cada vez mais do conhecimento que os termos e as relações entre os termos podem dar a conhecer. Ao afirmar que o conceito é uma unidade de compreensão e conhecimento⁷⁷ multifacetada, entra em contradição, quanto a nós, com o carácter invariável e estável do conhecimento que uma ontologia pretende ter, e mais de

⁷⁶ Esta constatação não põe em causa o facto de um termo não se limitar a uma entidade lexical.

⁷⁷ “[N]ous proposons de définir le concept comme une unité de compréhension (...) au sens d’une unité permettant d’appréhender la réalité dans sa diversité (...) le concept est une connaissance.”(Roche, 2009, pp. 3-4)

acordo com a complexidade cognitiva que a terminologia apreende quando analisa os termos e as suas relações nas suas actualizações discursivas. Se o conceito é uma unidade de compreensão, embora Roche se afaste deliberadamente da perspectiva cognitiva de Temmerman, não pode recusar-se ou excluir-se aos meios dessa manifestação, que são, entre outros, o texto.

Outra das grandes questões colocadas por esta abordagem da inteligência artificial, é saber se podemos ou não extrair ontologias a partir de texto, ou, por outras palavras, se um ponto de partida semasiológico é válido na especificação de conceptualizações com objectivos de manipulação computacional de dados na forma de ontologias⁷⁸. Concordamos com a rejeição dos pressupostos expostos por Roche⁷⁹ para fins de construção de ontologias. No entanto, a análise de candidatos a termo ligados por relações semânticas pode fornecer dados importantes para a ontologia, apesar de estarmos situados na actualização discursiva do texto. Como afirmam Costa e Silva, “(...) s’il est vrai que l’organisation des connaissances qui se trouvent à un niveau extralinguistique est un des objectifs primordiaux de la recherche en Terminologie, il n’est pas moins vrai que c’est à travers l’acte de la parole, c’est-à-dire le discours, que nous pouvons accéder à la représentation de ces connaissances.” (Costa & Silva, 2009, p. 5) Se o plano conceptual não está expresso no texto, ou seja, está omisso, as duas estruturas não podem ser totalmente sobreponíveis, porque uma é muito mais complexa que a outra, embora a lexical contenha elementos da conceptual.

Dado que o texto não é uma tradução fiel em língua natural do conhecimento de domínio, não podemos aplicar a hipótese do mundo fechado. Entramos, igualmente, com muito conhecimento implícito, tanto por parte do autor, como do leitor. Partilhamos, por isso, da opinião de Costa e Silva:

⁷⁸ Recordemos as metodologias explanadas no início do presente capítulo, nomeadamente por Hjelm, que defendem a ‘aprendizagem de ontologias’.

⁷⁹ Os pressupostos que normalmente se colocam na eventual extracção de ontologias a partir de texto são: i) o conhecimento dos especialistas estaria patente no corpus, uma espécie de mundo fechado; ii) o texto seria uma estrutura superficial dessa estrutura mais profunda de conhecimento; iii) construir-se-ia a estrutura lexical a partir do texto (= rede de palavras ligadas por relações semânticas – hiperonímia/hiponímia, meronímia, sinonímia); iv) seguidamente, proceder-se-ia à construção da estrutura conceptual a partir da lexical; v) a estrutura conceptual seria validada por especialistas e considerada uma ontologia.

“Le lecteur à qui se dirige ce type de texte se trouve à un niveau de connaissance proche de celui de l’auteur, car théoriquement, il est en mesure d’appréhender les connaissances et l’intention de ce qui lui est communiqué. Ce nivellement de connaissances a de l’influence dans la façon dont l’auteur écrit son texte : il se gère une espèce de complicité qui amène à une forte présence du *non-dit* qui souvent occupe une place capitale dans le texte, et qui, d’après nous, est une des propriétés les plus caractéristiques du texte de spécialité. Établir la relation entre le *dit* et le *non-dit*, entre l’explicite et l’implicite est une des tâches du lecteur spécialiste, puisque dans toutes les situations de communication, on a tendance à signifier plus que ce que l’on dit.” (Costa & Silva, 2009, p. 4)

O argumento de que as figuras de retórica utilizadas no texto iriam impedir a possibilidade de *matching* da estrutura lexical e da estrutura conceptual é também interessante. A utilização de figuras de retórica indica que o escritor e o leitor partilham a mesma conceptualização do mundo necessária para entender o texto. Este conhecimento é imanente, não está presente nos textos e por isso não pode ser extraído. Concordamos com esse facto. Mas acrescentamos: não pode ser extraído automaticamente, com uma transposição do texto para uma suposta ontologia. Um mesmo termo pode ter uma conceptualização totalmente diferente em vários domínios, e não estar dependente de um contexto linguístico. Os modelos conceptuais resultantes da análise de duas comunidades de especialistas sem recurso ao texto seriam necessariamente diferentes. A questão não está no texto, está no agente que o constrói.

A natureza incompleta do texto aplica-se da mesma forma à ontologia. Uma ontologia nunca está completa. Extrair ontologias de texto seria uma espécie de ‘Ovo de Colombo’, já que um dos grandes problemas da inteligência artificial é o factor tempo e o factor custo. As ontologias são artefactos construídos através de processos muito demorados de formalização. Embora o acesso ao texto em formato electrónico seja amplamente facilitado pela tecnologia actual, Roche entende que cometemos um erro ao encará-lo como uma fonte de extracção de ontologias. Pensamos que o erro está em encarar o texto como algo que contém ontologias ou conhecimento. Estamos a exigir do texto funcionalidades para as quais não foi concebido. Mais uma vez, estamos em níveis diferentes de análise. Somos também de opinião de que não podemos confundir a conceptualização de um domínio com os discursos a que essas

conceptualizações podem dar lugar. Uma coisa é o mundo, outra é conceptualizarmos sobre o mundo. Contudo, uma ontologia também pode ser encarada como uma outra forma de analisar os mesmos dados.

Roche afirma que “(...) les discours scientifiques et techniques relèvent de la langue de spécialité. Ils constituent, lorsqu’on étudie la terminologie d’un domaine sans en être un expert, la partie la plus visible et la plus directement accessible.» (Roche, 2007, p. 5) Pensamos que é aqui que se centra o argumento relativo à importância do especialista. O estudo do discurso sem a intervenção do especialista faz com que, não dispendo de conhecimento, nos concentremos no estudo dos termos em discurso, por isso não avançamos para o nível extralinguístico. Porém, os especialistas não constroem ontologias, estruturam conceitos e definem as relações entre conceitos. Por outro lado, temos de partir do princípio que existem ‘bons especialistas’, e ‘maus especialistas’. Para o terminólogo, a colaboração do especialista sempre foi, e continuará a ser, fundamental.

A língua natural é claramente um problema para a formalização. Porém, não concordamos em absoluto que seja considerada o pior veículo para transmissão de conhecimento. O texto, do nosso ponto de vista, continua a ser “(...) le moyen le plus efficace pour le spécialiste de communiquer avec les membres de sa communauté professionnelle.” (Costa & Silva, 2009, p. 4) O texto serve um objectivo, é um elemento operacional. O que podemos extrair dele depende directamente do que se procura. Aqui temos multiplicidade em tudo: na conceptualização do domínio, na estrutura conceptual, e na manifestação linguística. A análise terminológica no campo das ontologias pode adoptar o mesmo pressuposto, partilhando uma “(...) perspective de l’organisation des textes comme contribution déterminante pour l’organisation des connaissances.” (Costa & Silva, 2009, p. 3)

Por estarmos conscientes da profunda alteração tecnológica que a sociedade contemporânea sofreu nas últimas décadas, partilhamos com Roche a necessidade de readaptação da formação de um terminólogo. Actualmente, e dada a evolução exponencial das tecnologias de informação e das ferramentas de processamento de texto e de representação do conhecimento, todos os agentes envolvidos devem readaptar-se, nomeadamente o terminólogo, uma vez que tem de trabalhar com recursos informáticos por vezes muito complexos. O movimento, porém, não deve ser unilateral. Embora o engenheiro do

conhecimento entenda que é auto-suficiente, uma vez que domina a tecnologia, não deve sobrevalorizar a confiança que deposita num programa de tratamento de dados. Os resultados estatísticos podem ser até surpreendentemente elevados, mas, em termos de conhecimento, não representarem o que na realidade se gostaria de alcançar. Uma colaboração de sucesso entre especialista e engenheiro do conhecimento, tal como o próprio Roche afirma em algumas ocasiões, é mais visível nas ‘ciências duras’, onde a utilização de esquemas se revela muito útil, com eventual exclusão da língua natural. Mas, o que dizer das ‘ciências moles’? Vamos ignorá-las na representação do conhecimento? A presença de um linguista ou de um terminólogo, conhecedores do funcionamento da língua natural e do ‘significado’, como o próprio autor reconhece, não será, nesses casos, fundamental?

5.3 Notas conclusivas

O facto de, nos últimos anos, a ontologia ter conquistado um lugar preponderante na representação do conhecimento, não significa que a análise terminológica (dos termos e das relações entre os termos) tenha sempre como finalidade a extracção de conceptualizações para ontologias. A estrutura lexical que se pode retirar de um texto não representa, de facto, conhecimento (a não ser conhecimento de tipo semântico), por isso não pode servir como plataforma de transposição de dados. O texto pode ser um recurso indispensável para alguns objectivos, e dispensável para outros. A questão está em saber se o texto serve ou não determinada aplicação. Se serve, devemos utilizar as duas perspectivas, pois só a sua combinação poderá resultar num explorar ao máximo das potencialidades das duas abordagens. Nesse sentido, a metodologia de análise textual está dentro dos objectivos da ontologia, uma vez que vai contribuir para a ‘especificação informal da conceptualização’, construindo-se como um artefacto semântico, primeiro estágio do artefacto formal.

Parece-nos, pois, que a perspectiva onomasiológica não pode ser totalmente colada ao papel do especialista. Tanto o terminólogo como o especialista utilizam as duas perspectivas, dependendo do objectivo. O que acontece quando não temos acesso ao especialista? Não temos acesso ao conhecimento? As questões da extracção de ontologias e a divisão entre

perspectiva semasiológica e onomasiológica surgem, assim, como falsas questões. Utilizamos recursos para ajudar na construção de um artefacto de engenharia que pretende especificar conceptualizações numa linguagem formal para tentar representar o conhecimento de um domínio. Neste ponto de vista, o terminólogo tem uma maior sensibilidade para ajudar o especialista na extracção de termos, por isso utiliza, geralmente, o suporte do texto. Por seu turno, o especialista detém o conhecimento que, geralmente, não está no texto, e por isso tem maior sensibilidade para, no caso da utilização do recurso texto, ajudar o terminólogo na detecção dos candidatos a termo que vão compor a estrutura dos conceitos para organizar as conceptualizações com o objectivo de representar conhecimento através de uma linguagem formal. A função do terminólogo não se encontra assim desvirtuada. Antes se reforça, afirmando as competências específicas da sua profissão no novo enquadramento tecnológico da nossa era.

Encontramo-nos num estádio da ciência terminológica em que se registou uma evolução muito grande, conduzindo a um ambiente tecnológico substancialmente diferente do que era há dez anos atrás. Não podemos esquecer que Wüster defendia a onomasiologia porque não tinha como preocupação fundamental o texto, mas sim a comunicação eficiente em domínios de especialidade. Por outro lado, algumas correntes da terminologia reagiram aos princípios da Escola de Viena numa época em que não se vislumbrava um crescimento tão exponencial dos sistemas de informação. O recurso que efectivamente o terminólogo dispunha na altura era o texto. É natural que se evoluísse para um estudo mais completo da unidade termo, tal como era entendida pelas perspectivas estruturalistas, na sua parte sintáctica, e mais tarde pelas perspectivas sócio-cognitivas, na sua vertente do termo enquanto unidade de conhecimento. Aqui temos o grande cismo, uma vez que se promulgou muito essa vertente textual. A sociedade em que vivemos hoje, permite, através da tecnologia, transmitir conhecimento de muitas formas. Defendemos, por isso, que, do ponto de vista da terminologia, o termo constitui uma parte fundamental na representação do conhecimento, não como decalque, mas como manifestação linguística de potenciais conceitos.

Como tivemos oportunidade de observar, o linguista acha que o maior problema das ontologias é conceptual, e o engenheiro ontológico acha que o maior problema das ontologias é linguístico. A dicotomia na inteligência artificial é por vezes estabelecida entre linguística e

ontologia e não entre linguística e terminologia. A divergência encontra-se ao nível da língua natural, sendo que a terminologia assume os contornos de uma ontologia, uma ferramenta, que conta com a participação do especialista, do engenheiro e do terminólogo. Mas a questão permanece: “The problem is all the more difficult in that this knowledge is not directly accessible in usable form but is spread out through various bodies of knowledge, and more especially into documents.” (Roche, 2007a, p. 2) Concluimos que, apesar de todas as desconfianças em relação à língua natural, a inteligência artificial reconhece que os seus esforços vão no sentido de tentar suplantar, através de uma postura extralinguística, as qualidades que manifestamente a língua natural encerra.

6. Proposta de construção de um sistema conceptual

Sendo o nosso objectivo analisar o papel desempenhado pela terminologia na representação do conhecimento, e tendo expressado a nossa posição relativamente às abordagens descritas ao longo desta dissertação, interessa neste momento analisar que tipo de dados podemos extrair das metodologias que partem do linguístico e das metodologias que partem da observação do extralinguístico na construção de ontologias.

Baseamo-nos no pressuposto de que, sendo o conhecimento um elemento cognitivo, não é possível extraí-lo de texto ou de ontologias. Vamos trabalhar, portanto, com conceptualizações. A análise a desenvolver no presente capítulo situa-se num estágio anterior à representação formal das conceptualizações, permitindo construir o que já designámos por uma ‘especificação informal’. Isto porque entendemos que é nesse estágio que podemos contar com a colaboração fundamental da terminologia enquanto ciência e enquanto conjunto de práticas.

Encontramo-nos, pois, perante duas fontes de aquisição de dados: o discurso verbal, expresso, neste caso, através do texto de especialidade, e o especialista, enquanto entidade detentora de conhecimento de domínio. O ponto de partida do objecto de estudo – as conceptualizações – torna-se uma questão fundamental. Que papel desempenham o texto e o especialista na organização e representação do conhecimento?

Estamos conscientes das dificuldades inerentes à extracção de dados a partir de texto. Encarado como espaço de actualização constante, as conceptualizações ausentes e a forte presença do ‘não-dito’ constituem obstáculos difíceis de transpor. Por outro lado, a convicção de que o texto é um elemento operacional fundamental na comunicação humana, esteja ela ligada à inteligência artificial ou não, faz-nos questionar a sua importância na construção de ontologias. Procurou-se, portanto, seleccionar um corpus teoricamente enquadrado na tipologia de discurso expectável de encerrar elementos linguísticos que concorressem para a organização conceptual de um domínio.

Nesse sentido, para nos mantermos o mais próximos possível da metodologia da inteligência artificial defendida por Roche, foi tomada a decisão de construir em primeiro lugar um mapa conceptual de domínio com a colaboração exclusiva de especialistas, sem recurso à análise textual. Posteriormente, seleccionámos um corpus do mesmo domínio submetendo-o a técnicas de processamento de língua natural com o objectivo de construir uma rede lexical composta por termos e relações entre termos extraídos do texto e, com base nessa rede lexical, elaborar um mapa, utilizando a mesma ferramenta informática. Assim, as manifestações linguísticas do conhecimento susceptíveis de serem extraídas do corpus através da análise de texto foram representadas numa forma graficamente semelhante ao mapa elaborado directamente pelos especialistas com o intuito de observar, entre outras coisas, se a questão da representação irá desempenhar um papel decisivo junto do especialista, ou se, pelo contrário, não é a representação que está em causa, mas sim os dados representados.

Uma vez que partimos de uma postura baseada nas ciências da linguagem, é nossa intenção chegar a conclusões, embora não absolutas, sobre algumas questões, nomeadamente:

- Devemos optar pela utilização de uma única abordagem terminológica (semasiológica ou onomasiológica) na construção de ontologias?
- Os termos enquanto rede lexical representam conceitos?
- Um termo encontra-se ao nível do conceptual ou ao nível do discurso?
- A rede lexical construída a partir da análise de corpus representa ou não um sistema conceptual?
- Qual o valor do texto e o estatuto do corpus na representação de um sistema conceptual?
- A representação conceptual construída a partir da rede lexical influencia ou não a decisão do especialista?

6.1 Perspectiva extralinguística

A abordagem extralinguística utilizada no presente capítulo irá centrar-se na construção de um mapa conceptual de domínio.

Os princípios orientadores de uma abordagem conceptual baseiam-se no entendimento de que um sistema conceptual só poderá ser partilhado e reutilizado se for consensual e não ambíguo. Essas qualidades encontram-se, fundamentalmente, em dados extralinguísticos. À partida, seria possível assegurar deste modo, e com algum grau de certeza, a coerência e consistência necessárias à ontologia. Numa primeira fase, a forma mais utilizada para construir um sistema conceptual são suportes gráficos, nomeadamente mapas, onde, teoricamente, lidamos com conceitos e não com termos. A selecção dos conceitos e das relações entre os conceitos constituiria uma forma de pensar sobre o mundo, reflectindo uma organização de conhecimento especializado.

6.1.1. Mapa conceptual

Segundo Canas et al (2005), a visualização da informação tem sido um tópico de pesquisa recorrente, constituindo-se hoje como uma área formada com práticas estabelecidas. Pelo contrário, a visualização de conhecimento⁸⁰ é uma área relativamente nova que recebeu

⁸⁰ Canas et al distinguem entre informação e conhecimento: "Our interpretation of 'knowledge' is based on the theories of knowledge and learning from Ausubel *et al.* (1978) and Novak (1977) on which our research is supported. Novak (*ibid*) proposed that the primary elements of knowledge are *concepts* and relationships between *concepts* called *propositions*. Concepts are defined as 'perceived regularities in events or objects, or records of events or objects, designated by a label' (Novak, 1998). Propositions consist of two or more concept labels connected by a linking relationship that forms a semantic unit (Novak & Gowin, 1984). Most researchers agree that knowledge is a human creation. Using these propositions, we construct new knowledge by linking new concepts to knowledge we already have. Knowledge is constructed through meaningful learning (...) which takes place when the learner deliberately seeks to relate and incorporate new information to relevant knowledge he/she already possesses (Ausubel *et al.*, 1978)." Para o propósito do artigo, os autores consideram os mapas conceptuais como visualização de conhecimento, e todos os outros meios (imagens, vídeo, texto, som, gráficos, etc.) são entendidos como visualização de informação. Embora reconheçam que "some of the associated media (e.g., text) may depict knowledge", a distinção irá simplificar a discussão sobre a forma como os mapas conceptuais podem ser utilizados para integrar visualização de conhecimento e informação. (Cañas et al., 2005, pp. 2-4)

atenção devido ao interesse pela gestão do conhecimento. A visualização do conhecimento foi definida por Burkhard e Meier (2004) como “(...) the use of visual representations to transfer knowledge between at least two persons.” (Cañas et al., 2005, p. 1)

Os mapas conceptuais foram desenvolvidos por Novak em 1972 ao longo da sua investigação para dar conta das alterações do conhecimento infantil relativamente à ciência, e surgiram pela necessidade de representar esse conhecimento. Novak define os mapas conceptuais como “(...) graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts. Words on the line, referred to as linking words or linking phrases, specify the relationship between the two concepts.” (Novak & Cañas, 2006, p. 1) Para Novak, um conceito é definido como “(...) a perceived regularity in events or objects, or records of events or objects, designated by a label.”(Novak & Cañas, 2006, p. 1) Para além da sua função didáctica enquanto mecanismo de aprendizagem, os mapas conceptuais funcionam também como uma ferramenta de avaliação que designa por “meaningful learning processes”(Novak, 2000, p. 1), de resto semelhantes aos processos utilizados por cientistas e matemáticos na construção de novo conhecimento.

Os mapas conceptuais incluem frequentemente as chamadas “cross-links”, definidas como relações (proposições) entre conceitos em domínios diferentes do mapa conceptual. A etiqueta é, na maioria dos conceitos, representada por uma palavra ou por um símbolo. As proposições são constituídas por dois ou mais conceitos ligados por outras palavras para formarem uma declaração com sentido. Nos mapas conceptuais os conceitos são representados hierarquicamente: os mais gerais encontram-se no topo do mapa e os mais específicos estão dispostos abaixo, também de forma hierárquica. Normalmente os conceitos são representados por substantivos, e as frases de ligação são normalmente verbos, recomendando-se que se restrinja ao máximo o número de palavras. As frases de ligação podem expressar qualquer tipo de relação e não estão limitadas a nenhum grupo pré-definido (por exemplo, *is_a*, *part_of*, etc.). Contudo, é esta liberdade na construção das frases de ligação que impede os mapas conceptuais de serem automaticamente traduzidos para uma representação formal.

O contexto no qual o conhecimento é aplicado vai determinar de igual forma a estrutura hierárquica do domínio em causa. Para facilitar a construção do mapa, formula-se, normalmente uma questão central inicial relativa ao seu objectivo que, por sua vez, irá despoletar todo o raciocínio. Os mapas conceptuais construídos por pessoas diferentes sobre o mesmo domínio serão também necessariamente diferentes, pois cada um dos mapas representa o conhecimento de quem o construiu. Isso não significa que só uma representação esteja correcta.

De acordo com Cañas et al, num mapa conceptual bem construído, cada par de conceitos, juntamente com as frases de ligação, pode ser lido como uma declaração ou proposição individual com sentido. A estrutura é hierárquica e o nó mais elevado deve ser um bom representante do tópico do mapa, funcionando assim como “browsable knowledge models” (Cañas et al., 2005, p. 4) que representam visualmente o conhecimento dos especialistas.

Para o nosso propósito, é curioso notar que Novak sugere que, para a construção de um mapa conceptual, partamos de algo que desejamos compreender, como um texto: “Since concept map structures are dependent on the context in which they will be used, it is best to identify a segment of a text, a laboratory activity, or a particular problem or question that one is trying to understand. This creates a context that will help to determine the hierarchical structure of the concept map.” (Novak & Cañas, 2006, p. 3) O passo seguinte é identificar os principais conceitos do domínio através da elaboração inicial de uma lista, ordenando-os do mais geral para o mais específico. Trata-se de uma ordem aproximada, mas ainda assim útil no início do processo da construção do mapa. Parte-se depois para a construção do primeiro mapa, manualmente ou utilizando um programa informático.

É importante reconhecer que um mapa conceptual nunca se encontra finalizado. As suas várias versões devem ser revistas, posicionando os conceitos de tal forma que ajudem a clarificar o mapa final. Utilizar mapas conceptuais ajuda a que o conhecimento se torne “conceptually transparent”. (Novak & Cañas, 2006, p. 5) Novak e Canas defendem ainda que a utilização crescente de mapas conceptuais se deve à sua capacidade para capturar o conhecimento tácito dos especialistas: “Experts know many things that they often cannot articulate well to others. This tacit knowledge is acquired over years of experience and derives in part from activities of the expert that involve thinking, feeling and acting.” (Novak & Cañas, 2006, p. 24)

6.1.2 Metodologia

A construção de um mapa conceptual de um domínio de conhecimento requer que se preencham os seguintes pressupostos:

- Justificar e delimitar o domínio de conhecimento que se pretende analisar
- Pesquisar fontes de informação que possam coadjuvar essa delimitação
- Seleccionar um ou vários especialistas de domínio
- Seleccionar uma ou várias ferramentas informáticas

O domínio de conhecimento seleccionado para o nosso propósito foi 'Tratamentos Biológicos de Águas Residuais'. A relevância de um domínio é tanto mais oportuna quanto o avanço científico da área em causa e o Ambiente surgiu como um campo vasto de pesquisa. A selecção do domínio foi facilitada pela proximidade profissional com especialistas ligados ao tratamento de efluentes, suscitando o interesse pela observação do comportamento conceptual e linguístico dessa área do saber.

O crescimento exponencial da população mundial, a par de uma sociedade largamente urbanizada e industrializada, geraram problemas graves relacionados com a gestão das águas residuais urbanas e industriais, tornando a gestão dos recursos hídricos uma prioridade fundamental de todos os governos. Os resíduos resultantes dos efluentes agro-pecuários, domésticos e industriais constituem a fonte principal de poluição da água, verificando-se uma crescente atenção relativamente à preservação das fontes naturais deste recurso. Tem-se registado ao longo das últimas décadas uma evolução significativa na pesquisa e desenvolvimento de processos e tecnologias que permitam reciclar as águas residuais. Do mesmo modo, tenta-se substituir os tratamentos químicos e físicos, tradicionalmente utilizados, por sistemas biológicos, amigos do ambiente. A capacidade metabólica dos microrganismos de degradarem quase todos os compostos orgânicos que ocorrem naturalmente no meio ambiente tem incentivado os especialistas nesta matéria, impulsionados, recentemente, pelas descobertas da era 'pós-genómica', com as infinitas possibilidades de manipulação molecular. O tratamento biológico de efluentes congrega uma

panóplia de saberes oriundos de diversas ciências: biologia, química, física, biotecnologia, microbiologia, microbiologia ambiental, entre outras.

Foram consultadas diversas fontes documentais como tesouros, publicações científicas, fontes em linha e curricula de várias licenciaturas, assim como mestrados ministrados em universidades portuguesas, nomeadamente: Engenharia do Ambiente, Engenharia Biológica, Engenharia dos Recursos Hídricos, Engenharia Civil, Eco-Turismo e Eco-Agroturismo.

Encetámos uma colaboração estreita com especialistas da área do ambiente do Departamento de Ambiente e Ordenamento da Universidade de Aveiro. Ao longo dessa colaboração, assistimos às aulas da cadeira de Técnicas de Tratamento de Efluentes ministrada nesse Departamento. Contámos igualmente com a colaboração pontual das seguintes universidades:

- Universidade do Minho – Departamento de Biotecnologia Ambiental
- Instituto Superior Politécnico de Leiria – Departamento de Ambiente
- Instituto Superior de Engenharia de Coimbra – Departamento de Engenharia Química e Biológica
- Universidade Federal do Rio de Janeiro – Departamento de Microbiologia

Para este trabalho, utilizaram-se duas obras de referência na área como apoio documental para a elaboração do mapa conceptual, a saber:

- Tchobanoglous, George and Burton, Franklin L., *Wastewater Engineering: Treatment and Reuse*, Metcalf & Eddy, 2004
- Greenberg, Arnold et al, *Standard Methods for the Examination of Water & Wastewater*, American Water Works Association, 2005

Optámos pela visualização dos mapas através do programa CMapTools. Esta ferramenta colaborativa de software foi desenvolvida pelo Institute for Human and Machine Cognition, e contém funcionalidades adaptadas à internet. É de utilização intuitiva e amigável, permitindo a

fácil construção e alteração de mapas de conceitos semelhante a um processador de texto. Permite ainda que os utilizadores colaborem à distância, publiquem e acessem a mapas, se liguem a fontes para pormenorizar conteúdos e pesquisem informação na Web. Os mapas assim construídos podem ser armazenados em servidores e partilhados através da internet, fornecendo uma integração entre conhecimento e visualização de informação. Na ferramenta CMapTools um conjunto de mapas conceptuais e os recursos a ele associados sobre um determinado domínio do conhecimento é considerado um “knowledge model”. (Cañas et al., 2005, p. 6)

6.1.2.1 Construção do mapa conceptual

Um mapa conceptual não pretende ser uma ontologia. As relações entre conceitos estabelecidas numa ontologia requerem conhecimentos sobre linguagens de formalização que ultrapassam o objectivo deste projecto. O mapa inicial não continha, por isso, a especificação das relações entre os conceitos, embora fosse visualmente verificável que existiam relações hierárquicas e associativas entre os mesmos.

Sabemos que as relações conceptuais mais utilizadas em ontologias são a relação *is_a* e a relação *part_of*, correspondendo a relações hierárquicas genéricas e partitivas. A relação *part_of* é também conhecida como part-whole relation⁸¹, classificada como uma relação de mereologia, “(...) a comprehensive theory of parts and wholes (...)”⁸² assente em axiomas.

Foi sentida, por parte dos especialistas, a necessidade de complementar essas duas relações com outras que conseguissem expressar melhor a riqueza conceptual do domínio. Tendo em conta que o mapa a elaborar não obedeceu a regras de formalização, optámos por seleccionar as relações que estão na base da Unified Medical Language System (UMLS) propostas por

⁸¹ Guarino aprofunda esta relação em (Guarino, Artale, Franconi, & Pazzi, 1996a).

⁸² <http://www.britannica.com/EBchecked/topic/46274/axiomatizationU>

Budin⁸³ para o estudo do domínio ambiental, disponíveis em www.eea.eionet.europa.eu/Public/irc/enviowindows.

Figura 1 – Tipos de relações para o domínio ambiental: proposta

isa associated_with physically_related_to part_of consists_of contains connected_to interconnects branch_of tributary_of ingredient_of spatially_related_to location_of adjacent_to surrounds traverses functionally_related_to affects manages treats disrupts complicates interacts_with prevents brings_about produces causes	[associated_with] (continued) [functionally_related_to] (continued) performs carries_out exhibits practices occurs_in process_of uses manifestation_of indicates result_of temporally_related_to co_occurs_with precedes conceptually_related_to evaluation_of degree_of analyzes assesses_effect_of measurement_of measures diagnoses property_of derivative_of developmental_form_of method_of conceptual_part_of issue_in
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

O conjunto de conceitos e as relações entre conceitos que podem ser visualizadas nos mapas elaborados não reflectem conceptualizações formalizadas. Constituem, antes, uma representação informal do conhecimento do domínio, com a particularidade de ter sido construída sem recurso à análise de texto. A transformação do mapa conceptual em ontologia necessitaria, forçosamente, de encontrar relações conceptuais formais de substituição.

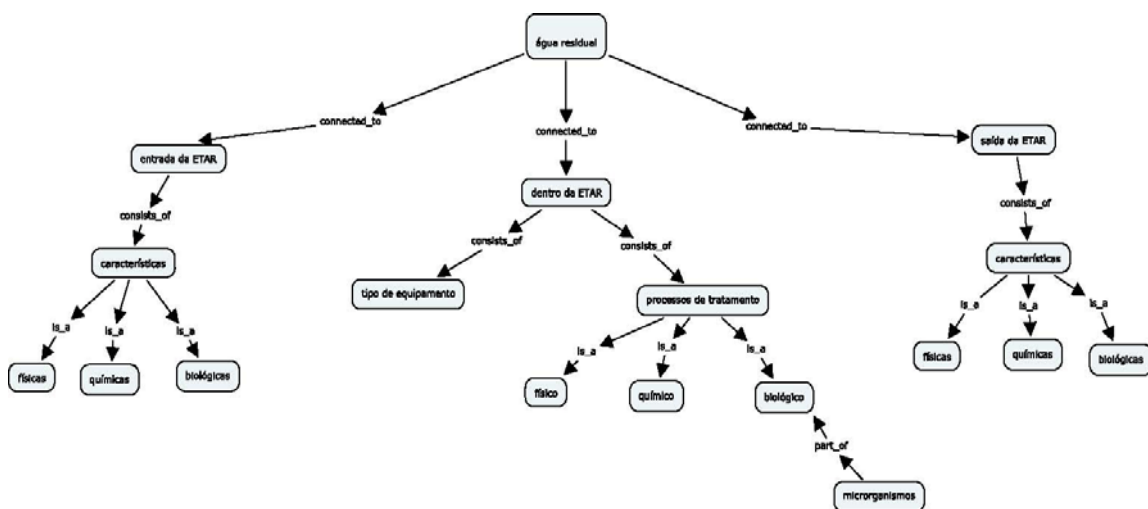
O passo inicial foi seleccionar um conjunto alargado de conceitos relacionados com o domínio em formato Excel. Esses conceitos passaram por um processo de filtragem manual e foram organizados em grandes grupos.

Para transpor essa informação para o mapa conceptual, foi pensada a seguinte conceptualização do domínio: como os tratamentos biológicos ocorrem em estações de tratamento, optámos por situá-los dentro de uma unidade designada por ETAR (acrónimo de Estação de Tratamento de Águas Residuais). Tendo igualmente como base conceptual

⁸³ Gerhard Budin, 2004.

subjacente uma sequência lógica de tecnologias e processos de tratamento dos efluentes no tempo e no espaço físico, dividiu-se o mapa em três áreas: 1. À entrada da ETAR; 2. Dentro da ETAR; 3. À saída da ETAR. A área ‘Dentro da ETAR’ foi subdividida em ‘Tipo de Equipamento’ e ‘Processos de Tratamento’.

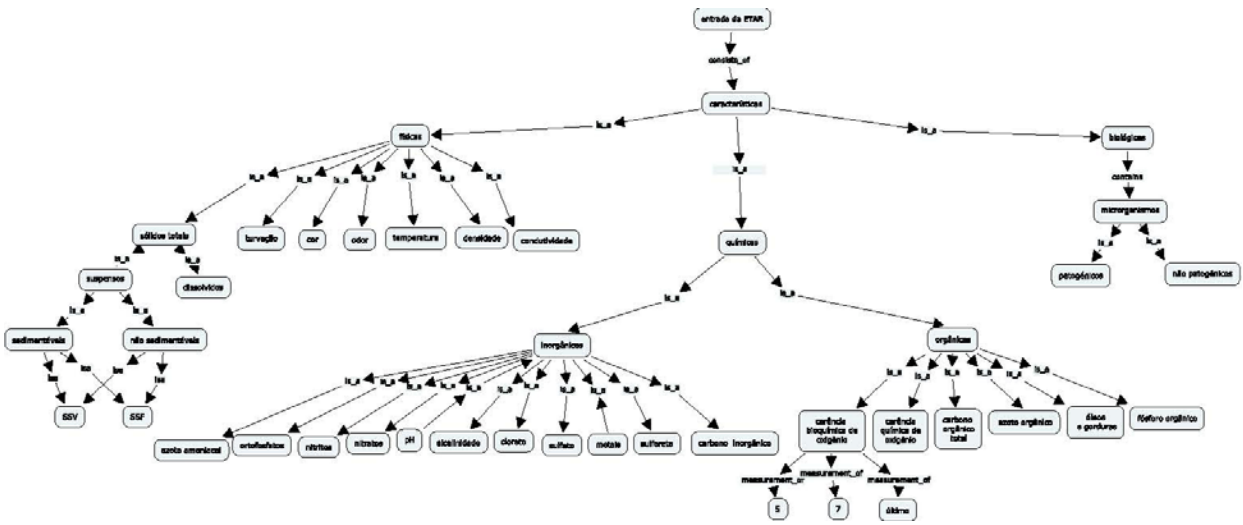
Figura 2 – Top Level Tree



A árvore central representa os conceitos mais gerais e superordenados do mapa conceptual. Apesar dos tratamentos biológicos de águas residuais estarem conceptualmente ligados aos tratamentos físicos e químicos, foi privilegiada sempre que possível a sua representação em detrimento das outras duas.

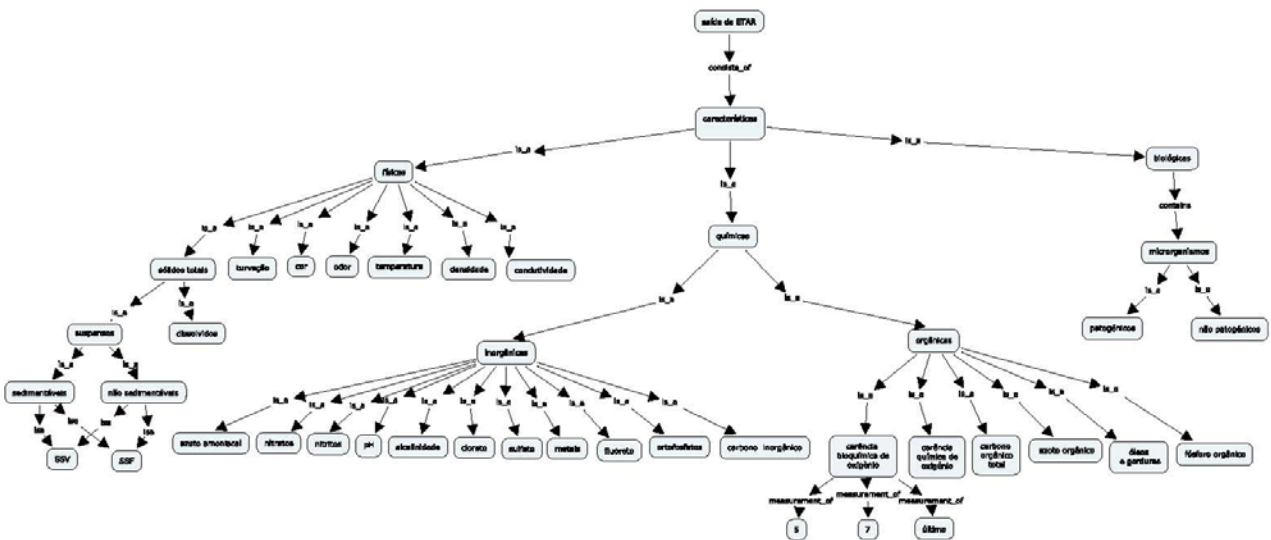
⁸⁴ À medida que avançávamos no grau de complexidade, os diversos mapas foram adquirindo dimensões difíceis de transpor para uma página A4. Serão, por isso, apresentados no Anexo a esta dissertação em formato legível.

Figura 3 – Entrada da ETAR



Na parte consignada à 'Entrada da ETAR' descreveram-se as propriedades da água residual em termos físicos, químicos e biológicos.

Figura 4 – Saída da ETAR



A parte 'Saída da ETAR' contém somente ligeiras diferenças em relação à parte 'Entrada da ETAR', uma vez que, na generalidade, os parâmetros para analisar a qualidade do efluente são idênticos.

A parte constituída pela área ‘Dentro da ETAR’ foi dividida da seguinte forma: Tipo de Equipamento e Processos de Tratamento. Os Processos de Tratamento foram divididos em Químico, Físico e Biológico, sendo que dentro do Biológico incluímos os Microrganismos.

No Tipo de Equipamento estão representados os conceitos relativos a alguns componentes físicos, partes essenciais do tratamento físico prévio que tem que ser efectuado, antes de se poder aplicar o tratamento biológico.

A parte respeitante aos Processos de Tratamento pretende representar os conceitos e suas relações por, simultaneamente, grau de complexidade e sequência. Este subdomínio provocou um debate interessante entre os especialistas envolvidos, dado que se trata de uma área do conhecimento bastante complexa, versátil e sujeita a uma actualização e renovação constantes, uma vez que a evolução tecnológica eliminou a estanquidade conceptual que existia há alguns anos. Hoje podemos combinar vários processos de tratamento obedecendo a sequências diversas, conforme o objectivo que se quer atingir no efluente a tratar. Por essa razão, conceitos como ‘tratamento primário’, ‘tratamento secundário’ e ‘tratamento terciário’, bastante comuns entre os especialistas, foram suprimidos do mapa conceptual.

Figura 5 – Tipo de Equipamento

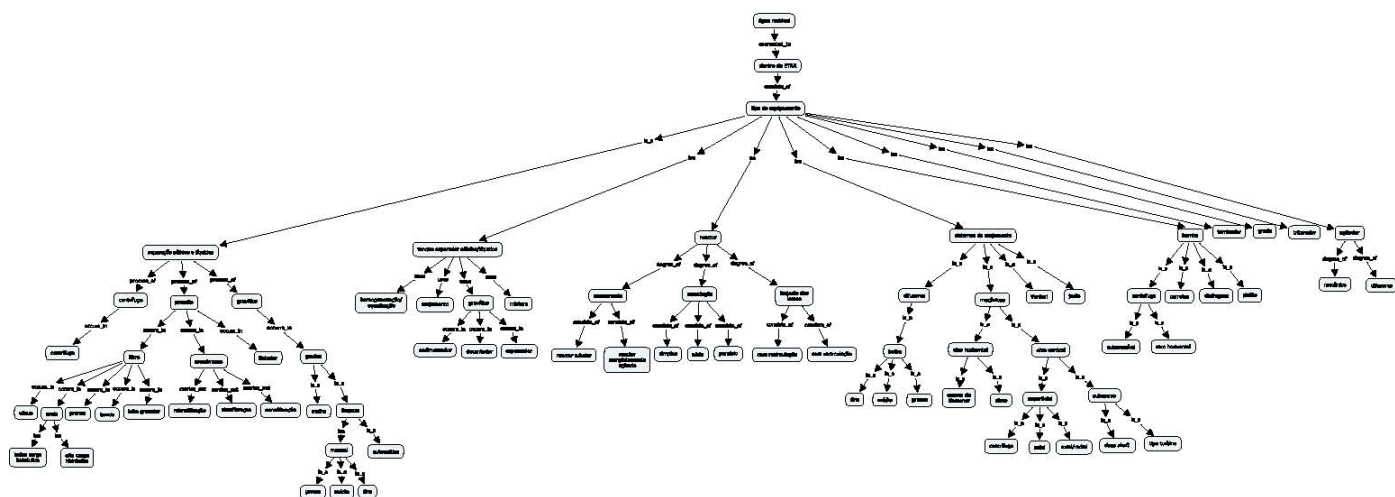


Figura 6 – Processos de Tratamento

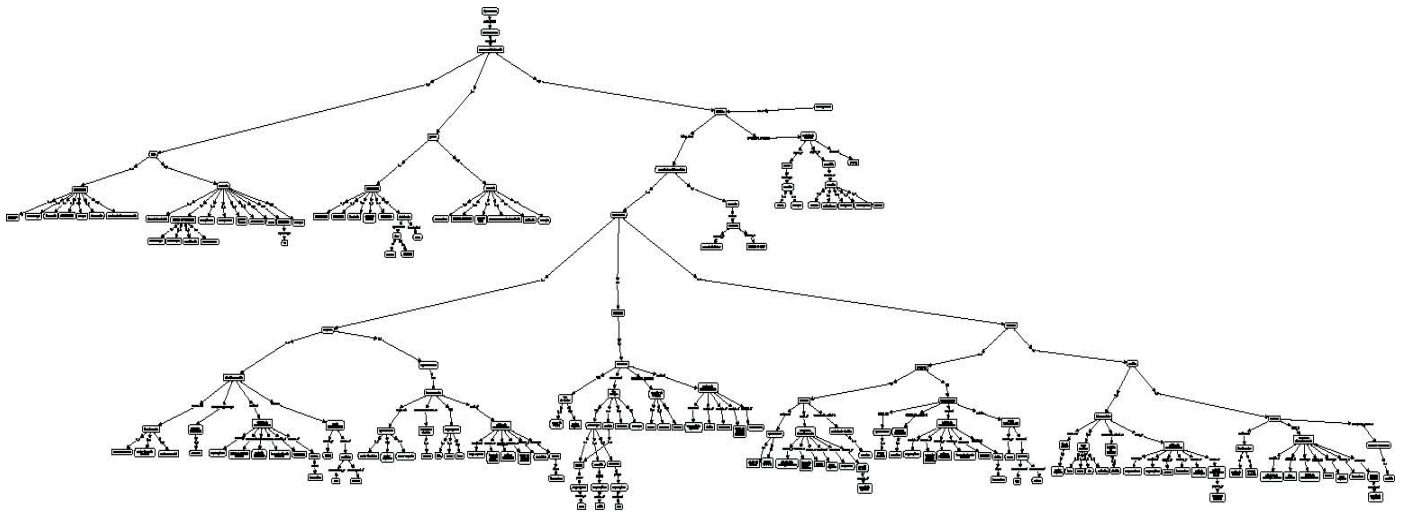
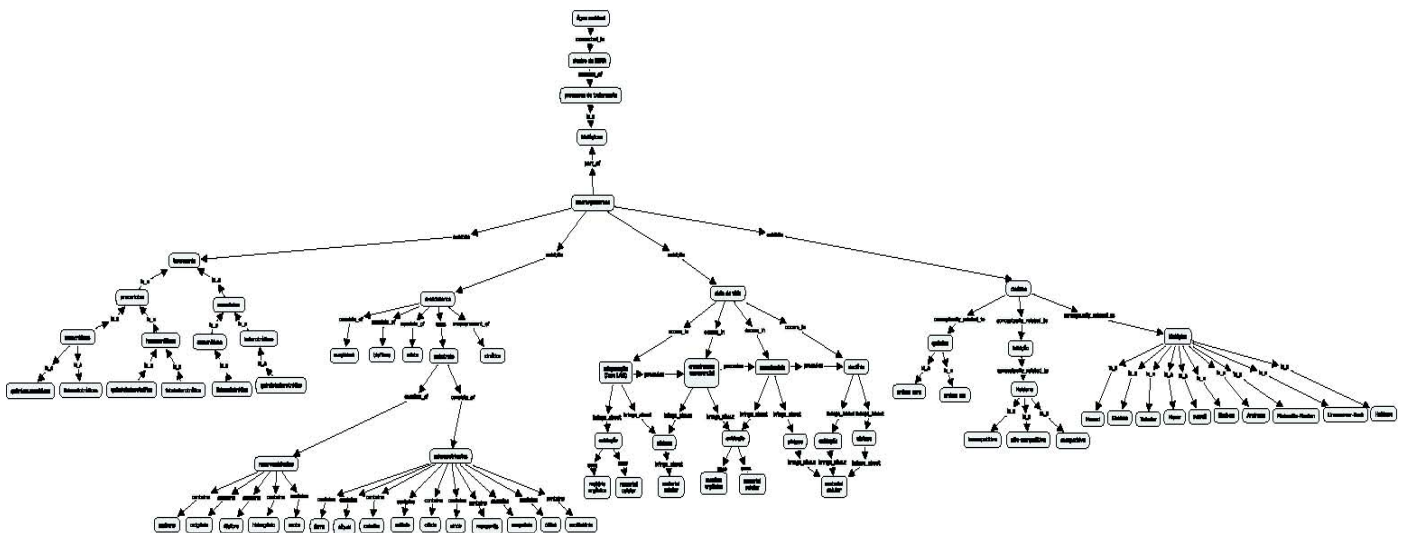


Figura 7 – Microrganismos



A construção dos mapas conceptuais exigiu longas horas de reflexão, na tentativa de agrupar de forma mais concisa e coerente possível o largo grupo de conceitos que podem representar este domínio de especialidade. Trata-se de um mapa conceptual que pretende representar tecnologias e processos envolvidos nos tratamentos biológicos de águas residuais. Não existiria ETAR se não existisse água residual, daí a opção por esse espaço onde decorrem as tecnologias e os processos.

Existem certamente conceitos cujo posicionamento poderia ser aperfeiçoado. Trata-se apenas de uma proposta que teria de ser sujeita a um conjunto mais vasto de opiniões. Contudo, a sua construção teve a preocupação de poder vir a ser um instrumento a utilizar pelos técnicos nas Estações de Tratamento de Águas Residuais. Assim, a construção do mapa e a classificação hierárquica e associativa utilizada reflectem uma conceptualização possível do domínio neste momento. O conceito com uma superordenação mais elevada é o de 'Água Residual', permitindo referenciar a parte física e química, necessária, por vezes, aos tratamentos biológicos de águas residuais.

A direcção das setas foi uma opção dos especialistas, sem preocupação com questões de ordem de representação formal. Verificou-se a necessidade de consultar regularmente as publicações de carácter geral para confirmar classificações de conceitos, a sua organização e posicionamento.

Durante a construção do mapa, constataram-se os seguintes factos e tomaram-se as seguintes decisões:

- Optámos por tratar como sinónimos os conceitos tecnologias e equipamentos, sistemas de tratamento e processos de tratamento.
- O mapa pretende analisar a linha da água, excluindo, por conseguinte, a linha de sólidos. Trata-se de uma decisão importante, uma vez que não vão ser representados, de forma exhaustiva, conceitos como por exemplo 'lamas', um dos produtos dos processos de tratamento. Quer a estabilização de lamas, quer os tratamentos biológico, aeróbio ou anaeróbio de lamas, não aparecem desenvolvidos. Numa Estação de Tratamento existem sempre lamas e o seu tratamento também pode ser químico,

físico ou biológico, com equipamentos associados. Esses sólidos perfazem uma nova linha dentro da ETAR. Note-se que, independentemente desta decisão, o conceito de recirculação de lamas faz parte integrante do processo de lamas activadas. Sem recirculação, o processo não funcionaria, encontrando-se, por isso, referenciado no mapa.

- Existem, naturalmente, inúmeros conceitos que apontam para conhecimento implícito. Podemos referir, a título de exemplo, que, relativamente ao mapa sobre 'Tipo de Equipamento', o especialista sabe que as bombas centrífugas de eixo horizontal encontram-se em poço seco e as submersíveis estão dentro do próprio líquido. Do mesmo modo, o conceito de 'difusor' está sempre associado à propriedade 'submerso'. Optámos por isso, em consciência, por omitir determinados conceitos que se tornariam redundantes. Outro exemplo que se poderia referenciar prende-se com o conceito de 'torre biológica'. Este conceito está implicitamente presente quando falamos de carga orgânica muito alta no leito percolador. Nesse caso, utiliza-se um material plástico, ou seja, uma torre biológica.
- Na designação dos microrganismos, optou-se pela classificação patogénicos e não patogénicos, em detrimento de entéricos e não entéricos, uma vez que estávamos mais preocupados com o efeito e não tanto com a proveniência.
- Nos processos avançados não foram referenciados os reactores avançados, que são, basicamente, modificações do reactor UASB. Esses reactores podem estar presentes nos tratamentos avançados ou convencionais. Os reactores, enquanto tecnologias, podem combinar-se de várias formas com os processos tratamento de base, ou seja, os convencionais. As designações de reactor contínuo e reactor descontínuo são conceitos superordenados e hiperónimos da maioria dos reactores.
- Foi também uma opção reflectida representar os diversos tipos de reactores com os termos em Português, muito embora os especialistas utilizem as siglas em Inglês. Para facilitar a correspondência de terminologia, elaborámos uma tabela com base em

informação fornecida pelo especialista, onde se pode verificar que o mesmo termo em Inglês pode corresponder a mais do que um termo em Português⁸⁵:

Tabela 1 – Reactores

SIGLA	INGLÊS	PORTUGUÊS
n/tem	batch reactor	reactor descontínuo
SBR	sequential batch reactor	reactor sequencial
PFR	plug flow reactor	reactor de pistão
		reactor tubular
		reactor tubular com fluxo tipo pistão
		reactor contínuo
FBR	fluidized biological reactor	reactor de leito fluidizado
FBBR	fluidized biological biofilm reactor	reactor biológico de leito fluidizado
MBR	membrane bio-Reactor	bio-reactor de membranas
		reactor de membranas
MBBR	membrane batch bio-reactor	bio-reactor de membranas sequencial
	moving bed biofilm reactor	reactor biológico de leito móvel
CSTR	continuous stirring tank reactor	reactor de mistura
		reactor completamente agitado
UASB	upflow anaerobic sludge blanket	digestor anaeróbio de fluxo ascendente
		reactor de manto de lamas
ASBR	anaerobic sequential batch reactor	reactor sequencial anaeróbio
RBC	rotating biological contactor	biodisco
AnRBC	anaerobic rotating biological contactor	biodisco anaeróbio
n/tem	anaerobic contact reactor	reactor anaeróbio de contacto
		digestor de contacto
AF	anaerobic filter	filtro aneróbio
TF	trickling filter	filtro percolador
PF	percolating filter	leito percolador
AS	activated sludge	lamas activadas

⁸⁵ Optou-se por colocar nesta e em outras tabelas todos os termos em letra minúscula, apesar de na literatura poder aparecer a designação com letra maiúscula, uma vez que, embora possa tratar-se de uma estratégia discursiva, não corresponde a uma regra ortográfica na língua inglesa e portuguesa.

Na totalidade do mapa foram utilizados 226 conceitos. A tabela que se segue exhibe somente os conceitos que aparecem três ou mais vezes no mapa conceptual total.

Tabela 2 – Mapa conceptual total: conceitos

Conceitos no mapa conceptual total	Frequência dos conceitos
carga orgânica	11
cinética	9
temperatura	9
aceitador de electrões	8
critérios de dimensionamento	7
tipo de reactor	7
aeróbio	6
anaeróbio	6
reactor de pistão	6
tempo de retenção hidráulico	6
alcalinidade	5
biológicas	5
média	5
microrganismos	5
reactor de mistura	5
síntese	5
água residual	4
características	4
dentro da ETAR	4
físicas	4
oxidação	4
químicas	4
razão de recirculação	4
alta	3
avançado	3
baixa	3
convencional	3
processos de tratamento	3
rugosidade específica	3
tempo de retenção de sólidos	3

Teríamos de considerar vários exemplos como não sendo conceitos, como ‘características’ ou ‘dentro da ETAR’, uma vez que não fariam parte da conceptualização do domínio, mas sim de uma estratégia de construção subjacente ao mapa conceptual para que fosse possível ligar as várias partes. Num projecto futuro de formalização, seria obrigatório proceder a substituições adequadas.

A tabela seguinte corresponde ao conjunto de relações empregues na totalidade do mapa conceptual por ordem decrescente de utilização.

Tabela 3 – Mapa conceptual total: relações

Tipo de relação no mapa conceptual total	Frequência das relações
is_a	234
consists_of	73
contains	20
occurs_in	17
brings_about	13
conceptually_related_to	12
carries_out	9
degree_of	9
uses	9
measurement_of	8
part_of	8
result_of	7
connected_to	6
exhibits	4
affects	3
ingredient_of	3
precedes	3
process_of	3
property_of	3
derivative_of	1

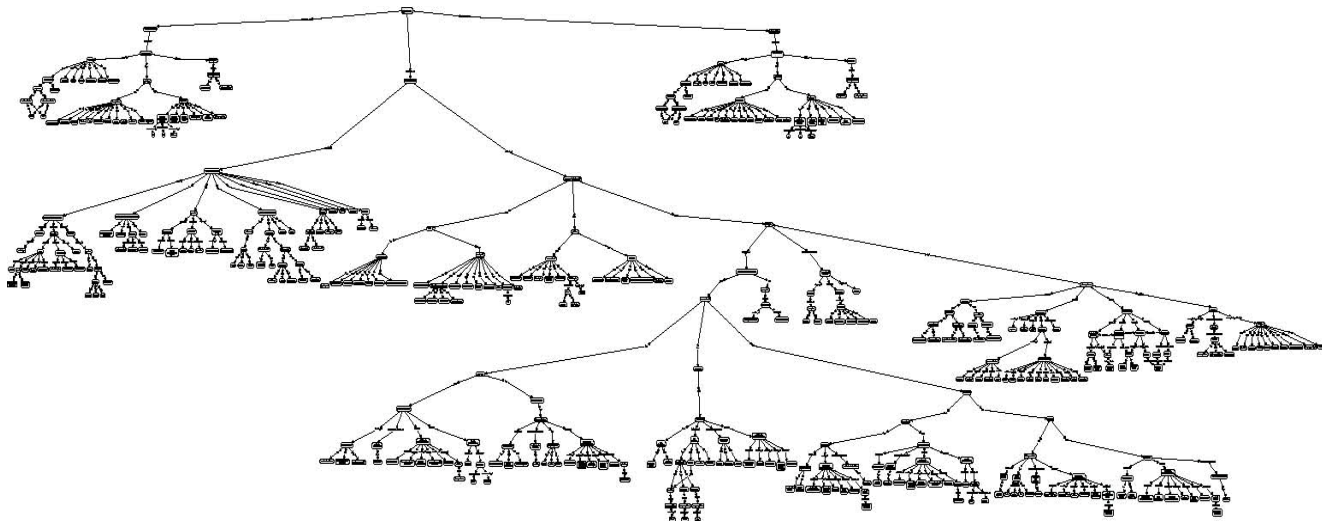
A relação genérica *is_a* destaca-se de todas as outras, confirmando o tipo de classificação tradicional nas ciências naturais.

As relações *consists_of*, *contains*, *part_of*, *connected_to* e *ingredient_of*, encontram-se muito próximas, pelo que, se considerarmos a sua soma, verificamos que esta relação hierárquica partitiva aparece 110 vezes, destacando-se como a segunda relação mais utilizada.

As relações de causa/efeito são *brings_about*, *occurs_in*, *carries_out*, *result_of*, *affects*, *process_of*, *uses*, *exhibits*, perfazendo 65 utilizações.

O trabalho desenvolvido em conjunto com os especialistas permitiu assim construir um mapa do domínio 'Tratamentos Biológicos de Águas Residuais', constituindo-se como uma proposta de organização conceptual que poderá, posteriormente, servir de base para projectos futuros, como, por exemplo, ao nível do Comité Técnico 207 da ISO:

Figura 8 – Mapa conceptual total



7. Proposta de construção de uma rede lexical

7.1 Perspectiva linguística

A perspectiva linguística encara o texto como ‘conhecimento em acção’. No texto podemos encontrar termos que, neste caso, se constituem maioritariamente como unidades lexicais. Interessa portanto analisar, por um lado, como se processa a manifestação do conhecimento através de elementos linguísticos presentes em discurso e, por outro, se essas manifestações contribuem de forma válida para a construção de uma rede lexical que reflecta uma possível organização conceptual do domínio.

Tal como na observação de dados na primeira parte, iremos construir um mapa graficamente semelhante ao utilizado pelos especialistas, contando para isso com os elementos a extrair de sequências de texto seleccionadas do corpus de especialidade e directamente relacionadas com as conceptualizações representadas no mapa conceptual sobre água residual anteriormente elaborado.

Partimos, assim, com vantagem analítica: embora não sendo especialistas, já possuímos algum conhecimento sobre o domínio, fruto do trabalho anteriormente realizado, encontrando-nos mais sensíveis em relação à selecção e utilização correcta do corpus e atentos à organização lexical que poderá advir da análise linguística.

A construção da rede lexical foi elaborada sem a colaboração dos especialistas, embora posteriormente tenha sido submetida à sua apreciação.

7.1.1 Análise do texto

A língua natural é ambígua por natureza, mas constitui-se como o elemento comunicativo mais eficiente na transmissão de conhecimento. Não temos conceitos no texto, temos termos e relações entre termos. Nesse sentido, vamos trabalhar sobre dados de natureza linguística.

Cruse afirma que as propriedades semânticas dos elementos lexicais se reflectem nas relações que estabelecem com os seus contextos actuais e potenciais, incluindo contextos situacionais extralinguísticos: “(...) first, the relation between a lexical item and extra-linguistic contexts is often crucially mediated by the purely linguistic contexts (...); second, any aspect of an extra-linguistic context can in principle be mirrored linguistically; and, third, linguistic context is more easily controlled and manipulated.”(Cruse, 1986, p. 1) Um termo isolado não pode ser verdadeiro ou falso, tem de co-existir com outras unidades lexicais para formar uma proposição, simples ou complexa. Assim, para o autor, “(...) the meaning of a word is fully reflected in its contextual relations (...)” (Cruse, 1986, p. 16) Os termos são analisados tendo em conta a situação contextual discursiva.

De acordo com Conceição, é através dos enunciados linguísticos que assistimos também a processos de reformulação, importantes para a construção de um sistema conceptual de domínio: “La reformulation, plus qu’une ré-expression d’une dénomination, est une nouvelle conceptualisation qui établit des liens (d’équivalence ou pas) sémantiques, cognitifs et même pragmatique entre deux unités de discours.” (Conceição, 2005, p. 74) Segundo o autor, os conceitos enquanto unidades cognitivas que representam conhecimento constroem-se no discurso através de redes semânticas estabelecidas pelas reformulações e pelas unidades linguísticas que designam e relacionam. (Conceição, 2005, p. 82) Estimulando os processos cognitivos de inferência e dedução, as reformulações completam-se, nomeadamente, pela interpretação do implícito, do que não se encontra no texto, desempenhando uma função importante, não só na construção do discurso, como na transmissão de conhecimento. O contexto determina em grande parte a delimitação das unidades terminológicas, sendo possível retirar do mesmo informações de vária ordem, linguística e extralinguística. Para Conceição, é no discurso que se constrói o conhecimento de domínio, permitindo uma aquisição de traços conceptuais a partir de dados textuais:

“L’organisation conceptuelle exprimée par le domaine est la condition *sine qua non* pour qu’une unité lexical soit considérée comme terme, puis qu’elle ne le devient que parce qu’elle appartient à un domaine dans lequel elle désigne un concept et a une définition. Cette organisation de concepts est donc aussi une structuration de connaissances, une conceptologie.”(Conceição, 2005, pp. 136-137)

Para Sager, os conceitos são construções da mente humana, produtos de processos cognitivos que ajudam na classificação dos objectos de forma sistemática e abstracta. A estrutura conceptual dos domínios de especialidade distingue-se, entre outras coisas, pela natureza dos conceitos que predominam nessa área do conhecimento. Sager entende assim que a teoria terminológica ordena os conceitos de acordo com um esquema de classificação conceptual, onde estes são caracterizados pelas relações que estabelecem: “(...) terminology relates terms to concepts (and not vice versa) and is therefore not concerned with absolute conceptual systems but only with systems created for the specific purpose of assisting communication.” (Sager, 1990, p. 29) Esse conteúdo cognitivo pode realizar-se linguisticamente através de um código perceptível e interpretável pelos locutores envolvidos no acto de comunicação. No caso concreto, o código presente no texto é a língua natural.

Segundo Barros (2007), um corpus pode constituir um sistema conceptual e as relações semânticas podem ser consideradas como um mecanismo essencial de estruturação do léxico. Referenciando Cruse, Barros comenta que as relações podem ser paradigmáticas e sintagmáticas. As primeiras reflectem o modo como a realidade é apreendida através de categorizações, as segundas estão ligadas à estruturação do discurso, manifestando a forma como as unidades lexicais se relacionam para criar significação. As relações semânticas ajudam a estruturar o léxico de uma língua, uma vez que funcionam como mecanismo linguístico, mas também constituem a base da construção de um sistema conceptual.

Nas últimas décadas assistimos à construção de inúmeras ferramentas de processamento de língua natural. A extracção automática de candidatos a termo e de relações léxico-semânticas entre candidatos a termo constitui uma tarefa automatizada e de fácil implementação. Esse pré-estágio semântico pode constituir um dado importante na construção de uma ontologia. Vamos por isso centrar a nossa análise do corpus em relações semânticas e/ou marcadores linguísticos que poderão contribuir para a construção de um mapa conceptual graficamente semelhante ao construído pelos especialistas. A nossa hipótese teórica é de que o mapa daqui resultante irá conter manifestações linguísticas de conhecimento que se aproximarão das conceptualizações reflectidas no mapa conceptual anteriormente elaborado.

7.1.2 Relações semânticas, marcadores linguísticos e reformulações

A Norma ISO/FDIS 704 – 2000, *Terminology Work – Principles and Methods* estabelece as seguintes relações para modelar um sistema conceptual: i) relações hierárquicas (genéricas e partitivas); ii) relações associativas. De acordo com a mesma Norma, “(...) in a hierarchical relation, concepts are organized into levels where the superordinate concept is subdivided into at least one subordinate concept.” (ISO, 2000b, p. 6) Uma relação genérica existe entre dois conceitos “(...) when the intension⁸⁶ of the subordinate concept includes the intension of the superordinate concept plus at least one additional delimiting characteristic” (ISO, 2000b, p. 6), enquanto uma relação partitiva existe “(...) when the superordinate concept represents a whole, while the subordinate concept represents parts of that whole.” (ISO, 2000b, p. 6) A relação partitiva é designada também na Norma ISO/FDIS 1087 2000 Parte I por ‘part-whole relation’.

As relações associativas são relações não-hierárquicas. De acordo com a Norma ISO/FDIS 704 – 2000, temos uma relação associativa “(...) when a thematic connection can be established between concepts by virtue of experience.” (ISO, 2000b, p. 12) A Norma ISO/FDIS 1087 2000 Parte I designa igualmente as relações associativas de ‘relações pragmáticas’ (ISO, 2000c, p. 5), dividindo-as em sequenciais, temporais e causais. Sager propõe em *A Practical Course in Terminology Processing* (1990) um conjunto de fórmulas para as relações de tipo hierárquico genéricas e partitivas, a saber:

Relações genéricas (Sager, 1990, p. 30):

Tabela 4 – Relações genéricas

Relações genéricas
<i>X is a type of A</i>
<i>X, Y and Z are types of A</i>
<i>A has the specific concepts X, Y and Z</i>
<i>A has the subtype X</i>

⁸⁶ A Norma ISO/FDIS 1087-1 de 2000 fornece as seguintes definições: “Intension – set of characteristics which makes up the concept.(...) Extension – totality of objects to which a concept corresponds.”(ISO, 2000c, p. 3)

Relações partitivas (Sager, 1990, p. 32):

Tabela 5 – Relações partitivas

Relações partitivas
<i>X is a constituent part of Y</i>
<i>X, Y and Z are constituent parts of A</i>
<i>A consists of X</i>
<i>A consists of X, Y and Z</i>

Sager considera ainda um outro tipo de relações, designadas por relações complexas, dado que “(...) concepts are often seen as being inter-related in a complex manner which cannot be conveniently captured by straightforward generic and partitive structures.” (Sager, 1990, p. 34)

Propõe então a seguinte lista de tipos de relações complexas:

Tabela 6 – Relações complexas

Relações complexas	
<i>is caused by</i>	<i>cause-effect</i>
<i>is a product of</i>	<i>material-product</i>
<i>is a property of</i>	<i>material-property</i>
<i>is a quantitative measure of</i>	<i>material-state</i>
<i>is an instrument for</i>	<i>process-product</i>
<i>is a counteragent of</i>	<i>process-instrument</i>
<i>is a container for</i>	<i>process-method</i>
<i>is a method of</i>	<i>phenomenon-measurement</i>
<i>is a material for</i>	<i>object-counteragent</i>
<i>is a place for</i>	<i>object-container</i>
	<i>object-material</i>
	<i>object-quality</i>
	<i>object-operation</i>
	<i>object-characteristic</i>
	<i>object-form</i>
	<i>activity-place</i>

As reformulações presentes nos enunciados fornecem igualmente informações de carácter linguístico e extralinguístico, uma vez que, segundo Conceição, são as segundas formulações de conteúdo cognitivo e linguístico da primeira formulação. (Conceição, 2005, p. 73) Através das reformulações criam-se ligações de equivalência entre X, o reformulado, e Y, a nova

formulação. Podemos partir de axiomas simples para detectar, por exemplo, fenómenos de sinonímia, como referido no quadro seguinte:

Tabela 7 – Axiomas

Axiomas
X = Y
X > Y
X < y
X diferente de Y

Para Conceição, as equivalências podem ser fornecidas pela definição textual, que é forçosamente uma reformulação. As reformulações são introduzidas por marcadores de identificação, tais como verbos, advérbios ou expressões de diferenciação, entre outros. Os marcadores lexicais agem como indicadores de relações semânticas (“Ces marqueurs de reformulations, aussi appelés connecteurs reformulatifs (...)” (Conceição, 2005, p. 91)) O estatuto dos marcadores de reformulação depende da utilização concreta em cada situação e a sua tipologia tem em conta a natureza sintáctica dos elementos que a constituem. Esta abordagem considera fundamental para a análise terminológica não somente os termos, mas também outras unidades lexicais que, com eles, constituem o discurso, pois “(...) la recherche en terminologie est un travail sur la langue en tant que système et non seulement sur certaines unités de ce système que, d’ailleurs, ne peuvent être utilisées en communication qu’avec d’autres unités.” (Conceição, 2005, p. 163)

Baseados em (Conceição, 2005), foi possível construir a seguintes tabela:

Tabela 8 – Marcadores linguísticos: conectores discursivos e verbos

Conectores discursivos	Verbos
alguns autores	abranger
aliás	aplicar
apenas	aplicar
apesar de	causar
assim	chamar
bem como	classificar
como se disse	consistir
conforme	constituir
dado que	dar/ter o nome de
de uma maneira geral	denominar
distingue-se de	designar
dos quais	diferir
e	distinguir
em resumo	dizer
em síntese	entender-se por
em suma	enumerar
em termos de	estudar
embora	incluir
este, esta, estes, estas	nomear
excepto	ocupar-se de
inclusive	pertencer
isto é	poder ser
já que	provir
logo	provocar
mas	referir-se a
melhor dizendo	remeter para
na verdade	ser
não... mas	significar
no entanto	
numa palavra	
o que quer dizer	
o seguinte	
ou	
ou seja	
pelo contrário	
por outras palavras	
quer dizer	
resumindo	
um tipo de	

A pontuação é igualmente considerada um marcador linguístico importante.

Baseados nos pressupostos teóricos anteriormente expostos, iremos proceder a uma análise de sequências extraídas de texto do domínio de interesse recorrendo à tipologia de relações e marcadores linguísticos apresentados.

7.1.3 Metodologia

Pretende-se analisar o mesmo domínio de especialidade: “Tratamentos Biológicos de Águas Residuais”. O objectivo consiste em extrair dados do texto de especialidade que permitam a construção de uma rede lexical composta por termos e relações entre termos. A selecção e delimitação do domínio encontram-se justificadas, uma vez que esta abordagem decorre da investigação anterior.

A pesquisa de fontes de informação foi facilitada pela construção anterior do mapa conceptual por parte dos especialistas. Detínhamos dados suficientes para pesquisar um corpus⁸⁷ electrónico no domínio em causa em repositórios em linha nas diferentes universidades.

Entre as inúmeras teses disponíveis, foram seleccionadas quatro sobre tratamento de efluentes vinícolas em formato electrónico, a saber:

1. Maria Cândida da Silva Moreira, *Tratamento de Efluentes Vinícolas* – Dissertação de Mestrado em Tecnologia do Ambiente, Universidade do Minho, 1995.
2. Pedro Miguel Gonçalves, *Tratamento de Efluentes Vinícolas por Digestão Anaeróbia* – Dissertação de Mestrado em Tecnologia do Ambiente, Universidade do Minho, 1998.
3. Margarida Louro da Fonseca Pinho, *Aplicabilidade do reactor MBBR no tratamento de efluentes vínicos* – Dissertação de Mestrado, Universidade de Aveiro, 2007.

⁸⁷ Definições de corpus: “Corpus designe, de nos jours, un grand ensemble de données textuelles, normalement informatisées, pour que le traitement des informations, que peuvent en être extraites, soit facilite ou même simplement possible.” (Conceição, 2005, p. 124); Na Norma ISO 1087-2, 2000 – *Terminology work – Vocabulary, Part 2: Computer applications*, um corpus é definido como “(...) a systematic collection of machine-readable texts or parts of text prepared, coded and stored according to predefined rules.” Norma ISO 1087-2, 2000 (ISO, 2000a, p. 2)

4. Elsa Raquel Lages Almeida, *Avaliação da Biodegradabilidade Aeróbia de Efluentes Vinícolas* – Dissertação de Mestrado, Universidade de Aveiro, 2008.

Optámos pela análise de sequências de texto presentes na tese número 4, seleccionando como critério a data de defesa da dissertação.

7.1.3.1 Construção da rede lexical

Pretendia-se proceder a uma análise lexical e semântica de sequências do texto seleccionado. Esse processo de selecção socorreu-se das seguintes ferramentas informáticas: WordSmith, Concapp e ExtracTerm. Utilizaram-se as funcionalidades de concordâncias, listas de palavras e extracção de candidatos a termo baseada em regras morfossintácticas, tanto por ordem alfabética, como por frequência.

Os dados extraídos permitiram compor manualmente uma lista de candidatos a termo, mono- e multi-palavra, extraídos da tese *Avaliação da Biodegradabilidade Aeróbia de Efluentes Vinícolas*⁸⁸:

Tabela 9 – Candidatos a termo a partir do corpus: siglas

Candidatos a Termo – Siglas (selecção manual)	
FBBR	fixed bed biofilm reactor
RBC	rotating biological contactor
SBR	sequential batch reactor
MBBR	moving bed biofilm reactor
SSV	sólidos suspensos voláteis
CQO	carência química de oxigénio
gSSV	gramas sólidos suspensos voláteis
gCQO	gramas carência química de oxigénio

⁸⁸ Algumas tabelas são constituídas por duas colunas unicamente por razões de melhor visualização da mancha gráfica.

Tabela 10 – Candidatos a termo a partir do corpus

Candidatos a Termo (selecção manual)	
aeróbio	leveduras
água	macrófitas
anaeróbio	membranas
areia	metabolismo
atmosfera	microrganismos
bactérias	minerais
biofilme	moléculas
biofiltro	nutrientes
biológico	oxigénio
biomassa	partículas
bioquímica	processo
carbono	protozoários
carga	reactor
caudal	recirculação
concentração	retenção
decantador	saibros
decantação	sedimentação
degradação	sedimentador
depuração	separação
efluente	sequência
fertilizantes	sistema
filme	sólidos
filtração	solo
filtros	superfície
fungos	suportes
hidráulico	suspensão
hidrogénio	tanque
inertes	tratamento

Tabela 11 – Candidatos a termo multi-palavra a partir do corpus

Candidatos a Termo multi-palavra (selecção manual)	
águas residuais	nutrientes excedentários
águas residuais industriais	partículas em suspensão
águas residuais urbanas	processo de depuração biológico
alimentação contínua	processo de filtração por membranas
alimentação descontínua	processo de tratamento
armazenamento arejado	processo de tratamento biológico
bacia arejada	processo de tratamento biológico aeróbio
biomassa em suspensão	processo de tratamento biológico aeróbio extensivo
biomassa suspensa	processo de tratamento biológico aeróbio intensivo
bio-reactor	produção de lamas
bio-reactor de membranas	propriedades depurativas
caudal de recirculação	propriedades filtrantes
caudal de recirculação de lamas	reactor arejado
concentração de oxigénio	reactor de biofilme
culturas agrícolas	recirculação de lamas
culturas fixas	rotating biological contactor
degradação da matéria orgânica	sequential batch reactor
degradação da matéria orgânica do efluente	sistema de lamas activadas com sedimentador e recirculação
depuração da matéria orgânica	sistema de lamas activadas sem sedimentador
dióxido de carbono	sistema hidráulico
efluente tratado	sólidos em suspensão
energia bioquímica	tanque de arejamento
fertilizantes minerais	técnica de tratamento de efluentes
filme fixo	tecnologias de biomassa em suspensão
fixed bed biofilm reactor	tecnologias de filme fixo
função fotossintética	tempo de retenção hidráulico
lagoa arejada	tratamento aeróbio
lamas activadas	tratamento aeróbio extensivo
leito de macrófitas emergentes	tratamento aeróbio intensivo
leito percolador	tratamento anaeróbio
matéria orgânica	tratamento biológico aeróbio
moléculas de água	tratamento biológico anaeróbio
moving bed biofilm reactor	tratamento biológico com alimentação sequencial
moving bed reactor	tratamento por lamas activadas e alimentação descontínua

Utilizando a ferramenta Wordsmith4, procedeu-se a vários testes de levantamento de concordâncias, utilizando não só candidatos a termo, como também relações semânticas (genéricas, partitivas e associativas) e outro tipo de marcadores linguísticos que denotassem contextos de reformulação. Selecionou-se um conjunto de candidatos a termo directamente relacionados com o princípio orientador do mapa conceptual anteriormente elaborado:

Tabela 12 – Candidatos a termo a partir do corpus: frequência

Candidato a termo	Frequência
remoção	132
tratamento	124
degradação	113
processo	100
concentração	68
eficiência	38
operação	37
parâmetros	23

Optou-se pelo candidato a termo *processo*, tendo-se procedido ao levantamento dos contextos no documento a analisar. Seguidamente, e na tentativa de não nos distanciarmos do mapa conceptual já construído, foram seleccionadas algumas sequências de texto versando sobre 'Tratamento Biológico Aeróbio' que a seguir se transcrevem:

TRATAMENTO BIOLÓGICO AERÓBIO

Nos processos de tratamento aeróbios, a matéria orgânica, que corresponde à carga poluente, é oxidada levando à formação de dióxido de carbono e à libertação de hidrogénio. Este, por sua vez, reage com o oxigénio levando à produção de moléculas de água. Todo este processo é caracterizado por produzir uma elevada quantidade de energia bioquímica, armazenada na forma molecular, que é utilizada pela biomassa para se reproduzir (com a consequente produção de lamas) e pelo próprio metabolismo do processo (produção de CO₂). (E. R. L. Almeida, 2008, p. 26)

Os processos de tratamento biológico aeróbios podem ser classificados em intensivos ou extensivos, tal como os processos anaeróbios já apresentados. Nos processos intensivos, os microrganismos necessitam de fornecimento artificial de oxigénio, através de sistemas mecânicos e recorrendo a tecnologias de filme fixo ou biomassa em suspensão, enquanto os processos extensivos utilizam o oxigénio que existe na atmosfera, independentemente do modo como o processo de tratamento se processa (...). (E. R. L. Almeida, 2008, p. 26)

Os processos de tratamento aeróbio extensivos que se destacam são: - distribuição no solo (...); - armazenamento arejado que é uma adaptação do processo de depuração biológico conhecido

por lamas activadas. É caracterizado por possuir um elevado tempo de retenção hidráulico, ausência de recirculação de lamas e reduzida necessidade de mão-de-obra especializada (...). Neste processo de tratamento ocorre uma sucessão de etapas de onde decorre o desenvolvimento de vários microrganismos, como leveduras, fungos, bactérias e protozoários, e ao declínio e morte desta mesma biomassa, quando a degradação da matéria orgânica do efluente está praticamente completa. O sistema é caracterizado por funcionar de modo descontínuo, geralmente designado por batch, com três fases distintas (arejamento, sedimentação e decantação), todas elas processadas no mesmo local (...); - leito de macrófitas emergentes (...). (E. R. L. Almeida, 2008, p. 27)

Os processos de tratamento aeróbio intensivos têm por objectivo o conhecimento e a estabilização da razão matéria orgânica e biomassa (F/M) a aplicar ao processo biológico, conduzindo ao equilíbrio do funcionamento do reactor. A maior parte destes processos são baseados no sistema das lamas activadas, a técnica de tratamento de efluentes mais utilizada no tratamento de águas residuais urbanas e industriais. Este tipo de processos apresenta duas classes que se distinguem pela presença ou não de sedimentador que separe as lamas e um circuito que assegure a sua recirculação. Assim, as variantes deste processo de tratamento são as seguintes: (E. R. L. Almeida, 2008, p. 27)

-sistema de lamas activadas com sedimentador e recirculação, caracterizado por possuir duas fases distintas: uma fase em que a matéria orgânica é degradada por acção da biomassa num reactor arejado e uma segunda fase, em que a biomassa é separada do efluente tratado, através de um sedimentador. Este processo utiliza microrganismos mantidos em suspensão numa bacia arejada e alimentada de forma contínua com efluente e biomassa. Este processo, ao operar de forma contínua, contribui para que haja uma elevada percentagem de sólidos à saída do reactor, o que é colmatado com a passagem deste efluente por um decantador (...). A eficiência deste processo depende muito do caudal de recirculação de lamas para o tanque de arejamento e também da concentração de oxigénio no meio. (...) (E. R. L. Almeida, 2008, p. 28)

-sistema de lamas activadas sem sedimentador é caracterizado por possuir, tal como o sistema anteriormente apresentado, duas fases de tratamento distintas: uma primeira em que a matéria orgânica é degradada por acção da biomassa num reactor arejado e uma segunda fase, em que a biomassa é separada do efluente tratado através da sua decantação na ausência de arejamento. A grande diferença deste sistema com o anterior é que a alimentação ao sistema é efectuada de forma descontínua. As técnicas de tratamento por lamas activadas e alimentação descontínua são: tratamento biológico com alimentação sequencial sequential batch reactor (SBR) e o bio-

reactor de membranas. (...) O bio-reactor de membranas é uma técnica de tratamento que deriva do processo de filtração por membranas. Consiste na associação de um reactor biológico a um sistema de separação do efluente tratado por filtração que é conseguido através de membranas. Estas têm a função de actuar como barreiras para a matéria orgânica e microrganismos, o que proporciona uma maior separação do efluente tratado da biomassa e conseqüentemente, o aumento da concentração desta no reactor (...). (E. R. L. Almeida, 2008, p. 28)

7.1.3.1.1 Sequências de texto

Com base nos autores e obras anteriormente citados bem como nas suas propostas de relações semânticas e marcadores linguísticos, foi feita uma anotação manual das sequências, onde ML corresponde a Marcador Linguístico (a vermelho) e TR corresponde a Tipo de Relação (a azul):

Nos processos de tratamento aeróbios, a matéria orgânica, que [ML corresponde à] [TR generic] carga poluente, [ML é] [TR cause/effect] oxidada [ML levando à] [TR cause/effect] formação de dióxido de carbono [ML e à] [TR cause/effect] libertação de hidrogénio. Este, por sua vez, [ML reage com] [TR reacts with] o oxigénio [ML levando à] [TR cause/effect] produção de moléculas de água. Todo este processo [ML é caracterizado por] [TR process/product] produzir uma elevada quantidade de energia bioquímica, armazenada [ML na forma] [TR generic] molecular, que [ML é utilizada pela] [TR is a material for] biomassa [ML para] [TR cause/effect] se reproduzir ([ML com a conseqüente] [TR cause/effect]) produção de lamas) [ML e] [TR cause/effect] pelo próprio metabolismo do processo (produção de CO₂).

Os processos de tratamento biológico aeróbios [ML podem ser classificados em] [TR generic] intensivos ou extensivos, [ML tal como] [TR generic] os processos anaeróbios [ML já apresentados]. Nos processos intensivos, os microrganismos [ML necessitam de] [TR process/product] fornecimento artificial de oxigénio, [ML através de] [TR is caused by] sistemas mecânicos [ML e] [TR process/product] recorrendo a tecnologias de filme fixo [ML ou] [TR process/product] biomassa em suspensão, enquanto os processos extensivos [ML utilizam] [TR is a material for] o oxigénio que existe na atmosfera, independentemente do modo como o processo de tratamento se processa.

Os processos de tratamento aeróbio extensivos que se destacam [ML são] [Tr generic]: - distribuição no solo (...), armazenamento arejado (...) leito de macrófitas emergentes. (...) -armazenamento arejado [ML que é uma adaptação do] [TR generic] processo de depuração biológico [ML conhecido por] [TR generic] lamas activadas. [ML É caracterizado por] [TR partitive] possuir um elevado tempo de retenção hidráulico [ML ,] [TR partitive] ausência de recirculação de lamas [ML e] [TR partitive] reduzida necessidade de mão-de-obra especializada (...). Neste processo de tratamento [ML ocorre] [TR partitive] uma sucessão de etapas [ML de onde decorre] [TR is a product of] o desenvolvimento de vários microrganismos, [ML como] [TR process/product] leveduras [ML ,] [TR process/product] fungos [ML ,] [TR process/product] bactérias [ML e] [TR process/product] protozoários, [ML e] [TR is a product of] ao declínio e morte desta mesma biomassa, [ML quando][TR phenomenon/measurement] a degradação da matéria orgânica do efluente [ML está] praticamente completa. O sistema [ML é caracterizado por] [TR process/method] funcionar de modo descontínuo, [ML geralmente designado por] [TR generic] batch, [ML com três fases distintas] [TR process/method] (arejamento, sedimentação e decantação), todas elas processadas no mesmo local (...).

[Os processos de tratamento aeróbio intensivos têm por objectivo o conhecimento e a estabilização da razão matéria orgânica e biomassa (F/M) a aplicar ao processo biológico, conduzindo ao equilíbrio do funcionamento do reactor.] A maior parte destes processos são baseados no sistema das lamas activadas [ML ,][TR generic] a técnica de tratamento de efluentes mais utilizada no tratamento de águas residuais urbanas e industriais. Este tipo de processos apresenta duas classes que se distinguem pela presença ou não de sedimentador que separe as lamas e um circuito que assegure a sua recirculação. Assim, as variantes deste processo de tratamento [ML são as seguintes] [TR generic]: -sistema de lamas activadas com sedimentador e recirculação, [ML caracterizado por] [TR partitive] possuir duas fases distintas: uma fase em que a matéria orgânica [ML é degradada por acção da] [TR is caused by] biomassa [ML num] [TR object/container] reactor arejado [ML e] [TR partitive] uma segunda fase, [ML em] [TR material/state] que a biomassa é separada do efluente tratado, [ML através de] [TR object/container] um sedimentador. Este processo [ML utiliza] [TR is a product of]

microrganismos mantidos em suspensão [ML numa] [TR object/container] bacia arejada e alimentada [ML de forma] [TR process/instrument] contínua [ML com] [TR is a material for] efluente e biomassa. Este processo, ao operar de forma contínua, [ML contribui para que haja] [TR cause/effect] uma elevada percentagem de sólidos à saída do reactor, [ML o que é colmatado com] [TR material/method] a passagem deste efluente por um decantador (...). A eficiência deste processo [ML depende muito] [TR cause/effec] do caudal de recirculação de lamas [ML para] [TR is a place for] o tanque de arejamento [ML e também] [TR cause/effect] da concentração de oxigénio no meio.

- sistema de lamas activadas sem sedimentador [ML é caracterizado por] [TR partitive] possuir, tal como o sistema anteriormente apresentado, duas fases de tratamento distintas: uma primeira em que a matéria orgânica [ML é degradada] [TR process/product] [ML por acção da] [TR is caused by] biomassa [ML num] [TR object/container] reactor arejado e uma segunda fase, em que a biomassa [ML é separada] [TR material/state] do efluente tratado [ML através] [TR is a method of] da sua decantação na ausência de arejamento. A grande diferença deste sistema com o anterior [ML é que] [TR process/method] a alimentação ao sistema [ML é efectuada de forma] descontínua. As técnicas de tratamento por lamas activadas e alimentação descontínua [ML são] [TR generic]: tratamento biológico com alimentação sequencial - sequential batch reactor (SBR) [ML e] [TR generic] o bio-reactor de membranas.

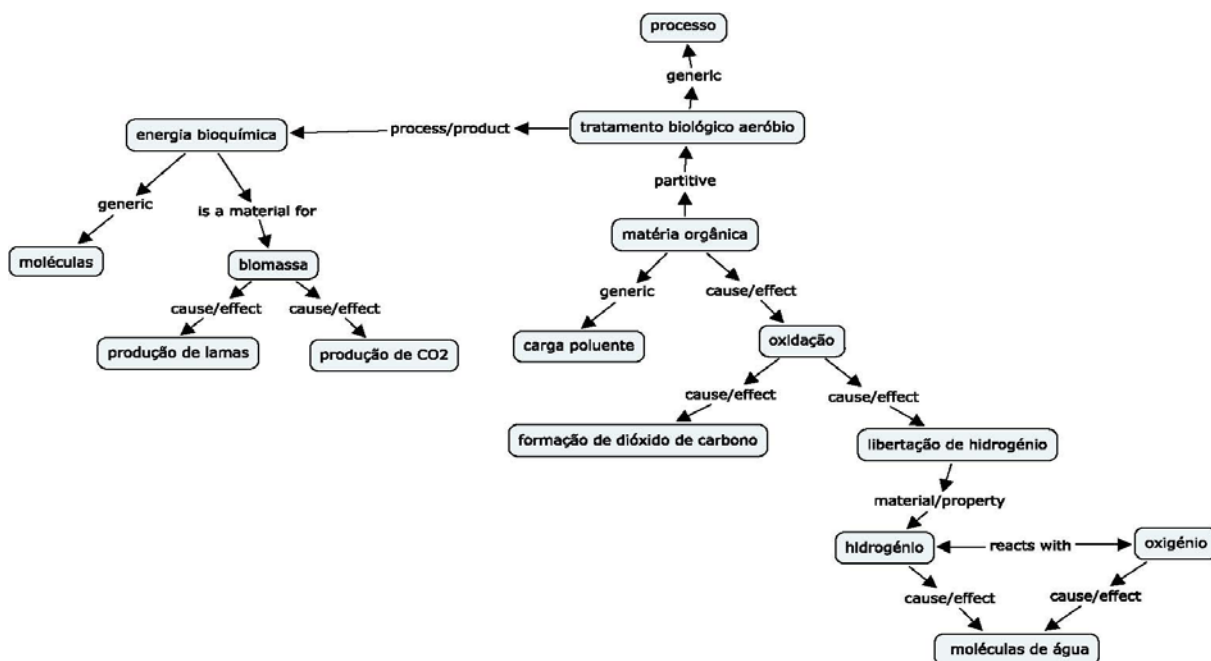
O bio-reactor de membranas [ML é] [TR generic] uma técnica de tratamento [ML que deriva] [TR is a product of] do processo de filtração por membranas. [ML Consiste na associação de um] [TR partitive] reactor biológico [ML a um] sistema de separação do efluente tratado por filtração [ML que é conseguido através de] [TR is a material for] membranas. Estas [ML têm a função de actuar como] [TR generic] barreiras para a matéria orgânica e microrganismos, [ML o que proporciona] [TR cause effect] uma maior separação do efluente tratado da biomassa [ML e conseqüentemente], o aumento da concentração desta no reactor (...).

Para não nos distanciarmos da metodologia seguida na construção do mapa conceptual sobre água residual, optámos por aplicar igualmente a ferramenta informática CMapTools na construção gráfica da rede lexical.

Com base na análise anteriormente apresentada relativamente aos marcadores linguísticos e reformulações, foram construídos vários mapas compostos por redes lexicais correspondentes a sequências ou partes de sequências, como a seguir se demonstra:

Sequência 1: *Nos processos de tratamento aeróbios, a matéria orgânica, que corresponde à carga poluente, é oxidada levando à formação de dióxido de carbono e à libertação de hidrogénio. Este, por sua vez, reage com o oxigénio levando à produção de moléculas de água. Todo este processo é caracterizado por produzir uma elevada quantidade de energia bioquímica, armazenada na forma molecular, que é utilizada pela biomassa para se reproduzir (com a conseqüente produção de lamas) e pelo próprio metabolismo do processo (produção de CO₂).*

Figura 9 – Sequência 1



A rede lexical traduz um processo cognitivo de transposição entre os dados de natureza linguística e os dados de natureza conceptual graficamente representados através da figura 9. Apresentamos, a título exemplificativo, o raciocínio utilizado para a construção do mapa correspondente à Sequência 1:

O termo⁸⁹ central seleccionado foi ‘tratamento biológico aeróbio’, sendo que, tal como em outras sequências que teremos oportunidade de observar, é considerado um ‘processo’, encontrando-se este último superordenado em relação à totalidade da sequência de texto. Note-se que, na sequência em particular, o termo ‘tratamento biológico aeróbio’ não se encontra ao nível textual. No entanto, possuímos informação linguística anterior que permite afirmar que se trata de um processo de tratamento biológico. O mesmo se aplica ao termo ‘processo’, tratando-se de um termo inferido da totalidade das sequências de texto seleccionadas. Por motivos de registo e de estilo, esta é uma situação recorrente nos textos de especialidade.

Da análise da sequência, foi possível dividi-la em duas sub-sequências, tendo como base os termos ‘processo’ e ‘tratamento biológico aeróbio’:

a) **Sub-sequência 1:** (...) a matéria orgânica, que corresponde à carga poluente, é oxidada levando à formação de dióxido de carbono e à libertação de hidrogénio. Este, por sua vez, reage com o oxigénio levando à produção de moléculas de água.

b) **Sub-sequência 2:** Todo este processo é caracterizado por produzir uma elevada quantidade de energia bioquímica, armazenada na forma molecular, que é utilizada pela biomassa para se reproduzir (com a consequente produção de lamas) e pelo próprio metabolismo do processo (produção de CO₂).

Na sub-sequência 1 é estabelecida uma relação genérica entre ‘matéria orgânica’ e ‘carga poluente’, através do marcador linguístico [ML corresponde à], inserido numa oração de tipo relativo. Do mesmo modo, estabelece-se uma relação de causa/efeito a partir do termo ‘matéria orgânica’ através dos marcadores linguísticos verbais [ML é (oxidada)] e [ML levando à (formação de)], bem como [ML e levando à (libertação de)]. Para uma melhor harmonização ao nível da classificação gramatical dos termos a relacionar, optámos por substituir a forma verbal ‘é oxidada’ pelo substantivo ‘oxidação’. Obedecendo ao mesmo critério, incluímos o termo ‘hidrogénio’ enquanto elemento lexical independente, estabelecendo uma relação de

⁸⁹ A partir do momento em que analisamos e estabelecemos relações semânticas, deixamos de trabalhar com candidatos a termo e passamos a trabalhar com termos.

material/property com o termo 'libertação de hidrogénio', permitindo uma relação mais satisfatória com o termo 'oxigénio' através de uma forma gramatical idêntica⁹⁰. O pronome 'Este' remete para a última referência na oração anterior, facilitando o estabelecimento da relação entre termos. O marcador linguístico verbal [ML levando à (produção de)] foi de alguma forma suprimido, embora a relação de causa/efeito permita ao especialista uma inferência clara relativamente ao processo de produção quando a relaciona com o termo 'moléculas de água'.

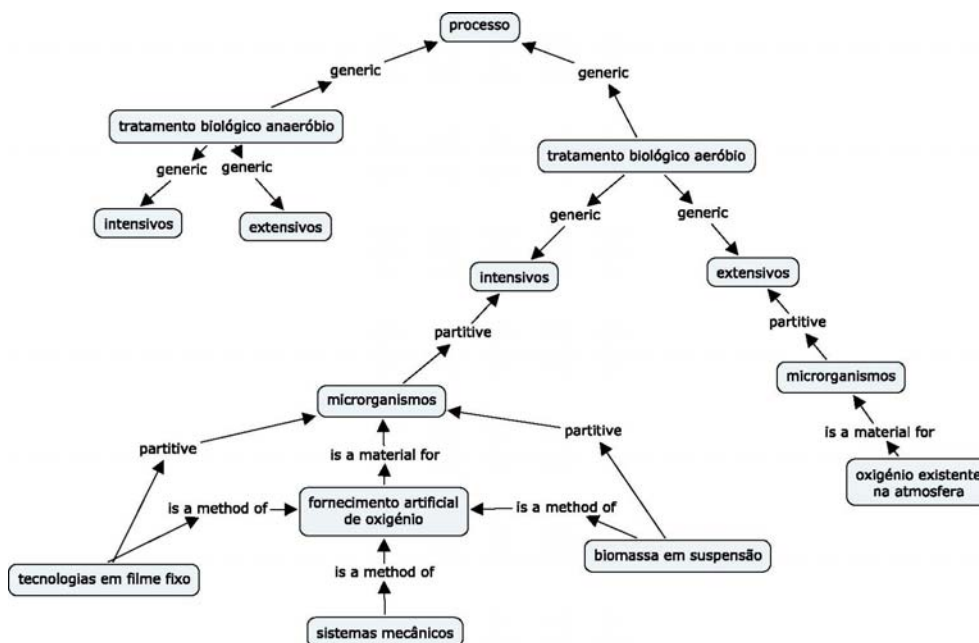
A sub-sequência 2 manifesta uma relação directa com a sub-sequência 1 através da referência explícita a 'todo este processo' que se encontra anteriormente descrito e que, de acordo com as nossas opções na construção da rede lexical, será o processo de tratamento biológico aeróbio. O emprego do marcador linguístico [ML é caracterizado por (produzir)], expressa uma relação de tipo associativo process/product. Optámos por alterar a forma gramatical da expressão 'forma molecular' por 'moléculas', ficando assim em consonância com o termo que se encontra no mesmo nível - neste caso 'biomassa' - não causando ruído em termos conceptuais. Mais uma vez, os marcadores linguísticos [ML é utilizado pela], [ML para] e [ML com a consequente (produção de)] apontam para uma relação de causa/efeito.

Observa-se que os termos e as relações entre termos extraídos da sequência permitem a construção de uma pequena rede lexical que representa uma organização discursiva, assim como aponta para uma organização conceptual. Ao nível linguístico, a sequência encontra-se correcta, uma vez que é composta por duas orações completas. Do ponto de vista conceptual, embora se encontre necessariamente incompleto face ao conteúdo cognitivo do domínio em causa, é portador de conexões que se estabelecem no texto com as sequências posteriores.

⁹⁰ Este procedimento teria de ser analisado de forma mais aprofundada e efectuado num âmbito mais alargado caso pretendessemos transpor os resultados obtidos através da construção da rede lexical para uma aplicação informática.

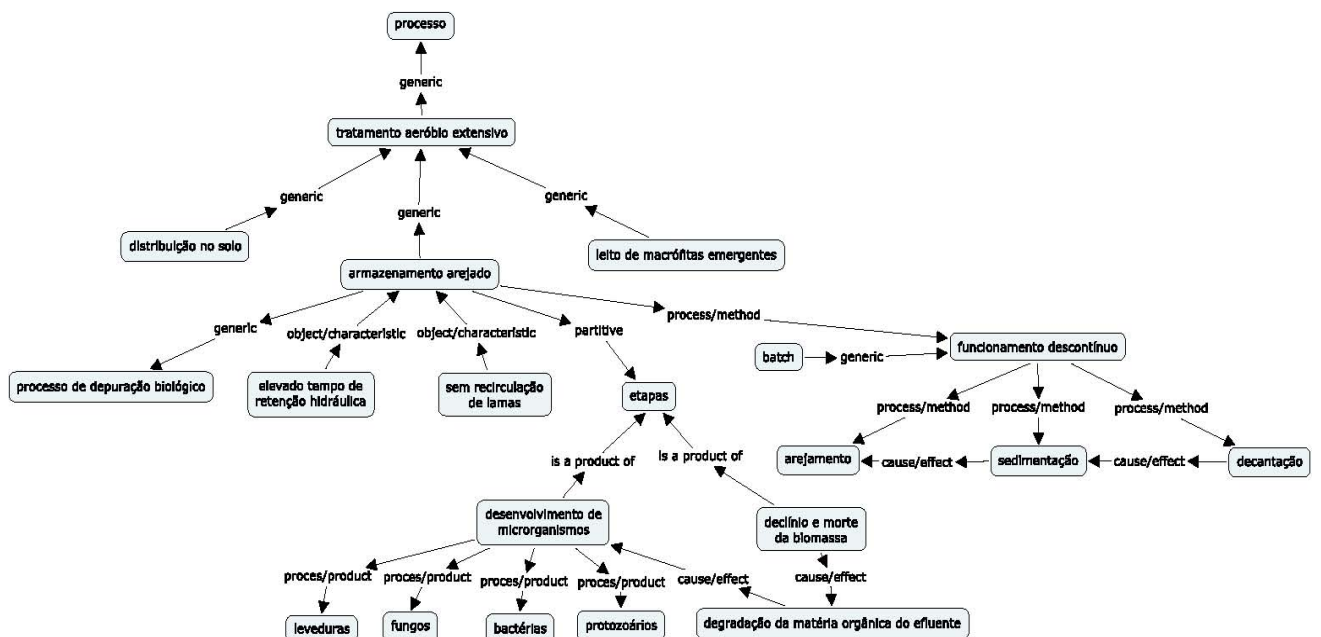
Sequência 2: Os processos de tratamento biológico aeróbios podem ser classificados em intensivos ou extensivos, tal como os processos anaeróbios já apresentados. Nos processos intensivos, os microrganismos necessitam de fornecimento artificial de oxigénio, através de sistemas mecânicos e recorrendo a tecnologias de filme fixo ou biomassa em suspensão, enquanto os processos extensivos utilizam o oxigénio que existe na atmosfera, independentemente do modo como o processo de tratamento se processa.

Figura 10 – Sequência 2



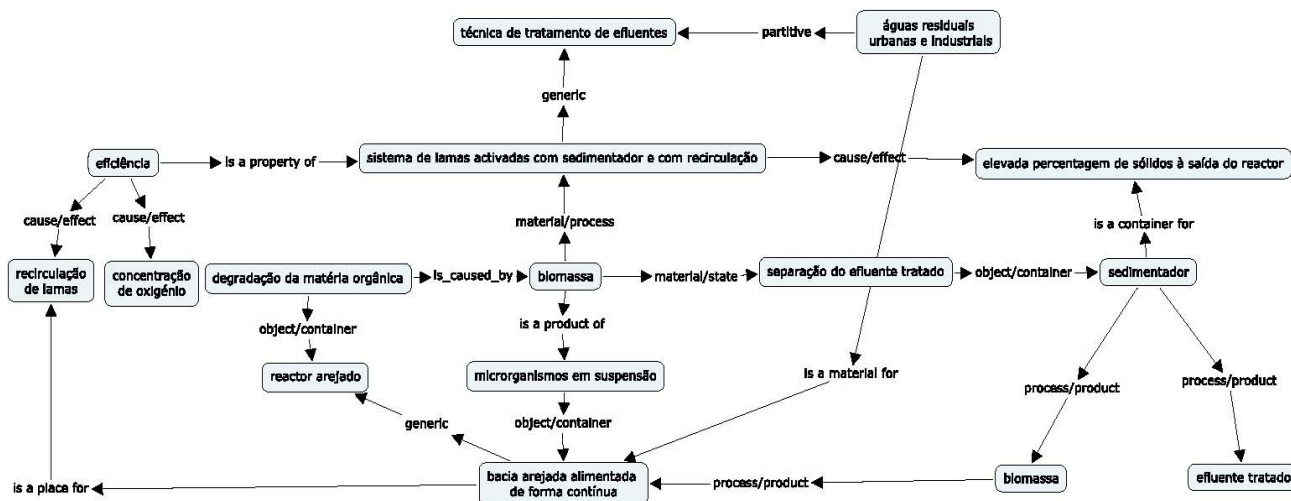
Sequência 3: Os processos de tratamento aeróbio extensivos que se destacam são: - distribuição no solo (...); -armazenamento arejado que é uma adaptação do processo de depuração biológico conhecido por lamas activadas. É caracterizado por possuir um elevado tempo de retenção hidráulico, ausência de recirculação de lamas e reduzida necessidade de mão-de-obra especializada (...). Neste processo de tratamento ocorre uma sucessão de etapas de onde decorre o desenvolvimento de vários microrganismos, como leveduras, fungos, bactérias e protozoários, e ao declínio e morte desta mesma biomassa, quando a degradação da matéria orgânica do efluente está praticamente completa. O sistema é caracterizado por funcionar de modo descontínuo, geralmente designado por batch, com três fases distintas (arejamento, sedimentação e decantação), todas elas processadas no mesmo local. (...) - leito de macrófitas emergentes (...)

Figura 11 – Sequência 3



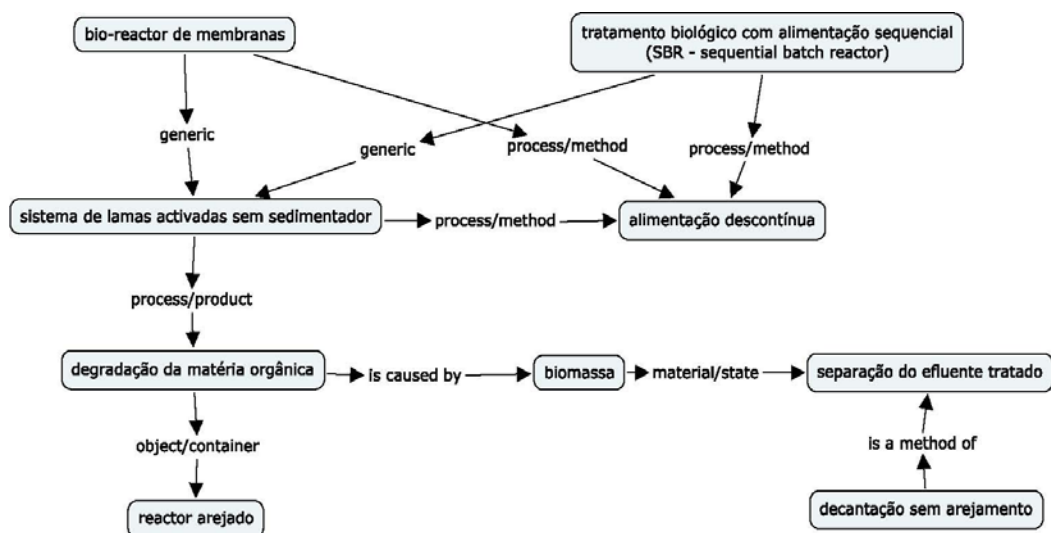
Sequência 4: Os processos de tratamento aeróbio intensivos têm por objectivo o conhecimento e a estabilização da razão matéria orgânica e biomassa (F/M) a aplicar ao processo biológico, conduzindo ao equilíbrio do funcionamento do reactor. A maior parte destes processos são baseados no sistema das lamas activadas, a técnica de tratamento de efluentes mais utilizada no tratamento de águas residuais urbanas e industriais. Este tipo de processos apresenta duas classes que se distinguem pela presença ou não de sedimentador que separe as lamas e um circuito que assegure a sua recirculação. Assim, as variantes deste processo de tratamento são as seguintes: - sistema de lamas activadas com sedimentador e recirculação, caracterizado por possuir duas fases distintas: uma fase em que a matéria orgânica é degradada por acção da biomassa num reactor arejado e uma segunda fase, em que a biomassa é separada do efluente tratado, através de um sedimentador. Este processo utiliza microrganismos mantidos em suspensão numa bacia arejada e alimentada de forma contínua com efluente e biomassa. Este processo, ao operar de forma contínua, contribui para que haja uma elevada percentagem de sólidos à saída do reactor, o que é colmatado com a passagem deste efluente por um decantador (...). A eficiência deste processo depende muito do caudal de recirculação de lamas para o tanque de arejamento e também da concentração de oxigénio no meio.

Figura 12 – Sequência 4



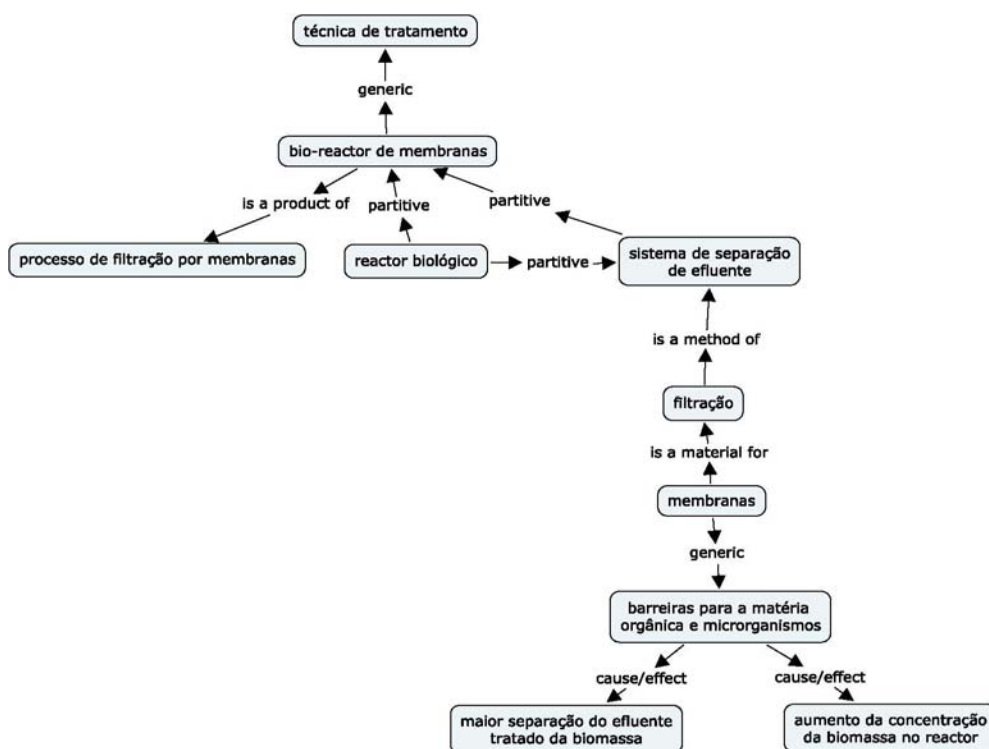
Sequência 5: -sistema de lamas activadas sem sedimentador é caracterizado por possuir, tal como o sistema anteriormente apresentado, duas fases de tratamento distintas: uma primeira em que a matéria orgânica é degradada por acção da biomassa num reactor arejado e uma segunda fase, em que a biomassa é separada do efluente tratado através da sua decantação na ausência de arejamento. A grande diferença deste sistema com o anterior é que a alimentação ao sistema é efectuada de forma descontínua. As técnicas de tratamento por lamas activadas e alimentação descontínua são: tratamento biológico com alimentação sequencial sequential batch reactor (SBR) e o bio-reactor de membranas.

Figura 13 – Sequência 5



Sequência 6: O bio-reactor de membranas é uma técnica de tratamento que deriva do processo de filtração por membranas. Consiste na associação de um reactor biológico a um sistema de separação do efluente tratado por filtração que é conseguido através de membranas. Estas têm a função de actuar como barreiras para a matéria orgânica e microrganismos, o que proporciona uma maior separação do efluente tratado da biomassa e consequentemente, o aumento da concentração desta no reactor (...).

Figura 14 – Sequência 6



Após a construção da rede lexical, foi possível contabilizar o tipo de relações e marcadores linguísticos utilizados tendo em conta a sua frequência:

Tabela 13 – Relações e marcadores linguísticos da rede lexical

Tipo de relação	Frequência no corpus
generic	26
cause/effect	15
partitive	11
process/method	7
is a material for	6
activity/place	5
is a method of	5
is a product of	4
object/container	4
proces/product	4
material/state	3
process/product	3
is_caused_by	2
material/property	2
is a place for	1
is a property of	1
material/method	1
phenomenon/measurement	1
process/material	1

Tal como verificado no mapa conceptual, a rede lexical apresenta a relação genérica com o maior número de frequências. A relação associativa de causa/efeito surge em segundo lugar, situação justificada, uma vez que foi seleccionado um contexto textual onde se descrevem processos de tratamento de efluentes. À semelhança do mapa conceptual, a relação partitiva aparece igualmente destacada. A rede lexical reflecte assim, a este nível, o sistema conceptual previamente construído.

Tivemos já oportunidade de referir que nas sequências seleccionadas existiram diversas situações em que foi necessário acrescentar e/ou suprimir termos e relações entre termos, não só para tornar os mapas mais explícitos, como também para corrigir e enriquecer algumas sequências.

Do mesmo modo, tal como já referido, foram alteradas algumas formas gramaticais do enunciado textual, como a forma verbal do particípio passado 'foi oxidada' para 'oxidação'. Fenómenos desta natureza decorrem do facto de estamos a lidar com duas realidades distintas: de um lado, temos o texto presente nas sequências, do outro lado, temos a língua, que necessita frequentemente da supressão das marcas discursivas. A utilização da rede lexical numa implementação mais avançada dos dados necessitaria de uma revisão atenta desta questão. O termo 'oxigénio existente na atmosfera' presente na Sequência 2 poderia, por exemplo, ser hipoteticamente substituído por 'oxigénio atmosférico'.

Foi possível comprovar, igualmente, a detecção de conhecimento extralinguístico implícito, não só devido à omissão de referências no texto, como também através das observações dos especialistas à rede lexical elaborada.

Podemos, assim, tecer os seguintes comentários às sequências de texto analisadas:

a) Totalidade das sequências:

- O termo 'processo' surge mais do que uma vez ligado por uma relação genérica com os vários sistemas de tratamento. Embora não apareça por vezes nas sequências, sabemos que estas se referem a processos de tratamento.

b) Sequência 1:

- Foi acrescentada a relação partitiva entre os candidatos a termo 'matéria orgânica' e 'tratamento biológico aeróbio', uma vez que detemos conhecimento prévio sobre o domínio. A sequência de texto não refere explicitamente que a matéria orgânica é uma das partes constituintes do tratamento biológico aeróbio.
- Acrescentou-se a relação 'reacts with' com direcção bilateral entre os candidatos a termo hidrogénio e oxigénio. Apesar de não estar referido explicitamente no texto, sabemos que as substâncias reagem uma com a outra.

c) Sequência 2:

- Acrescentámos os termos ‘tratamento biológico anaeróbio’, ‘intensivo’ e ‘extensivo’, ligados por relações genéricas entre si, e por relação genérica ao termo ‘processo’, decorrente da informação fornecida pelo enunciado textual: “(...) tal como nos processos anaeróbios já apresentados (...)”.
- Foi adicionada a relação partitiva entre os termos ‘microrganismos’ e ‘intensivos’, uma vez que sabemos que os microrganismos fazem parte dos tratamentos intensivos. O texto não permite ligar os termos através de um marcador linguístico: “ (...) nos processos intensivos, os microrganismos necessitam de (...)”.
- Acrescentámos o termo ‘microrganismos’ através de uma relação partitiva aos tratamentos extensivos’, dado que está omissa no texto que são os microrganismos que utilizam o oxigénio existente na atmosfera.

d) Sequência 3:

- Acrescentou-se a relação genérica entre os termos ‘processo’ e ‘tratamento aeróbio extensivo’.
- Foi omitida no mapa parte discursiva “(...) que é uma adaptação do processo de depuração biológico conhecido por lamas activadas (...)”, uma vez que, conforme explícito no texto, o processo de armazenamento arejado é uma adaptação do processo de lamas activadas.
- Foi igualmente omitida no mapa a referência à mão-de-obra especializada, uma vez que não constitui um elemento conceptual importante. A autora inclui esta indicação no texto porque, no contexto dos efluentes vinícolas, a opção por processos extensivos encontra-se directamente relacionada com os custos de mão-de-obra.
- Podemos ser induzidos em erro pela afirmação “O sistema é caracterizado por funcionar de modo descontínuo (...)”. No contexto dos efluentes vinícolas, o processo de armazenamento arejado pode funcionar de forma descontínua, dado que uma adega tem várias fases de funcionamento que condicionam a quantidade de efluente.

- Está implícito no texto uma sequência temporal (“(...) o sistema é caracterizado por funcionar de modo descontínuo (...) com três fases distintas (arejamento, sedimentação e decantação).”) Optámos por incluir no mapa as relações de causa/efeito não só para transmitir a noção cronológica, mas também para indicar as relações de dependência entre os processos.

e) Sequência 4:

- Foi acrescentada a relação ‘material/process’ entre os termos ‘biomassa’ e ‘sistema de lamas activadas com sedimentador e com recirculação’.
- Acrescentámos a relação genérica entre os termos ‘bacia arejada e alimentada de forma contínua’ e ‘reactor arejado’, uma vez que podemos considerar, decorrendo do nosso conhecimento do domínio, que uma bacia arejada alimentada de forma contínua é um reactor arejado.
- Foi acrescentado um ‘cross-link’ entre o termo ‘águas residuais urbanas e industriais’ e ‘bacia arejada alimentada de forma contínua’, uma vez que, embora não explícito no texto, o tratamento só funciona se o reactor receber efluente e biomassa/microrganismos.
- Não está explícito no texto o facto de existirem dois tipos de efluente: o efluente bruto e o efluente tratado. O termo ‘águas residuais urbanas e industriais’ representa o efluente bruto, e o termo ‘efluente’, ligado a ‘sedimentador’, representa o efluente tratado. Optámos por não acrescentar essa designação no mapa.
- Foi omitida a referência ao termo ‘decantador’, dado que ‘sedimentador’ é sinónimo de ‘decantador’.

f) Sequência 5:

- A utilização da conjunção ‘e’ no final do parágrafo faz com que classifiquemos o bio-reactor de membranas como uma tecnologia que utiliza uma alimentação descontínua, quando, na generalidade, isso não acontece, a não ser na limpeza periódica das membranas.

- Não foi necessário referir a presença ou ausência de recirculação, uma vez que o texto explicita que não temos sedimentador.

g) Sequência 6:

- Acrescentou-se a relação partitiva entre os termos 'sistema de separação do efluente' e 'bio-reactor de membranas'.
- Se analisarmos esta sequência separadamente, verificamos que está omissa o facto de se tratar de uma técnica de tratamento por lamas activadas que utiliza o processo de filtração por membranas para a separação da biomassa.

Como em todos os textos de especialidade de todas as áreas científicas, foram observadas algumas imprecisões textuais, nomeadamente através de marcas gráficas utilizadas de forma inadequada ou inclusão indevida de orações em determinado contexto. Pela natureza do texto científico e técnico, a imprecisão e a inexactidão – consciente ou inconsciente - constitui um fenómeno natural e frequentemente presente, decorrendo, por exemplo, da tentativa de redigir de forma esteticamente agradável, ou de elementos externos de pressão relacionados com prazos de conclusão. Situações desta natureza, embora detectáveis através de uma análise atenta do discurso, raramente comprometem a qualidade e validade do texto técnico e científico.

7.1.3.1.2 Relações semânticas

Relativamente às relações semânticas presentes nas sequências de texto, foi possível observar fenómenos de hiperonímia/hiponímia, holonímia/meronómia, sinonímia e quase-sinonímia.

Hiperonímia/hiponímia:

A tabela seguinte apresenta uma listagem não exaustiva das possíveis relações de hiperonímia/hiponímia e co-hiponímia:

Tabela 14 – Relações de hiperonímia/hiponímia

HIPERÓNIMO	HIPÓNIMO
carga poluente	matéria orgânica
energia bioquímica	moléculas
processo de tratamento aeróbio	processo de tratamento aeróbio intensivo
processo de tratamento aeróbio	processo de tratamento aeróbio extensivo
processo de tratamento aeróbio extensivo	distribuição no solo
processo de tratamento aeróbio extensivo	armazenamento arejado
processo de tratamento aeróbio extensivo	leito de macrófitas emergentes
processo de depuração biológico	lamas activadas
técnica de tratamento de efluentes	lamas activadas
microrganismo	leveduras
microrganismo	fungos
microrganismo	bactérias
barreiras	membranas
reactor biológico	bio-reactor de membranas

Tabela 15 – Relações de co-hiponímia

CO-HIPÓNIMO
processo de tratamento aeróbio intensivo processo de tratamento aeróbio extensivo
distribuição no solo leito de macrófitas emergentes

Já constatámos que as relações hierárquicas genéricas e partitivas se encontram entre as mais utilizadas. Existem, no entanto, particularidades discursivas que nem sempre permitem estabelecer uma conexão directa entre o texto e o tipo de relação.

Por exemplo, na sequência “(...) a matéria orgânica, que corresponde à carga poluente (...)”, indica aparentemente uma relação genérica, expressa no mapa. Mas o nosso conhecimento do domínio indica que a carga poluente não é constituída somente por matéria orgânica. Daí estar a relação genérica de baixo para cima, ou seja, a matéria orgânica é carga poluente, mas a carga poluente não é igual a matéria orgânica. Na realidade, deveríamos ter uma relação partitiva.

Outro exemplo pode ser observado aquando da referência ao desenvolvimento de microrganismos. Apesar de, pelo texto, deduzirmos que todos os elementos referenciados são microrganismos, tal inferência não é linear. Os protozoários podem não ser considerados microrganismos, uma vez que se trata de seres mais desenvolvidos, surgindo somente quando se verifica um nível baixo de matéria orgânica e um nível elevado de oxigénio.

Podemos comentar igualmente o facto de um reactor biológico e um bio-reactor de membranas apresentarem, pelo texto, uma relação partitiva, quando na verdade um reactor biológico é hiperónimo de bio-reactor de membranas.

Do mesmo modo, outro exemplo a reter é a relação de sinonímia entre ‘sedimentador’ e ‘decantador’⁹¹ na sequência 4, cuja distinção importa não pelo objecto físico, mas pela função que pretendo dar ao objecto relativamente ao efluente a tratar. O recurso às duas designações prende-se com o registo discursivo, na tentativa de evitar a repetição da mesma designação.

Holonímia/meronímia:

No que diz respeito à holonímia/meronímia, a tabela seguinte apresenta exemplos de relações de que poderíamos extrair das sequências analisadas:

⁹¹ Existe ainda o termo ‘clarificador’, não referenciado nas sequências.

Tabela 16 – Relações de holonímia/meronímia

HOLÓNIMO	MERÓNIMO
tratamento biológico	matéria orgânica
processo de tratamento biológico aeróbio	biomassa em suspensão
processo de tratamento biológico aeróbio	microrganismos
armazenamento arejado	tempo de retenção hidráulico
tratamento de efluentes por lamas activadas	sedimentador
tratamento de efluentes por lamas activadas	recirculação

A relação hierárquica partitiva contempla igualmente um número considerável de conexões na rede lexical elaborada. À semelhança do ponto anterior, existem seqüências de texto que não revelam explicitamente o tipo de relação entre os candidatos a termo. Por exemplo, não sabemos, pelo texto, que a matéria orgânica é uma parte do tratamento biológico aeróbio.

Sinonímia:

Podemos observar dois casos de fenómenos de sinonímia. O primeiro estabelece-se entre os termos ‘dióxido de carbono’ e ‘CO2’. Embora não esteja explícito, sabemos por conhecimento do domínio e por conhecimento vulgarizado que são sinónimos neste contexto (expressão linguística/ fórmula química). Um segundo caso prende-se com os termos ‘bacia arejada’, ‘tanque de arejamento’ e ‘reactor arejado’. A relação de sinonímia não está expressa no encadeamento textual, mas o nosso conhecimento de domínio permite inferir esse tipo de relação.

Tabela 17 – Relações de sinonímia

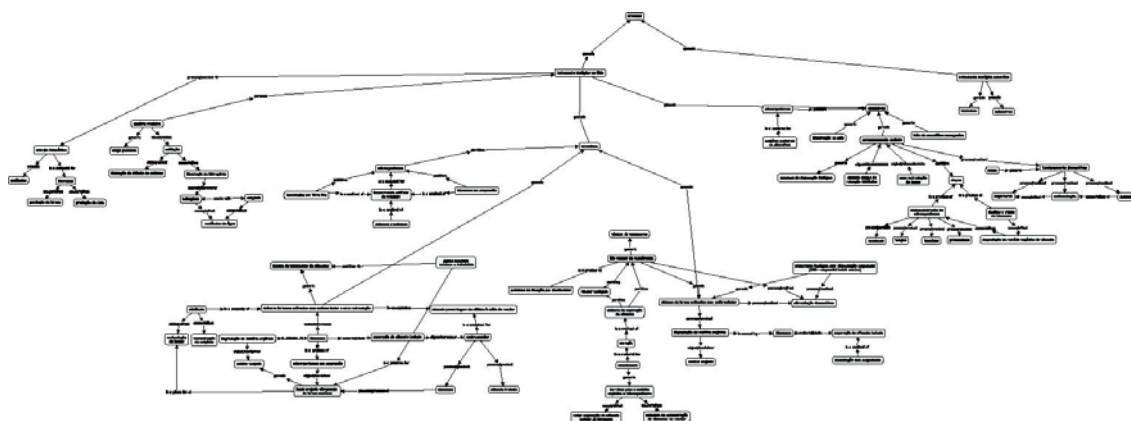
SINÓNIMO	SINÓNIMO	SINÓNIMO
dióxido de carbono	CO2	
bacia arejada	tanque de arejamento	reactor arejado

Quase-sinonímia:

Os termos 'biomassa' e 'microrganismos' podem ser considerados quase-sinónimos⁹². Como se trata de uma cultura mista e dinâmica, designa-se o conjunto dos microrganismos por 'biomassa'. Esta condição fez com que tivéssemos optado no mapa elaborado pelos especialistas pelo termo 'microrganismos'. A utilização por vezes indistinta dos dois termos está frequentemente relacionada com questões do foro discursivo.

A análise linguística das sequências seleccionadas permitiu a elaboração de várias redes lexicais com termos ligados por relações hierárquicas (genéricas e partitivas) e associativas (relações complexas). Os mapas respeitantes a cada sequência e o mapa conjunto foram submetidos à apreciação dos especialistas do Departamento de Ambiente e Ordenamento da Universidade de Aveiro, que validaram o conhecimento representado após análise crítica. Apresentamos, assim, uma proposta graficamente semelhante ao mapa conceptual elaborado directamente pelos especialistas, com a inclusão de todas as sequências de texto numa rede lexical única:

Figura 15 – Mapa total da rede lexical



⁹² A quase-sinonímia verifica-se quando dois termos designam o mesmo conceito, embora não sejam substituíveis em todos os contextos. O termo 'biomassa' encontra-se muito próximo do termo 'matéria orgânica', cuja remoção constitui a tarefa dos microrganismos. A matéria orgânica é o objecto do tratamento do efluente e o conceito pode sofrer alterações na conceptualização, dependendo da sua localização na estação de tratamento.

As decisões tomadas em relação à construção da rede lexical apontam para uma colaboração activa por parte do terminólogo, sustentadas, em larga medida, pelo conhecimento conceptual do domínio previamente adquirido. Essa atitude interventiva deve pautar-se não só pela constatação de eventuais imprecisões, como pela sua correcção, nomeadamente através da proposta de inclusão ou exclusão de determinado termo ou relação entre termo, ou da proposta de alteração da forma gramatical, da designação ou, eventualmente, da relação semântica entre termos presente no texto. A validação posterior do especialista é, naturalmente, fundamental neste processo.

8. Análise comparativa entre sistema conceptual e rede lexical

A análise dos contextos linguísticos permitiu elaborar várias tabelas com indicadores sobre possíveis intersecções entre termos extraídos das redes lexicais construídas a partir das sequências de texto e de conceitos seleccionados no mapa conceptual sobre tratamentos biológicos de águas residuais elaborado pelos especialistas:

Figura 16 – Intersecção entre termos e conceitos: ‘Processos’, ‘Tratamentos’, ‘Tecnologias’



Figura 17 – Intersecção entre termos e conceitos: ‘Processos de Filtração por Membranas’



Figura 18 – Intersecção entre termos e conceitos: ‘Lamas’

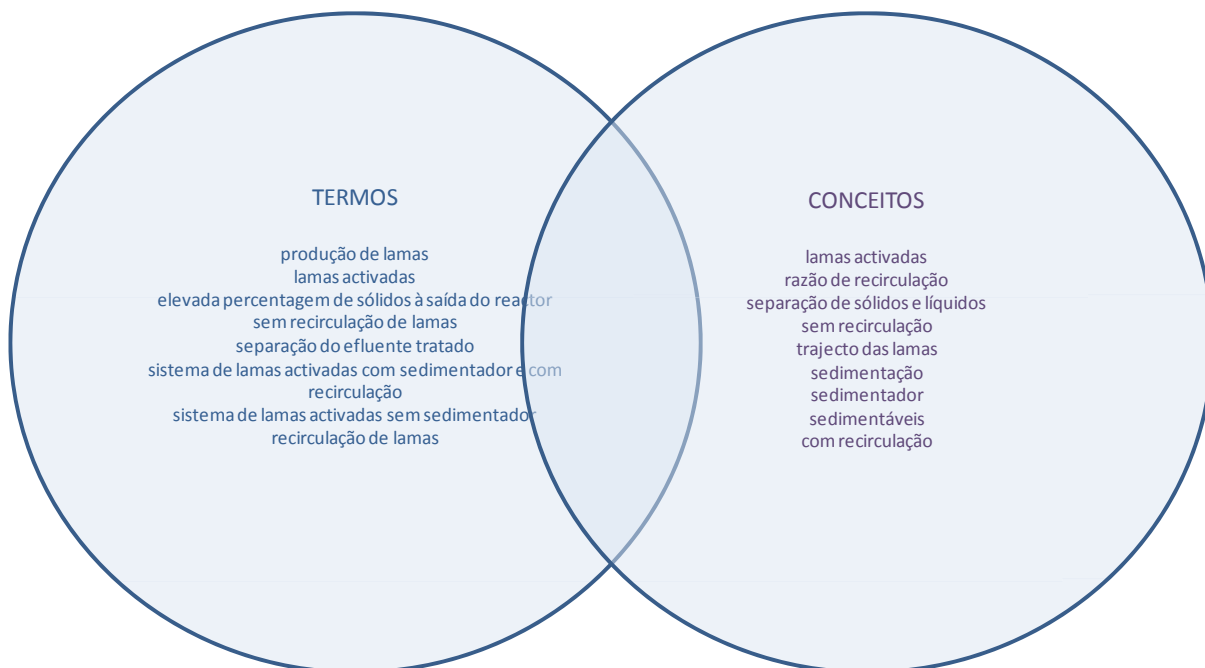
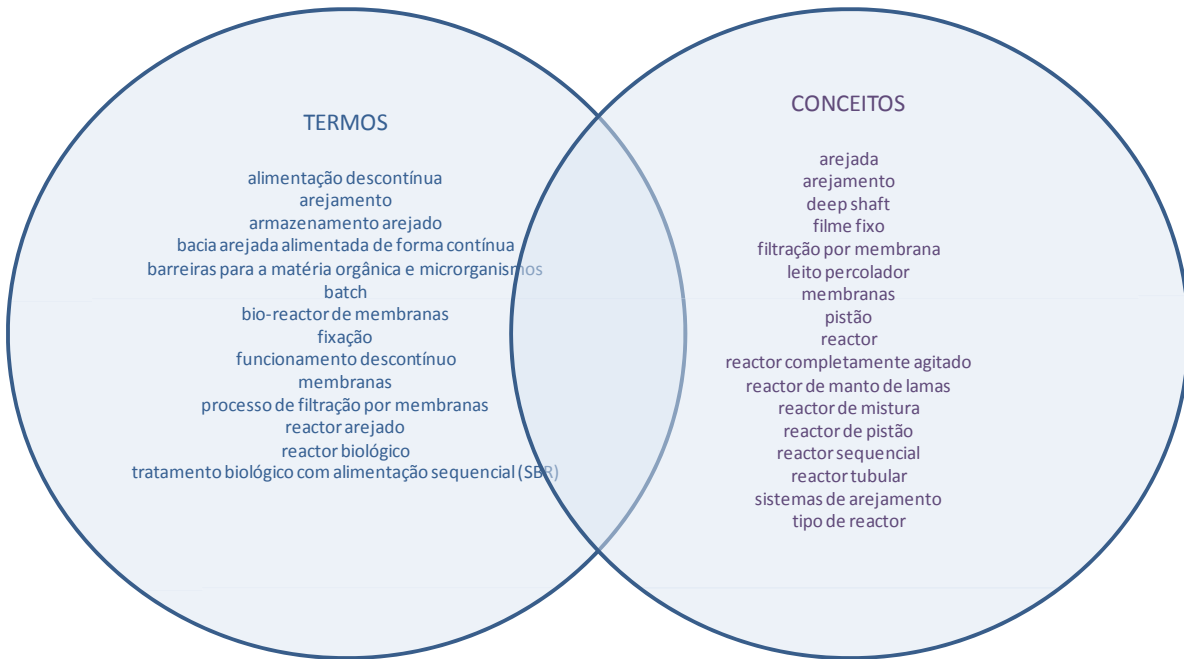
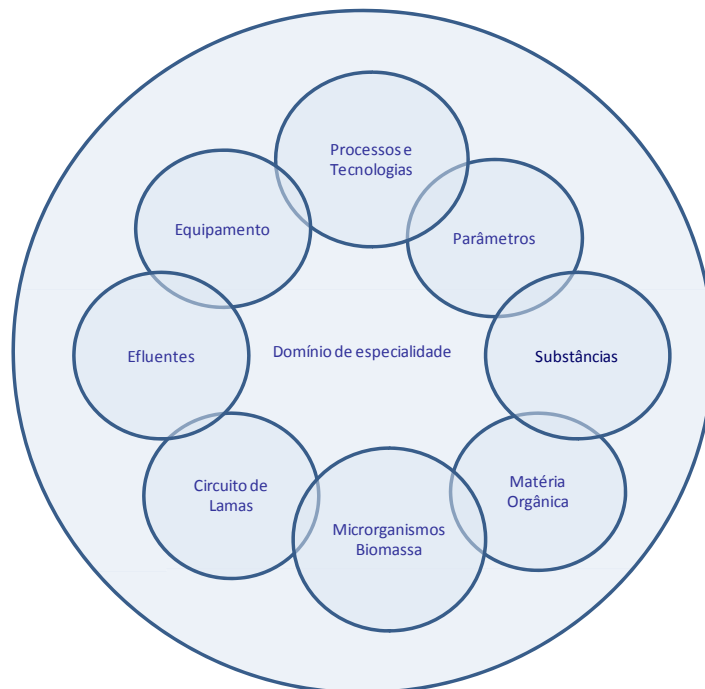


Figura 19 – Intersecção entre termos e conceitos: ‘Reactor’



Nos dados sobre as conceptualizações de âmbito mais geral que poderiam ser representadas pelas redes lexicais, é possível observar igualmente várias intersecções:

Figura 20 – Intersecção de conceptualizações a partir da rede lexical



O domínio de especialidade é comum à rede lexical e ao mapa conceptual - tratamentos biológicos de águas residuais - constituindo uma área do saber que versa sobre processos e tecnologias ambientais no tratamento de efluentes líquidos.

A rede lexical contém termos e relações entre termos que são comuns a todas as áreas conceptualizadas no mapa: parâmetros de análise de efluentes, composição dos efluentes, equipamentos utilizados nos processos de tratamento, produtos e sub-produtos, substâncias e microrganismos envolvidos, sendo que o local físico do tratamento de um efluente vinícola é igualmente uma estação de tratamento.

Além da intersecção conjunta e multi-direccional de todos os elementos no domínio de especialidade, podemos observar, por exemplo, que os termos e as relações entre termos directamente relacionados com o equipamento – parte integrante das tecnologias de tratamento – formam uma rede lexical que poderia quase sobrepor-se à parte do mapa conceptual correspondente aos reactores.

O alcance de constatações desta natureza por parte do terminólogo deve-se, sobretudo, ao trabalho conjunto desenvolvido na construção prévia do mapa conceptual e na relação de proximidade estabelecida com o registo discursivo e com a terminologia utilizada pelos especialistas nas aulas ministradas sobre esta matéria. Dispondo de conhecimento de especialidade, o terminólogo consegue fazer a distinção em termos de conceptualização, apreendendo a maior ou menor abrangência da parte discursiva em termos de organização conceptual válida do domínio. Essa capacidade é fundamental para separar as boas e más representações conceptuais que possam advir de uma abordagem semasiológica.

8.1. Notas conclusivas

Os dados terminológicos que pudemos observar, seleccionar e extrair, tanto na abordagem que parte do especialista como na abordagem que parte do texto, podem contribuir de forma significativa para a construção de sistemas conceptuais, nomeadamente de ontologias.

Embora a perspectiva que partiu da conceptualização directa dos especialistas não pretendesse utilizar análise de texto através de um corpus de especialidade, verificou-se a necessidade básica da mediação da língua natural na especificação informal de conceptualizações. Foram recorrentemente consultadas as obras já referenciadas enquanto manuais de boas práticas e documentos que reflectem uma organização consensual do domínio relativamente aos conceitos e à sua classificação. Verifica-se, pois, que mesmo centrando a nossa atenção na vertente conceptual, já acedemos pelo menos a um discurso verbal por parte do especialista, quer seja de carácter didáctico ou outro, o que nos permite, por via dele, aceder também a uma representação das conceptualizações.

Concluiu-se que uma selecção cuidada do corpus e uma análise aturada de sequências de texto podem fornecer dados que permitam esclarecer e coadjuvar a construção e organização das conceptualizações. Observou-se também que a reformulação em terminologia é importante não só para analisar a estrutura textual, mas também para verificar como se constrói o conhecimento no discurso específico de um domínio. Nesse sentido, constitui-se como um elemento semiótico fundamental na apreensão de dados extralinguísticos, uma vez que permitem realizar processos informais de inferência.

O resultado das redes lexicais construídas a partir das sequências revela uma proximidade esclarecedora relativamente ao mapa anteriormente elaborado, apontando para a possibilidade do texto de especialidade reflectir de alguma forma uma organização conceptual. No número reduzido de sequências seleccionadas foi possível observar que, proporcionalmente, dispomos de um número elevado de termos que correspondem a conceitos do mapa construído pelos especialistas. O processamento de língua natural através

da análise de texto aponta, assim, para uma activação simultânea da dimensão linguística e conceptual do termo.

A não limitação da selecção no tipo de relações permitiu construir redes lexicais e conceptuais complexas. Já em 1990 Sager advertia que “ (...) the simplistic view of the past that concepts are adequately represented by three types of relationships (generic, partitive, other) has been generally abandoned (...)” (Sager, 1990, p. 29), sendo que é possível e desejável estabelecer, por razões práticas, qualquer tipo de relação, adaptada à necessidade em causa, fazendo com que a dimensão dos sistemas conceptuais e a complexidade das relações sejam uma questão muito pragmática, directamente dependentes da análise do domínio. Pensamos que do ponto de vista da inteligência artificial, e particularmente na construção de ontologias, as restrições impostas às linguagens de representação fazem com que, em termos de formalização, não seja tão alargado o leque de possibilidades das relações a seleccionar. Como foi possível observar, tivemos à disposição um conjunto de opções que permitiu, sem dificuldade, encontrar uma relação linguística considerada válida para cada uma das ligações entre os termos.

Embora consideremos que o conhecimento prévio de domínio adquirido junto dos especialistas tenha contribuído de forma decisiva para a construção da rede lexical, observámos que as sequências analisadas revelaram manifestações de conhecimento ricas em conexões e ‘cross-links’ conceptuais. Este facto pode constituir uma condição importante para o âmbito de conhecimento que uma ontologia pretende representar. Foi portanto possível trabalhar a representação conceptual a partir da análise lexical, ajustando os dados aos resultados da última.

O facto de termos assistido a aulas sobre o domínio de especialidade em causa e a proximidade de registo linguístico entretanto estabelecida com os especialistas, potenciaram o sucesso dos resultados obtidos na abordagem semasiológica. Desta forma foi possível inferir grande parte do conhecimento implícito ausente das sequências seleccionadas. O especialista deve assim fazer parte do processo de construção de ontologias desde a sua fase inicial, independentemente da utilização de processamento de língua natural.

Foi muito positiva e bem sucedida a colaboração do terminólogo e do especialista na construção dos mapas conceptuais de domínio e validação das redes lexicais, confirmando a

importância das valências de cada uma das especialidades. Assim, o trabalho prático desenvolvido permite concluir que é possível, e aconselhável, utilizar uma metodologia mista, aplicando a abordagem onomasiológica e a abordagem semasiológica da terminologia na construção de ontologias, sempre que a aplicação o justifique, embora consideremos que a ordem de utilização pode não ser arbitrária.

Os especialistas que colaboraram mais de perto neste projecto entendem que, apesar de partilharem das conclusões acima expostas, e após uma análise final dos dois mapas resultantes das abordagens centradas no ponto de partida extralinguístico e linguístico, privilegiariam a aplicação da perspectiva extralinguística. Verificamos que, mesmo acompanhada de uma análise crítica, o texto pode conter imprecisões que afectam, em menor ou maior grau, a conceptualização do domínio, como, por exemplo, uma marca gráfica erradamente aplicada, ou uma conjunção indevidamente utilizada, como foi observado. Por outro lado, uma abordagem onomasiológica não necessita preocupar-se com os elementos do acto comunicativo, nomeadamente com o autor e com o leitor, bem como com a sintonia que deve existir entre os dois. Para um especialista, a construção de um mapa conceptual a partir de um conjunto de textos de especialidade revelar-se-ia mais difícil de concretizar, uma vez que os próprios especialistas têm dificuldade em serem consensuais na terminologia que aplicam. Um mapa conceptual elaborado sem recurso a análise linguística, embora necessite da colaboração de vários especialistas, obedece a uma uniformidade de construção estratégica mais coerente, facilitando a integração de contribuições posteriores. A probabilidade de encontrar inconsistências é maior no texto de especialidade do que na construção directa do mapa conceptual, visto como algo mais fechado e por isso mais salvaguardado de eventuais imprecisões.

Para os especialistas, é possível construir um mapa conceptual directamente de texto, mas só com recurso a obras de referência com amplo reconhecimento e com uma grande uniformidade ao nível da organização dos conceitos, passíveis, por isso, de transmitirem conceptualizações com relações explícitas e consensuais. Contudo, é reconhecido que cada obra de referência iria resultar num mapa conceptual diferente, embora aceite pela comunidade. Um documento dessa natureza reflectiria as conceptualizações de um grupo muito alargado de especialistas. É necessário ter presente, porém, que recorreremos sempre à

língua natural para elaborar um mapa conceptual de domínio. Não basta fazer esquemas, é necessário recorrer ao texto, embora obedecendo a critérios rigorosos de selecção. Será eventualmente conveniente recorrer à abordagem semasiológica depois da abordagem onomasológica, no sentido de confirmar a representação das conceptualizações. A possibilidade de acesso a mapas conceptuais de domínio ajudaria inclusivamente o autor do texto de especialidade a ser mais preciso no seu discurso.

Podemos concluir que a rede lexical construída a partir de sequências de texto classificado como sendo de especialidade reflecte uma determinada organização conceptual, dada a proximidade com o mapa anteriormente elaborado. A representação gráfica da rede lexical, os termos e as relações entre os termos foram aceites pelos especialistas, pelo que consideramos válida a representação das manifestações linguísticas. Observámos que, decorrendo do texto, foi possível representar um conjunto de relações complexas que podem enriquecer o mapa conceptual inicial. Por outro lado, os marcadores linguísticos utilizados permitiram, em muitos casos, clarificar a conceptualização, fazendo com que o texto possa desempenhar um papel decisivo. Contudo, conforme já observado nesta dissertação, o texto de especialidade nem sempre obedece a uma natureza prototípica.

De uma forma geral, constata-se as dificuldades referidas pelos engenheiros do conhecimento relativamente à utilização da língua natural na construção de ontologias. O que é importante, do nosso ponto de vista, é reter que não devemos entender uma rede lexical como um conjunto de dados susceptível de ser transposto de forma (semi)automática para uma estrutura formal. A observação final dos mapas permite concluir que o que está em causa pode ser não só a representação dos dados, como os próprios dados. Assim, o texto tanto pode surgir como um elemento operacional e colaborativo extremamente útil e facilitador do processo de representação do conhecimento, como constituir-se como um elemento perturbador e causador de ruído.

As críticas por vezes contundentes em relação à utilização dos corpora advêm, a nosso ver, da aplicação de metodologias menos correctas. Podemos e devemos fazer uso dos corpora através de uma metodologia mista, acedendo ao conhecimento de domínio linguístico e extralinguístico através do especialista, considerando que a organização conceptual ou cognitiva do especialista é, pela sua natureza, quase sempre de ordem esquemática. Assim, o

lexical não deve substituir o conceptual, uma vez que é através deste que conseguimos apreender a organização das conceptualizações.

10. Conclusão

Partindo de um ponto de vista das ciências da linguagem, pretendeu-se com esta dissertação analisar o papel da terminologia na representação do conhecimento através da construção de ontologias. Para tanto, explicitou-se o que se pode entender por uma ontologia e por conhecimento, estabelecendo-se posteriormente como base teórica uma análise comparativa entre duas metodologias: a abordagem onomasiológica, que se baseia no extralinguístico, e a abordagem semasiológica, em cuja orientação prevalece o ponto de vista linguístico.

A terminologia encontra-se ligada à inovação em todas as suas vertentes, sendo que existe uma relação directa entre terminologia e conhecimento. Entendemos que uma ontologia, enquanto ferramenta de representação do conhecimento, necessita da colaboração fundamental da terminologia na fase informal de 'especificação de conceptualizações', isto é, antes da formalização dos dados através de uma linguagem de representação. A natureza da aplicação irá condicionar a utilização das duas abordagens.

Apesar de oriunda das orientações teóricas de Wüster, marcadamente onomasológicas, a terminologia enquanto ciência e prática tem desenvolvido nas últimas décadas uma postura mais centrada na importância e análise do termo em discurso, enquanto manifestação linguística do conceito. Na perspectiva linguística, um termo constitui-se normalmente como uma unidade lexical de especialidade.

Contudo, a utilização cada vez mais frequente de ferramentas de partilha e reutilização de conhecimento entre sistemas de informação suscitou o interesse não só pelo termo enquanto unidade não necessariamente lexical, como por uma postura que valoriza e salienta de forma renovada a primazia do conceito, um dado extralinguístico que, pela sua natureza, pode eventualmente dispensar o contributo da língua natural.

Uma ontologia no sentido computacional é um artefacto, um produto, construído com o propósito de representar formalmente especificações de conceptualizações de forma consensual, de modo a poderem ser reutilizadas e partilhadas entre seres humanos e sistemas

de informação. Nesse sentido, só existe o que pode ser formalizado numa linguagem de representação. A preocupação central do engenheiro do conhecimento é a representação e a sua utilidade, onde questões relacionadas com interpretação e 'verdade' não são prioritárias. Baseada nos princípios de classificação aristotélicos, constrói-se uma rede formal de conceitos ligados por relações que representam um compromisso ontológico, uma visão do mundo, restrita a um domínio específico de conhecimento. Aceitar a ontologia é aceitar esse compromisso ontológico, cujo universo discursivo engloba um conjunto de entidades com determinadas características relacionadas de forma mais ou menos complexa e que tenta aproximar-se das conceptualizações da comunidade de especialistas. Idealmente, uma ontologia não depende de nenhuma língua natural, procurando constituir-se como uma ferramenta computacional de partilha não ambígua.

É precisamente no recurso ou não à língua natural, nomeadamente através do texto de especialidade, que se centram duas grandes orientações na construção de ontologias: por um lado defende-se que é possível construir ontologias de forma (semi)automática através da extracção de dados de texto, uma vez que, de acordo com teorias distribucionais, os elementos sintácticos contíguos ou semi-contíguos tendem a pertencer ao mesmo *cluster* conceptual; por outro entende-se que, embora se possam extrair dados de texto, estes não podem constituir-se como elementos válidos na construção de ontologias, não só porque o discurso se encontra permeado de conhecimento implícito não detectável por métodos computacionais, como a própria construção sintáctica enferma de princípios estilísticos e imprecisões que impossibilitam uma clara apreensão do significado dos termos. A extrema dificuldade em automatizar tarefas de inferência fez com que a inteligência artificial optasse por adoptar restrições rígidas na representação formal das conceptualizações, concentrando-se mais na natureza do raciocínio do que no conteúdo, evitando assim a interferência da língua natural.

De facto, a língua natural é intrinsecamente ambígua. Graças a essa propriedade, consegue contemplar a extrema complexidade da comunicação humana constituída por uma riqueza inesgotável de nuances cognitivas. Esse facto representa uma dificuldade difícil de transpor para um sistema de informação. Contudo, pensamos que a construção de uma ontologia não pode subtrair-se à questão linguística. Apesar das proposições matemáticas que constituem

uma ontologia computacional não estarem presas a línguas naturais, o processo de selecção e organização de conceitos necessita do contributo da vertente linguística do termo. As preocupações com o consenso do conhecimento representado existem tanto na fase da especificação informal, como na fase da especificação formal. O sucesso da partilha e reutilização formal do conhecimento não se encontra directamente relacionado com o recurso à língua natural na sua fase informal inicial. Tem antes a ver com o entendimento alargado a que uma comunidade de especialistas chega sobre um determinado domínio e aos limites computacionais da linguagem de representação utilizada.

O discurso verbal pode e deve ser visto como uma plataforma de análise, onde termos e outras entidades lexicais se cruzam, mobilizando os recursos da língua natural para transmitir conhecimento. Uma terminologia na perspectiva linguística não é prioritariamente um conjunto de conceitos, mas sim um conjunto de expressões da língua natural que manifesta, através das relações que estabelece em discurso, traços semânticos que apontam para uma eventual organização conceptual. No texto, unidades lexicais gerais e especializadas e outras unidades não linguísticas formam um todo coerente que ultrapassa o contexto do texto, embora a sua significação se constitua em discurso.

É nossa convicção de que o recurso ao texto de especialidade na construção de ontologias deve obedecer não só a critérios rigorosos e supervisionados de selecção do corpus, como também ao pressuposto de que um texto de especialidade pode não reflectir um conhecimento organizado de um domínio. Por se encontrar de alguma forma cativo da orientação cognitiva do autor, os termos e as relações entre termos ficam dependentes de estruturas léxico-semânticas que podem divergir da orientação conceptual expectável. Simultaneamente, a presença de conhecimento implícito é fortemente sentida. Consideramos portanto incorrecto presumir que as redes lexicais que podem ser extraídas de texto constituam ontologias, nem que possam ser transpostas de forma mais ou menos automática para linguagens de representação, não só porque o texto pode conter imprecisões discursivas, como também a presença do 'não-dito' vai necessariamente suprimir conexões importantes. As redes lexicais constituem eventualmente pré-estágios de redes conceptuais.

Porém, não consideramos que seja necessário existir uma descontextualização linguística para apreendermos um conceito. Apesar da lógica matemática utilizada nas linguagens de

representação ser ontologicamente neutra, podemos melhorar as representações das conceptualizações através da análise do discurso, capturando, de forma mais exacta, as relações entre conceitos e optimizando uma maior profundidade analítica e rigor de inferência.

Existe por parte do especialista a necessidade efectiva de consultar texto de especialidade, assim como a certeza de que determinados textos contêm os requisitos necessários para podermos extrair, com um elevado grau de confiança, elementos que permitam organizar o conhecimento de um domínio. Observámos que, decorrendo do texto, foi possível representar relações complexas que podem enriquecer mapas conceptuais elaborados directamente por especialistas.

Pensamos, ainda, que os estudos e práticas terminológicas actuais devem concentrar-se nas relações entre termos e nas relações entre conceitos. Embora o conjunto de conceitos e o conjunto de termos que perfazem um domínio de conhecimento constituam os nós de uma rede, são as relações entre conceitos e as relações entre termos que lhe dão corpo e que tecem e fortalecem a rede, seja ela lexical ou conceptual. É nessas relações que se encontra a capacidade de inferência de raciocínio.

O afastamento em relação ao texto por parte de determinadas correntes da inteligência artificial dá-se, também, pela sua noção muito clara do tempo e da satisfação do cliente. Optam, assim, por metodologias que minimizem as dificuldades de formalização. O contacto com o texto por parte do especialista iria realçar muitos aspectos cognitivos que, de contrário, não seriam porventura tratados. Os resultados apresentados podem ser por isso resultados consensuais e estatisticamente relevantes. Mas temos dúvidas se serão conceptualmente mais válidos, mesmo restritos a uma determinada aplicação.

Foi igualmente possível concluir, decorrente do trabalho desenvolvido, que, embora complementares, a ordem de utilização das metodologias semasiológicas e onomasiológicas não é arbitrária. Podemos alcançar resultados semelhantes partindo do linguístico ou do extralinguístico. Contudo, a obtenção de resultados mais satisfatórios é facilitada pela adopção da abordagem onomasiológica em primeiro lugar, uma vez que permite construir uma rede conceptual com a colaboração directa dos especialistas que nos faz automaticamente

detentores de conhecimento de domínio. Essa condição permitirá enriquecer e sustentar a posterior metodologia semasiológica, sempre que o recurso a esta se justifique. O âmbito e a profundidade do conhecimento a representar estão assim directamente relacionados com a simplificação ou não do objecto de estudo.

Na construção de ontologias não existem, portanto, abordagens exclusivamente extralinguísticas ou linguísticas. Existem abordagens que podem partir do extralinguístico ou do linguístico. O facto da inteligência artificial realçar o ponto de partida extralinguístico, não significa que destitua de importância a terminologia. Adopta, porém, uma postura que tenta isolar a terminologia do contacto com o discurso verbal, separando o termo da sua vertente discursiva. Se partirmos do princípio que não existe uma relação de um para um entre texto e ontologia, nem entre especialista e ontologia, podemos concluir que as ontologias não são extraíveis, são artefactos que se constroem em colaboração interdisciplinar. Por esse motivo, um texto não vai fornecer dados que viabilizem a construção automática de ontologias, uma vez que a sua génese não obedeceu a esses princípios. Pode, sim, conter manifestações linguísticas de conhecimento que contribuam para a organização conceptual informal de um domínio. O conhecimento é, na sua essência, um elemento cognitivo. Representar conhecimento é especificar conceptualizações, formal ou informalmente. Nesse sentido, tanto o texto como a ontologia podem desempenhar esse papel.

Entendemos que a metodologia mista utilizada nesta dissertação salientou resultados promissores na construção de sistemas conceptuais de domínio, apontando para a necessidade de re-equacionar o estatuto do corpus nas abordagens onomasiológicas. A quase sobreposição existente entre o mapa conceptual e a rede lexical revela, de forma esclarecedora, a necessidade de aceder previamente às conceptualizações do domínio por via do especialista. Curiosamente, ou talvez não, o acesso a essas conceptualizações faz-se precisamente através do discurso, da língua natural, sempre presente em todas as etapas da especificação informal da construção de uma ontologia.

Não esqueçamos, também, que, actualmente, a partilha de conceptualizações inclui a relação exclusiva entre sistemas, onde o conceito de conhecimento adquire contornos muito específicos. Conceptualizar é construir um modelo abstracto do mundo, identificando os conceitos e relações entre conceitos relevantes desse fenómeno. Conhecimento é a

capacidade de agir com base nas representações do mesmo. Trata-se de uma aproximação ao real, cujos limites são entendidos por todos os intervenientes com o propósito de partilha e reutilização eficiente entre sistemas. A formalização dos dados através de uma linguagem matemática nas ontologias não significa que represente conhecimento de forma mais abrangente do que o texto. Representa, sim, um tipo de conhecimento susceptível de ser totalmente partilhado entre sistemas.

Na construção de ontologias, a terminologia deve assim, na nossa opinião, elevar-se em relação a metodologias defendidas pelas diversas escolas, estejam elas ligadas à Escola de Viena, a abordagens sócio-cognitivas ou outras correntes teóricas, e centrar-se nas necessidades da aplicação e nos elementos disponíveis para a sua construção, seleccionando com rigor as fontes dos dados – linguísticas e extralinguísticas - e adaptando os resultados às necessidades das representações.

O papel da terminologia na construção de ontologias não deve passar exclusivamente pelo texto, embora pensemos que a extracção de dados semânticos de um texto de especialidade pode contribuir para esse processo. Devemos sobretudo ter presente o tipo de manifestações de conhecimento que uma ferramenta informática pode representar e o tipo de manifestações de conhecimento que uma análise de texto pode fornecer. A terminologia não é por isso algo externo ao texto de especialidade, antes faz parte integrante dele, constituindo-se como um dos seus elementos fundamentais. Podemos dispensar o texto em função da aplicação. Não podemos dispensar, porém, a terminologia.

Tendo como base de trabalho os propósitos utilitários que norteiam as duas práticas, o terminólogo, o especialista e o engenheiro do conhecimento/ontologias devem servir-se do texto e de outras fontes disponíveis para extracção e confronto de dados, beneficiando das suas potencialidades e utilizando as metodologias de análise que entendam adequadas.

Fruto da sua formação, o terminólogo detém uma grande sensibilidade relativamente à construção de sistemas conceptuais e à organização de redes léxico-semânticas. Representa, por isso, um elemento importante na construção de ontologias. O facto de o terminólogo encarar cada vez mais os termos como unidades que podem ou não ser lexicais é um avanço relativamente às práticas terminológicas anteriores, evidenciando uma formação cada vez

mais próxima da ciência computacional, reforçando e afirmando as suas competências técnicas abrangentes no novo enquadramento tecnológico deste século, à semelhança de todas as áreas profissionais. A construção e manutenção de ontologias requerem, porém, que o movimento seja tri-direccional, entre terminólogos, especialistas e engenheiros de ontologias e do conhecimento.

Qualquer perspectiva é necessariamente redutora em relação ao mundo que tenta representar. O que está em causa na construção de uma ontologia é a finalidade e os limites dessa representação. Entendemos que o recurso à terminologia através da língua natural faz parte integrante da fase inicial informal, onde é aconselhável, a bem da abrangência e profundidade das conceptualizações, que se recorra aos elementos válidos disponíveis, nomeadamente especialista e texto, em função dos resultados que pretendemos obter.

A representação do conhecimento acompanhou sempre a história da humanidade. Mas a sua antiguidade não a impede de continuar a estar na vanguarda da evolução tecnológica. Tanto assim é que, após a conclusão desta dissertação, torna-se difícil fazer previsões sobre o futuro das metodologias de construção de ontologias, ou mesmo sobre as próprias ontologias. Que futuro lhes estará reservado, enquanto artefacto de engenharia formalizado com objectivos de partilha de conhecimento consensual? A sua evolução dependerá em grande parte da capacidade das ferramentas informáticas conseguirem capturar, em maior ou menor grau, as subtilidades cognitivas das nossas conexões mentais.

11. Referências bibliográficas

11.1 Bibliografia citada

- Almeida, E. R. L. (2008). *Avaliação da Biodegradabilidade Aeróbia de Efluentes Vinícolas*. Universidade de Aveiro, Aveiro.
- Almeida, M. B., & Bax, M. P. (2003). Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, 32, 7-20.
- Antia, B. E. (2000). *Terminology and language planning: an alternative framework of practice and discourse*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Aussenac-Gilles, N. (2005). *Supervised text analysis for ontology and terminology engineering*. Paper presented at the Dagstuhl Seminar on Machine Learning for the Semantic Web, Germany.
- Bachimont, B. (2006, 25 Novembro 2008). Qu'est-ce c'est une ontologie ? , from http://www.technolanguen.net/imprimer.php3?id_article=280
- Barros, S. (2007). *A relação genérica num corpus de direito constitucional português: identificação de contextos para a delimitação de candidatos a termo.*, FCT-UNL - Universidade Nova de Lisboa, Lisboa.
- Beck, H., & Pinto, H. S. (2002). Overview of approach, methodologies, standards, and tools for ontologies. *The Agricultural Ontology Service (UN FAO)*, Sections 1, 4, and 6.
- Bendaoud, R., Rouane, M., Hacene, Y. T., Delecroix, B., & Napoli, A. (2007). *Text-based ontology construction using relational concept analysis* Paper presented at the International Workshop on Ontology Dynamics (IWOD), Innsbruck, Austria
- Borges, M. M. (2002). *De Alexandria a Xanadu* (Vol. 2). Coimbra: Quarteto Editora.
- Bourigault, D., & Aussenac-Gilles, N. (2003). *Construction d'ontologies à partir de textes*. Paper presented at the TALN 2003 Batz-sur-Mer.
- Bourigault, D., Aussenac-Gilles, N., & Charlet, J. (2002). Construction de ressources terminologiques ou ontologiques à partir de textes: un cadre unificateur pour trois études de cas. *Revue d'Intelligence Artificielle*, X, 1-X.
- Bourigault, D., Jacquemin, C., & (eds.), M.-C. L. H. (2001). *Recent Advances in Computational Terminology* (Vol. 2). Amsterdam: John Benjamins.

- Brewster, C. (2003). *Knowledge Maintenance and the Frame Problem*. Paper presented at the 6th Annual CLUK Research Colloquium, Edinburgh, United Kingdom.
- Brewster, C., Ciravegna, F., & Wilks, Y. (2003). *Background and Foreground Knowledge in Dynamic Ontology Construction*. Paper presented at the Proceedings of the Semantic Web Workshop -SIGIR 2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada.
- Brewster, C., Iria, J., Ciravegna, F., & Wilks, Y. (2005). *The Ontology: Chimaera or Pegasus*. Paper presented at the Dagstuhl Seminar Machine Learning for the Semantic Web,, Dagstuhl
- Brewster, C., & O'Hara, K. (2004). Knowledge Representation with Ontologies: The Present and Future. *IEEE Intelligent Systems*, 19(1), 72-81.
- Brewster, C., & Wilks, Y. (2004). *Ontologies, Taxonomies and Thesauri: Learning from Texts*. Paper presented at the The Use of Computational Linguistics in the Extraction of Keyword Information from Digital Library Content, King's College London
- Buitelaar, P. (2000). *Semantic lexicons: between ontology and terminology*. Paper presented at the OntoLex 2000, Sozopol, Bulgaria.
- Buitelaar, P., Cimiano, P., & Magnini, B. (2005). Ontology learning from text: an overview. In B. Buitelaar, P. Cimiano & B. Magnini (Eds.), *Ontology Learning from Text: Methods, evaluation and applications* (pp. 3-12): IOS Press.
- Cabré, T. (1998 (1992)). *Terminology: Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Cabré, T. (2000). Elements for a theory of terminology: Towards an alternative paradigm. In J. Benjamins (Ed.), *Terminology* (Vol. 6:1, pp. 1-23). Amsterdam: John Benjamins.
- Cabré, T. (2003). Investigar en terminología: posibilidades y líneas de trabajo. In E. Ortega (Ed.), *Panorama actual de la investigación e interpretación* (Vol. I).
- Cabré, T. (2003a). Theories of terminology - their description, prescription and explanation. In J. Benjamins (Ed.), *Terminology* (Vol. 9:2, pp. 163-199). Amsterdam: John Benjamins.
- Cabré, T. (2006). *Morfología y terminología*. Paper presented at the La morfología a debate, Jaén.
- Cabré, T., Condamines, A., & Ibekwe-SanJuan, F. (2005). Introduction - Application-driven terminology engineering. In J. Benjamins (Ed.), *Terminology* (Vol. 11:1, pp. 1-19): John Benjamins.
- Cabré, T., & Estopà, R. (2002). El conocimiento especializado y sus unidades de representación: diversidad cognitiva. *Sendébar*, 13, 141-153.

- Cabré, T., & Feliu, J. (2002). *Conceptual relations in specialized texts: new typology and an extraction system proposal*. Paper presented at the TKE 2002 - 6th International Conference on Terminology and Knowledge Engineering, Nancy.
- Cabré, T., Feliu, J., & Tebé, C. (2001). *Bases cognitivas de la terminología: hacia una visión comunicativa del concepto*. Paper presented at the II Congreso de la Asociación Española de Lingüística Cognitiva (AELCO), Madrid.
- Cañas, A., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T. C., et al. (2005). Concept maps: integrating knowledge and information visualization. In S. Verlag (Ed.), *Knowledge and Information Visualuzation* (pp. 205-219): Springer Ve.
- Candel, D. (2004, 2004). Wüster par lui-meme. *Cahier du CIEL, Des fondements théoriques de la terminologie*, 15-31.
- Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What Are Ontologies, and Why Do We Need Them? *IEEE Intelligent Systems*, 20-26.
- Clancey, W. J. (2007). The knowledge level reinterpreted: Modeling socio-technical systems. *International Journal of Intelligent Systems*, 8(1), 33-49.
- Cocchiarella, N. B. (2001). Logic and ontology. *Axiomathes*, 12, 117-150.
- Conceição, M. C. (2005). *Concepts, Termes et Reformulations*: Presses Universitaires de Lyon.
- Cornejo, M. (2003). *Unity, Value and Knowledge Communities*.
- Costa, R. (1993). *Terminologia da Economia Monetária: Relações conceptuais e semânticas numa sistemática terminológica e lexicográfica*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa.
- Costa, R. (2005). *Terminologia, linguística de corpora e ontologias*. Paper presented at the TKE 2005 - 7th International Conference on Terminology and Knowledge Engineering. Workshop Terminology and Content Development, Copenhagen Business School.
- Costa, R. (2006). Plurality of Theoretical Approaches to Terminology. In H. Picht (Ed.), *Modern Approaches to Terminological Theories and Applications*. Berlin: Peter Lang Verlag.
- Costa, R., & Silva, R. (2006). *Metodologia para a investigação aplicada em Terminologia*: Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa. Document Number)
- Costa, R., & Silva, R. (2009). *De la typologie à l'ontologie de textes*. Paper presented at the Terminologie & Ontologie: Théories et Applications, Annecy.
- Cowan, R., David, P. A., & Foray, D. (1999). *The explicit economics of knowledge codification and tacitness*. Paper presented at the 3rd TIPIK Workshop. from <http://www-econ.stanford.edu/faculty/workp/swp99027.pdf>

- Cruse, D. A. (1986). *Lexical Semantics*: Cambridge University Press.
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What is knowledge representation? *AI Magazine*, 14, 17-33.
- Depecker, L. (2002). *Entre Signe et Concept - Éléments de terminologie générale*. Paris: Presses Sorbonne Nouvelle.
- Desmet, I. M. (1994, 18 Outubro 2008). *Terminologia, desenvolvimento e ensino / aprendizagem das LSP. O valo heurístico do plano textual na investigação terminológica e no ensino do português especializado*. Paper presented at the IV Simposio Iberoamericano de Terminologia- Terminología y Desarrollo, Buenos Aires.
- Duguid, P. (2005). The Art of Knowing: Social and Tacit Dimensions of Knowledge and the Limits of the Community of Practice. *The Information Society* 21(2), 109-118.
- Feliu, J., Vivaldi, J., & Cabré, M. T. (2006). *SKELETON: Specialized knowledge retrieval on the basis of terms and conceptual relations*. Paper presented at the LREC 2006, Genoa, Italy.
- Fieser, J., & Dowden, B. H. (2008). *The Internet Encyclopedia of Philosophy - Aristotle*: University of Tennessee at Martin.
- Freitas, F. L. G. d. (2003). Ontologias e web semântica. In R. Vieira & F. Osório (Eds.), *Anais do XXIII Congresso da Sociedade Brasileira de Computação. Jornada de Mini-Cursos em Inteligência Artificial*. (Vol. 8, pp. 1-52). Campinas.
- Fuller, S. (2001). Knowledge R.I.P.? Resurrecting knowledge requires rediscovering the university. *Tamara: Journal of Critical Postmodern Organization Science*.
- Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps? Making sense of it all. *Ontopia* Retrieved 2 March 2007, from <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
- Gillam, L., Tariq, M., & Ahmad, K. (2005). Terminology and the construction of ontology. *Terminology*, 11, 55-81.
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering - with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*: Springer Verlag.
- Gresser, J. Y. (2008). *Terminology and Information Science(s)*. Paper presented at the Information and Communication Technologies: From Theory to Applications, 2008. ICTA 2008, Damascus.
- Gruber, T. (1993a, 12 January 2007). What is an ontology? , from <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

- Gruber, T. (1993b). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.
- Gruber, T. (2008a). Ontology. *Encyclopedia of Database Systems*. Retrieved 31 November 2008, from <http://tomgruber.org/writing/ontology-definition-2007.htm>
- Gualtieri, A., & Ruffolo, M. (2005). *An ontology-based framework for representing organizational knowledge*. Paper presented at the 5th International Conference on Knowledge Management, Graz, Austria.
- Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, 43, 625-640.
- Guarino, N. (1997). Understanding, building and using ontologies. *International Journal of Human Computer Studies*, 46, 293-310.
- Guarino, N. (1998). *Formal ontology and information systems*. Paper presented at the FOIS 98, Trento, Italy.
- Guarino, N., Artale, A., Franconi, E., & Pazzi, L. (1996a). Part-whole relations in object-centered systems: an overview. *Data & Knowledge Engineering*, 20, 347-383.
- Guarino, N., Gangemi, A., Masolo, C., & Oltramari, A. (2001). *Understanding top-level ontological distinctions*. Paper presented at the IJCAI 2001 - Workshop on Ontologies and Information Sharing, Seattle, USA.
- Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In N. Mars (Ed.), *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995* (pp. 25-32). Amsterdam: IOS Press.
- Guarino, N., & Welty, C. (2001). Supporting ontological analysis of taxonomic relationships. *Data & Knowledge Engineering*(39), 51-74.
- Hjelm, H. (2009). *Cross-language ontology learning. Incorporating and exploiting cross-languauge data in the ontology.*,
- ISO 1087-2:2000 (E/F) Terminology work - Vocabulary - Part 2: Computer Applications, (2000a).
- ISO/FDIS 704:2000 (E) Terminology work - Principles and Methods, (2000b).
- ISO/FDIS 1087-1:2000 (E/F) Terminology work - Vocabulary - Part 1: Theory and Application, (2000c).
- Kageura, K., Daille, B., Nakagawa, H., & Chien, L.-F. (2004). Recent trends in computational terminology. In J. Benjamins (Ed.), *Terminology 10:1* (pp. 1-21): John Benjamins.
- Kocourek, R. (1991). Textes et termes. *META - The Translator's Journal*, 36(1), 71-76.

- Koepsell, D. R. (1999). Introduction to applied ontology: the philosophical analyses of everyday objects. *The American Journal of Economics and Sociology*, 58(2), 217-217.
- L'Homme, M.-C. (2004). *La terminologie: principes et techniques*. Montréal: Les Presses de l'Université de Montréal.
- Lopes, A. C. M., & Rio-Torto, G. (2007). *Semântica: Caminho*.
- Madsen, B. N., & Thomsen, H. E. (2008). *Terminological Principles Used for Ontologies*. Paper presented at the TKE 2008 Terminology and Knowledge Engineering, Copenhagen, Denmark.
- Maia, B. (2003). Ontology, ontologies, general language and specialised language. In CLUP (Ed.), *Volume Comemorativo dos 25 anos da CLUP* (pp. 23-29). Porto.
- McGuinness, D. L. (2003). Ontologies come of age. In D. Fensel, J. Hendler, H. Lieberman & W. Wahlster (Eds.), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*: MIT Press.
- Mendes, O. B. (2008). *A gestão da informação na biblioteca especializada - as linhas de fronteira da pesquisa de informação*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa, Lisboa.
- Meyer, I. (1991). *Knowledge management for terminology-intensive applications: needs and tools*. Paper presented at the ACL SIG Workshop on Lexical Semantics and Knowledge Representation, Berkeley: California.
- Mineiro, A., Doria, M., Antunes, M., & Correia, M. (2004). *Hiponímia e meronímia num corpus da Náutica em português europeu*. Paper presented at the IX Simpósio Ibero-Americano de Terminologia RITERM, Barcelona.
- Musen, M. A. (1992). Dimensions of knowledge sharing and reuse. In *Computers and Biomedical Research* (Vol. 25, pp. 435-467). San Diego, CA, USA: Academic Press Professional, Inc.
- Newell, A. (1980). The Knowledge Level. *AI Magazine* 1-33.
- Newell, A. (1982). The Knowledge Level. *Artificial Intelligence*, 87-127.
- Novak, J. D. (2000). The theory underlying concept maps and how to construct them [Electronic Version],
- Novak, J. D., & Cañas, A. (2006). *The theory underlying concept maps and how to construct and use them*. Pensacola - Florida: Florida Institute for Human and Machine Cognition. Document Number)
- O'Hara, K. (2002). Plato and the internet. In *Postmodern Encounters* (pp. 79): Icon Books.

- O'Hara, K. (2004). Ontologies and technologies: knowledge representation or misrepresentation. In *ACM SIGIR Forum* (Vol. 38, pp. 11-17): ACM New York, NY, USA.
- Pearson, J. (1998). *Terms in Context* (Vol. 1). Amsterdam: John Benjamins.
- Pombo, O. (1998). Da classificação dos seres à classificação dos saberes. *Leituras. Revista da Biblioteca Nacional de Lisboa*, 2, 19-33.
- Rastier, F. (1995). Le terme: entre ontologie et linguistique. *Le banque des mots*, 7, 35-65.
- Rastier, F. (2004). Ontologie(s). *Revue des sciences et technologies de l'information - Révue d'intelligence artificielle*, 18, 15-40.
- Rees, R. v. (2003). *Clarity in the usage of the terms ontology, taxonomy and classification*. Paper presented at the CIB W78's 20th International Conference on Construction IT, Construction IT Bridging the Distance, Waiheke Island, New Zealand.
- Reinberger, M.-L., & Daelemans, W. (2005). Unsupervised text mining for ontology learning. Dagstuhl.
- Review, M. I. (2002). Ten Taxonomy Myths. Retrieved 22 November 2008, from <http://www.montague.com/review/myths.html>
- Roberto, M. T. C. G. (1999). *A linguistic analysis of the in-trade catalogue produced by industrial business concerns of the Aveiro district*. Aveiro University, Aveiro.
- Roche, C. (2003). *Ontology: a survey*. Paper presented at the 8th Symposium on Automated Systems Based on Human Skill and Knowledge - IFAC 2003, Göteborg, Sweden
- Roche, C. (2003a). *The differentia principle as a cornerstone of ontology*. Paper presented at the Knowledge Management and Philosophy - WM 2003 Conference, Luzern.
- Roche, C. (2005, Mars 2005). Terminologie & Ontologie. *Langages*, 11.
- Roche, C. (2006). *How words map concepts*. Paper presented at the VORTE 2006 - EDOC Conference, Hong Kong.
- Roche, C. (2006a). Lexical and Conceptual Structures in Ontology. In *Advances in Applied Artificial Intelligence - Lecture Notes in Computer Science* (pp. 1034-1041): Springer Berlin / Heidelberg.
- Roche, C. (2007). *Le terme et le concept: fondements d'une ontoterminologie*. Paper presented at the Terminologie & Ontologie: Théories et Applications, Annecy.
- Roche, C. (2007a). *Saying is not modelling*. Paper presented at the 9th International Conference on Enterprise Information Systems, Madeira.
- Roche, C. (2008). *Faut-il revisiter les Principes terminologiques ?* Paper presented at the TOTH 2008.

- Roche, C. (2009). *Faut-il revisiter les Principes terminologiques ?* Paper presented at the TOTH 2008.
- Rogers, M. (2004). Multidimensionality in concept systems. In J. Benjamins (Ed.), *Terminology* 10:2 (pp. 215-240): John Benjamins.
- Sager, J. C. (1990). A Practical Course in Terminology Processing. In (pp. 255). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Simões, M. d. G. (2008). *Da abstracção à complexidade formal – relações conceptuais no tesouro*. Coimbra: Almedina, S.A.
- Sintek, M., Buitelaar, P., & Olejnik, D. (2004). *A formalization of ontology learning fom text*. Paper presented at the Workshop on Evaluation of Ontology-based Tools Germany.
- Smith, B. (2003). Ontology: an introduction. Longer draft: Ontology and information systems [Electronic Version], 1-97, from <http://ontology.buffalo.edu/smith/articles/ontologies.htm>
- Smith, B., & Ceusters, W. (2007). Ontology as the Core Discipline of Biomedical Informatics: Legacies of the Past and Recommendations for the Future Direction of Research. In G. D. Crnkovic & S. Stuart (Eds.), *Computing, Information, Cognition* (pp. 104-122): Newcastle: Cambridge Scholars Press.
- Smith, B., Ceusters, W., & Temmerman, R. (2005). *Wüsteria*. Paper presented at the Medical Informatics Europe 2005, Geneva.
- Soares, A. L., & Pereira, C. (2008). Ontology development in collaborative networks as a process of social construction of meaning. *Lecture Notes in Computer Science - Proceedings of the OTM Confederated International Workshops and Posters on On the Move to Meaningful Internet Systems: 2008 Workshops: ADI, AWeSoMe, COMBEK, EI2N, IWSSA, MONET, OnToContent + QSI, ORM, PerSys, RDDS, SEMELS, and SWWS* 5333, 605-614.
- Sowa, J. (2005). The challenge of knowledge soup. In J. Ramadas & S. Chunawala (Eds.), *Research Trends in Science, Technology and Mathematics Education* (pp. 55-90). Mumbai: Homi Bhabha Centre.
- Sowa, J. (2006). Building, Sharing, and Merging Ontologies. Retrieved 20 June 2006, from <http://www.jfsowa.com/ontology/ontoshar.htm>
- Temmerman, R. (2000). *Towards New Ways of Terminology Description*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tricot, C., & Roche, C. (2006). *Visualisation of Ontology: a focus and context approach*. Paper presented at the International Conference on Multidisciplinary Information Sciences & Technologies, Merida, Spain.

- Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *knowledge Engineering Review*, 11, 96-136.
- Wilks, Y. (2008). The semantic web: apotheosis of annotation, but what are its semantics? *IEEE Intelligent Systems*, 23(3), 41-49.
- Wilson, R. (2004). *The role of ontologies in teaching and learning*. Paper presented at the TechWatch report.

11.2 Bibliografia consultada

- Ahmad, K., & Gillam, L. (2005). Automatic Ontology Extraction from Unstructured Texts. *Lecture notes in computer science* 1330-1346.
- Ahmad, K., Musacchio, M. T., & Palumbo, G. (2006). *Ontological Terminological Commitments and the Discourse of Specialist Communities*. Paper presented at the LREC 2006, Genoa, Italy.
- Akinci, S. (2003). *Ontological commitment*. Paper presented at the Twenty-first World Congress of Philosophy, Istanbul, Turkey.
- Alani, H., Kim, S., Millard, D. E., Weal, M. J., Hall, W., Lewis, P. H., et al. (2003). Natural Language Processing - Automatic ontology-based knowledge extraction from web documents *IEEE Intelligent Systems* 18, 14-21.
- Almeida, E. R. L. (2008). *Avaliação da Biodegradabilidade Aeróbia de Efluentes Vinícolas*. Universidade de Aveiro, Aveiro.
- Almeida, M. B., & Bax, M. P. (2003). Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção *Ciência da Informação*, 32, 7-20.
- Anstein, S., Kremer, G., & Reyle, U. (2006). *Identifying and Classifying Terms in the Life Sciences: The Case of Chemical Terminology*. Paper presented at the LREC 2006, Genoa, Italy.
- Antia, B. E. (2000). *Terminology and language planning: an alternative framework of practice and discourse*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Aussenac-Gilles, N. (2005). *Supervised text analysis for ontology and terminology engineering*. Paper presented at the Dagstuhl Seminar on Machine Learning for the Semantic Web, Germany.

- Aussenac-Gilles, N., Biébow, B., & Szulman, S. (2000). *Corpus analysis for conceptual modelling*. Paper presented at the EKAW'2000 - 12th International Conference on Knowledge Engineering and Knowledge Management, French Riviera.
- Bachimont, B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds.), *Ingénierie des connaissances, évolutions récentes et nouveaux défis*. Paris: Eyrolles.
- Bachimont, B. (2006, 25 Novembro 2008). Qu'est-ce c'est une ontologie ? , from http://www.technolangu.net/imprimer.php3?id_article=280
- Barrière, C. (2004). Building a concept hierarchy from corpus analysis. In J. Benjamins (Ed.), *Terminology 10:2* (pp. 241-264): John Benjamins.
- Barros, S. (2007). *A relação genérica num corpus de direito constitucional português: identificação de contextos para a delimitação de candidatos a termo.*, FCT-UNL - Universidade Nova de Lisboa, Lisboa.
- Beck, H., & Pinto, H. S. (2002). Overview of approach, methodologies, standards, and tools for ontologies. *The Agricultural Ontology Service (UN FAO)*, Sections 1, 4, and 6.
- Béjoint, H., & Thoiron, P. (2000). *Le Sens en Terminologie*. Lyon: PUL.
- Bendaoud, R., Rouane, M., Hacene, Y. T., Delecroix, B., & Napoli, A. (2007). *Text-based ontology construction using relational concept analysis* Paper presented at the International Workshop on Ontology Dynamics (IWOD), Innsbruck, Austria
- Bento, J. R. (2008). *Da Constituição do Corpus à Construção de uma Ontologia e Base de Conhecimento Terminológicas*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa, Lisboa.
- Biebow, M., & Szulman, S. (1999). *TERMINAE: a method and a tool to build of a domain ontology*. Paper presented at the 11th European Workshop on Knowledge Acquisition, Modeling, and Management, Dagstuhl, Germany.
- Borges, M. M. (2002). *De Alexandria a Xanadu* (Vol. 2). Coimbra: Quarteto Editora.
- Bourigault, D., & Aussenac-Gilles, N. (2003). *Construction d'ontologies à partir de textes*. Paper presented at the TALN 2003 Batz-sur-Mer.
- Bourigault, D., Aussenac-Gilles, N., & Charlet, J. (2002). Construction de ressources terminologiques ou ontologiques à partir de textes: un cadre unificateur pour trois études de cas. *Revue d'Intelligence Artificielle*, X, 1-X.
- Bourigault, D., Jacquemin, C., & (eds.), M.-C. L. H. (2001). *Recent Advances in Computational Terminology* (Vol. 2). Amsterdam: John Benjamins.

- Brewster, C. (2003). *Knowledge Maintenance and the Frame Problem*. Paper presented at the 6th Annual CLUK Research Colloquium, Edinburgh, United Kingdom.
- Brewster, C., Ciravegna, F., & Wilks, Y. (2003a). *Background and Foreground Knowledge in Dynamic Ontology Construction*. Paper presented at the Proceedings of the Semantic Web Workshop -SIGIR 2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada.
- Brewster, C., Iria, J., Ciravegna, F., & Wilks, Y. (2005). *The Ontology: Chimaera or Pegasus*. Paper presented at the Dagstuhl Seminar Machine Learning for the Semantic Web,, Dagstuhl
- Brewster, C., & O'Hara, K. (2004). Knowledge Representation with Ontologies: The Present and Future. *IEEE Intelligent Systems*, 19(1), 72-81.
- Brewster, C., & Wilks, Y. (2004a). *Ontologies, Taxonomies and Thesauri: Learning from Texts*. Paper presented at the The Use of Computational Linguistics in the Extraction of Keyword Information from Digital Library Content, King's College London
- Buitelaar, P. (2000). *Semantic lexicons: between ontology and terminology*. Paper presented at the OntoLex 2000, Sozopol, Bulgaria.
- Buitellar, P., Cimiano, P., & Magnini, B. (2005). Ontology learning from text: an overview. In B. Buitelaar, P. Cimiano & B. Magnini (Eds.), *Ontology Learning from Text: Methods, evaluation and applications* (pp. 3-12): IOS Press.
- Cabré, M., Condamines, A., & Ibekwe-SanJuan, F. (2005). Introduction - Application-driven terminology engineering. In J. Benjamins (Ed.), *Terminology* (Vol. 11:1, pp. 1-19): John Benjamins.
- Cabré, M. T. (1998 (1992)). *Terminology: Theory, Methods and Applications*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Cabré, M. T. (2000). Elements for a theory of terminology: Towards an alternative paradigm. In J. Benjamins (Ed.), *Terminology* (Vol. 6:1, pp. 1-23). Amsterdam: John Benjamins.
- Cabré, M. T. (2003). Investigar en terminología: posibilidades y líneas de trabajo. In E. Ortega (Ed.), *Panorama actual de la investigación e interpretación* (Vol. I).
- Cabré, M. T. (2003a). Theories of terminology - their description, prescription and explanation. In J. Benjamins (Ed.), *Terminology* (Vol. 9:2, pp. 163-199). Amsterdam: John Benjamins.
- Cabré, M. T. (2006). *Morfología y terminología*. Paper presented at the La morfología a debate, Jaén.
- Cabré, M. T., & Estopà, R. (2002). El conocimiento especializado y sus unidades de representación: diversidad cognitiva. *Sendébar*, 13, 141-153.

- Cabré, M. T., & Feliu, J. (2002a). *Conceptual relations in specialized texts: new typology and an extraction system proposal*. Paper presented at the TKE 2002 - 6th International Conference on Terminology and Knowledge Engineering, Nancy.
- Cabré, M. T., Feliu, J., & Tebé, C. (2001b). *Bases cognitivas de la terminología: hacia una visión comunicativa del concepto*. Paper presented at the II Congreso de la Asociación Española de Lingüística Cognitiva (AELCO), Madrid.
- Campenhoudt, M. v. (2006). *Que nous reste-t-il d'Eugen Wüster?* Paper presented at the Eugen Wüster et la Terminologie de l'École de Vienne, Paris.
- Cañas, A., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T. C., et al. (2005). Concept maps: integrating knowledge and information visualization. In S. Verlag (Ed.), *Knowledge and Information Visualuzation* (pp. 205-219): Springer Ve.
- Candel, D. (2004, 2004). Wüster par lui-meme. *Cahier du CIEL, Des fondements théoriques de la terminologie*, 15-31.
- Ceccaroni, L. (2001). *OntoWEDSS - An Ontology-based Environmental Decision-Support System for the management of Wastewater treatment plants*. Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.
- Ceccaroni, L., Cortés, U., & Sànchez-Marrè, M. (2002). *OntoWEDSS: an ontology-underpinned decision-support system for wastewater management*. Paper presented at the Special session iEMSs 2002 – BESAI: Binding Environmental Sciences and Artificial Intelligence, Lugano, Switzerland.
- Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What Are Ontologies, and Why Do We Need Them? *IEEE Intelligent Systems*, 20-26.
- Clancey, W. J. (2007). The knowledge level reinterpreted: Modeling socio-technical systems. *International Journal of Intelligent Systems*, 8(1), 33-49.
- Cocchiarella, N. (1996). Conceptual realism as formal ontology. In R. Poli & P. Simons (Eds.), *Formal Ontology* (pp. 27-60): Kluwer.
- Cocchiarella, N. B. (2001). Logic and ontology. *Axiomathes*, 12, 117-150.
- Conceição, M. C. (2005). *Concepts, Termes et Reformulations*: Presses Universitaires de Lyon.
- Corazzon, R. (2002, 26 Agosto 2006). Ontology - A resource guide for philosophers. from <http://www.formalontology.it/>
- Cornejo, M. (2003). *Unity, Value and Knowledge Communities*.
- Cortès, J. F. i. (2004). *Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica*. Universitat Pompeu Fabra.

- Costa, R. (1993). *Terminologia da Economia Monetária: Relações conceptuais e semânticas numa sistemática terminológica e lexicográfica*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa.
- Costa, R. (2001). *Pressupostos Teóricos e Metodológicos para a Extração Automática de Unidades Terminológicas Multilexémicas*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa, Lisboa.
- Costa, R. (2005). *Terminologia, linguística de corpora e ontologias*. Paper presented at the TKE 2005 - 7th International Conference on Terminology and Knowledge Engineering. Workshop Terminology and Content Development, Copenhagen Business School.
- Costa, R. (2006). Plurality of Theoretical Approaches to Terminology. In H. Picht (Ed.), *Modern Approaches to Terminological Theories and Applications*. Berlin: Peter Lang Verlag.
- Costa, R., & Silva, R. (2006a). *Metodologia para a investigação aplicada em Terminologia*: Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa. Document Number)
- Costa, R., & Silva, R. (2009). *De la typologie à l'ontologie de textes*. Paper presented at the Terminologie & Ontologie: Théories et Applications, Annecy.
- Cowan, R., David, P. A., & Foray, D. (1999). *The explicit economics of knowledge codification and tacitness*. Paper presented at the 3rd TIPIK Workshop. from <http://www-econ.stanford.edu/faculty/workp/swp99027.pdf>
- Cruse, D. A. (1986). *Lexical Semantics*: Cambridge University Press.
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What is knowledge representation? *AI Magazine*, 14, 17-33.
- Depecker, L. (2002). *Entre Signe et Concept - Éléments de terminologie générale*. Paris: Presses Sorbonne Nouvelle.
- Desmet, I. M. (1994, 18 Outubro 2008). *Terminologia, desenvolvimento e ensino / aprendizagem das LSP. O valor heurístico do plano textual na investigação terminológica e no ensino do português especializado*. Paper presented at the IV Simposio Iberoamericano de Terminologia- Terminología y Desarrollo, Buenos Aires.
- Duguid, P. (2005). The Art of Knowing: Social and Tacit Dimensions of Knowledge and the Limits of the Community of Practice. *The Information Society* 21(2), 109-118.
- Felber, H. (2001). *Allgemeine Terminologielehre, Wissenslehre und Wisstechnik - Theoretische Grundlagen und philosophische Betrachtungen*. Wien, Austria: TermNet.
- Feliu, J., Vivaldi, J., & Cabré, M. T. (2006). *SKELETON: Specialized knowledge retrieval on the basis of terms and conceptual relations*. Paper presented at the LREC 2006, Genoa, Italy.

- Fieser, J., & Dowden, B. H. (2008). *The Internet Encyclopedia of Philosophy - Aristotle*: University of Tennessee at Martin.
- Fragkou, P., Petasis, G., Theodorakos, A., Karkaletsis, V., & Spyropoulos, C. (2008). *BOEMIE Ontology-Based Text Annotation Tool*. Paper presented at the Sixth International Language Resources and Evaluation (LREC'08, Marrakech, Morocco).
- Freitas, F. L. G. d. (2003). Ontologias e web semântica. In R. Vieira & F. Osório (Eds.), *Anais do XXIII Congresso da Sociedade Brasileira de Computação. Jornada de Mini-Cursos em Inteligência Artificial*. (Vol. 8, pp. 1-52). Campinas.
- Fuller, S. (2001b). Knowledge R.I.P.? Resurrecting knowledge requires rediscovering the university. *Tamara: Journal of Critical Postmodern Organization Science*.
- Fuller, S. (2001c). Strategies of Knowledge Integration. In M.K.Tolba (Ed.), *Our Fragile World: Challenges, Opportunities for Sustainable Development* (pp. 1215-1228). Oxford: EOLSS Publishers for UNESCO.
- Gangemi, A., Guarino, N., & Oltramari, R. (2001). *Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level*. Paper presented at the FOIS 2001 - International conference on Formal Ontology in Information Systems, Ogunquit, Maine
- Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps? Making sense of it all. *Ontopia* Retrieved 2 March 2007, from <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
- Gillam, L., & Tariq, M. (2004). *Ontology via Terminology?* Paper presented at the Workshop on Terminology, Ontology and Knowledge Representation (Termino) 2004, Lyon, France.
- Gillam, L., Tariq, M., & Ahmad, K. (2005). Terminology and the construction of ontology. In J. Benjamins (Ed.), *Terminology 11:1* (pp. 55-81): John Benjamins.
- Giovannetti, E., Marchi, S., Montemagni, S., & Bartolini, R. (2008). *Ontology Learning and Semantic Annotation: a Necessary Symbiosis*. Paper presented at the Sixth International Language Resources and Evaluation (LREC'08), Marrakech, Morocco.
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering - with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*: Springer Verlag.
- Gonçalves, P. M. (1998). *Tratamento de Efluentes Vinícolas por Digestão Anaeróbia*. Universidade do Minho, Braga.
- Gresser, J. Y. (2008). *Terminology and Information Science(s)*. Paper presented at the Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008, Damascus.

- Gruber, T. (1993). Towards principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies* 43(5-6), 907-928.
- Gruber, T. (1993a, 12 January 2007). What is an ontology? , from <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- Gruber, T. (1993b). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.
- Gruber, T. (2004). Every Ontology is a Treaty- Interview for Semantic Web and Information Systems SIG of the Association for Information Systems *SIGSEMIS Bulletin*, 1.
- Gruber, T. (2008). Despite our Best Efforts, Ontologies are not the Hard Part. In A. S. Symposium (Ed.).
- Gruber, T. (2008a). Ontology. *Encyclopedia of Database Systems* Retrieved 31 November 2008, from <http://tomgruber.org/writing/ontology-definition-2007.htm>
- Gualtieri, A., & Ruffolo, M. (2005). *An ontology-based framework for representing organizational knowledge*. Paper presented at the 5th International Conference on Knowledge Management, Graz, Austria.
- Guarino, N. (1992). Concepts, attributes and arbitrary relations: some linguistic and ontological criteria for structuring knowledge bases. *Data & Knowledge Engineering*(8), 249-261.
- Guarino, N. (1995a). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, 43, 625-640.
- Guarino, N. (1997). Understanding, building and using ontologies. *International Journal of Human Computer Studies*, 46, 293-310.
- Guarino, N. (1998). *Formal ontology and information systems*. Paper presented at the FOIS 98, Trento, Italy.
- Guarino, N., Artale, A., Franconi, E., & Pazzi, L. (1996). Part-whole relations in object-centered systems: an overview. *Data & Knowledge Engineering*(20), 347-383.
- Guarino, N., Gangemi, A., Masolo, C., & Oltramari, A. (2001). *Understanding top-level ontological distinctions*. Paper presented at the IJCAI 2001 - Workshop on Ontologies and Information Sharing, Seattle, USA.
- Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In N. Mars (Ed.), *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995* (pp. 25-32). Amsterdam: IOS Press.
- Guarino, N., & Welty, C. (2001a). Supporting ontological analysis of taxonomic relationships. *Data & Knowledge Engineering*(39), 51-74.

- Hagengruber, R. (2005). *Knowledge in action - the use of formal philosophical ontologies in knowledge management*. Paper presented at the Philosophisches Seminar Universität Koblenz.
- Hahn, U., Klenner, M., & Schnattinger, K. (1996). Learning from texts - a terminological metareasoning perspective. In S. Wermter, E. Riloff & G. Scheler (Eds.), *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing* (pp. 453-468). Berlin: Springer.
- Hahn, U., & Schnattinger, K. (1998). *Ontology engineering via text understanding*. Paper presented at the 15th World Computer Congress - The Global Information Society on the Way to the Next Millennium, Viena.
- Hahn, U., & Schnattinger, K. (1998a). *Towards text knowledge engineering*. Paper presented at the AAAI'98 - 15th National Conference on Artificial Intelligence, Madison, Wisconsin.
- Hassell, L. (2003). *A Philosophical Look at Knowledge Management*. Paper presented at the Workshop on Philosophy and Knowledge Management - WM 2003 Luzern
- Heil, J. (2003). *From an Ontological Point of View*: Oxford University Press.
- Hitzler, P., Krötzsch, M., & Sure, M. E. a. Y. (2005). *What Is Ontology Merging? – A Category-Theoretical Perspective Using Pushouts*. Paper presented at the First International Workshop on Contexts and Ontologies (C&O-05), Pittsburgh, PA.
- Hjelm, H. (2009). *Cross-language ontology learning. Incorporating and exploiting cross-language data in the ontology.*,
- Horecky, J. (1997). Intension and extension of a term. *Human Affairs*, 7, 134-140.
- Hovy, E. (2005). Methodologies for the Reliable Construction of Ontological Knowledge. Conceptual Structures: Common Semantics for Sharing Knowledge - 13th Annual International Conference on Conceptual Structures (ICCS 2005) In F. Dau, M.-L. Mugnier & G. Stumme (Eds.), *Springer Lecture Notes in AI*. (Vol. 3596, pp. 91-106): Springer-Verlag.
- Hurford, J. R. (2007). *The Origins of Meaning. The Language in the Light of Evolution.*: Oxford University Press.
- Kageura, K., Daille, B., Nakagawa, H., & Chien, L.-F. (2004). Recent trends in computational terminology. In J. Benjamins (Ed.), *Terminology 10:1* (pp. 1-21): John Benjamins.
- Kashyap, V. (2001). Information modeling on the Web: the role of metadata, semantics and ontologies. In *Practical Handbook of Internet Computing* (pp. 7-28): CRC Press
- Kocourek, R. (1982). *La Langue Française de la Technique et de la Science*. Wiesbaden: Oscar Brandstetter Verlag GmbH & Co. Kg.
- Kocourek, R. (1991). Textes et termes. *META - The Translator's Journal*, 36(1), 71-76.

- Koepsell, D. R. (1999). Introduction to applied ontology: the philosophical analyses of everyday objects. *The American Journal of Economics and Sociology*, 58(2), 217-217.
- Köhler, J., Munn, K., Rüegg, A., Skusa, A., & Smith, B. (2006). Quality Control for Terms and Definitions in Ontologies and Taxonomies. *BMC Bioinformatics*, 7(212).
- L'Homme, M.-C. (2004). *La terminologie: principes et techniques*. Montréal: Les Presses de l'Université de Montréal.
- Leenheer, P. D., Moor, A. d., & Meersman, R. (2007). Context Dependency Management in Ontology Engineering: A Formal Approach In *Journal on Data Semantics VIII* (Vol. 4380/2007, pp. 26-56): Springer Berlin /Heidelberg.
- Lopes, A. C. M., & Rio-Torto, G. (2007). *O essencial sobre Língua Portuguesa: semântica: Caminho*.
- Madsen, B. N. (2005). *Proceedings of TKE2005: 7th Internatinal Conference on Terminology and Content Development. Terminology and Knowledge Engineering*. Paper presented at the TKE2005: 7th Internatinal Conference on Terminology and Knowledge Engineering, Copenhagen.
- Madsen, B. N., & Thomsen, H. E. (2008). *Terminological Principles Used for Ontologies*. Paper presented at the TKE 2008 Terminology and Knowledge Engineering, Copenhagen, Denmark.
- Maedche, A., & Staab, S. (2000). *Discovering Conceptual Relations from Text*. Paper presented at the 14th European Conference on Artificial Intelligence ECAI 2000, Berlim, Germany.
- Maedche, A., & Staab, S. (2000a). *Ontology learning*. Paper presented at the Ontology Learning ECAI-2000 Workshop, Berlin.
- Maia, B. (2003). Ontology, ontologies, general language and specialised language. In CLUP (Ed.), *Volume Comemorativo dos 25 anos da CLUP* (pp. 23-29). Porto.
- Malaisé, V., Zweigenbaum, P., & Bachimont, B. (2005). Mining defining contexts to help structuring differential ontologies. In J. Benjamins (Ed.), *Terminology 11:1* (pp. 21-53): John Benjamins.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics* Edinburgh: Edinburgh University Press.
- McGuinness, D. L. (2003). Ontologies come of age. In D. Fensel, J. Hendler, H. Lieberman & W. Wahlster (Eds.), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*: MIT Press.
- Mendes, O. B. (2008). *A gestão da informação na biblioteca especializada - as linhas de fronteira da pesquisa de informação*. Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa, Lisboa.

- Meyer, I. (1991). *Knowledge management for terminology-intensive applications: needs and tools*. Paper presented at the ACL SIG Workshop on Lexical Semantics and Knowledge Representation, Berkeley: California.
- Mineiro, A., Doria, M., Antunes, M., & Correia, M. (2004). *Hiponímia e meronímia num corpus da Náutica em português europeu*. Paper presented at the IX Simpósio Ibero-Americano de Terminologia RITERM, Barcelona.
- Montero, F. A., Sanchez, A. V., & Perez, F. S. (2008). *Conceptual Modeling of Ontology-based Linguistic Resources with a Focus on Semantic Relations*. Paper presented at the Sixth International Language Resources and Evaluation (LREC'08), Marrakech, Morocco.
- Musen, M. A. (1992). Dimensions of knowledge sharing and reuse. In *Computers and Biomedical Research* (Vol. 25, pp. 435-467). San Diego, CA, USA: Academic Press Professional, Inc.
- Newell, A. (1980). The Knowledge Level. *AI Magazine* 1-33.
- Nirenburg, S., & Wilks, Y. (2001). What's in a symbol: ontology, representation and language. *Journal of Experimental and Theoretical Artificial Intelligence (JETAI)* 1-18.
- Nonato, R. d. S., & Lima, G. Â. B. d. O. (2007). *A teoria do conceito aplicada à determinação de links hipertextuais: considerações sobre modelagem conceitual*. Paper presented at the VIII ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação. Salvador da Bahia, Brasil.
- Novak, J. D., & Cañas, A. (2006). *The theory underlying concept maps and how to construct and use them*. Pensacola - Florida: Florida Institute for Human and Machine Cognition. Document Number)
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Document Number)
- Nuopponen, A. (2003). Terminology. In W. Frawley (Ed.), *The International Encyclopedia of Linguistics* (Vol. 4): Oxford University Press.
- O'Hara, K. (2002). Plato and the internet. In *Postmodern Encounters* (pp. 79): Icon Books.
- O'Hara, K. (2004). Ontologies and technologies: knowledge representation or misrepresentation. In *ACM SIGIR Forum* (Vol. 38, pp. 11-17): ACM New York, NY, USA.
- Oldfield, P. (2002). Domain modelling [Electronic Version]. *Appropriate Process Movement*, 16, from www.aptprocess.com
- Pearson, J. (1998). *Terms in Context* (Vol. 1). Amsterdam: John Benjamins.
- Philosophy, S. E. o. (2001). Model Theory, *Stanford Encyclopedia of Philosophy*.

- Pinto, H. S., & Martins, J. P. (2004). Ontologies: How can They be Built? *Knowledge and Information Systems*, 6(4), 441-464.
- Pombo, O. (1998). Da classificação dos seres à classificação dos saberes. *Leituras. Revista da Biblioteca Nacional de Lisboa*, 2, 19-33.
- Quine, W. V. (2004). *Quintessence: Basic Readings from the Philosophy of W. V. Quine*. London: The Belknap Press of Harvard University Press.
- Ranchhod, E. M. (2001). *Tratamento das Línguas por Computador: Uma introdução à linguística computacional e suas aplicações*. Lisboa: Editorial Caminho.
- Rastier, F. (1995). Le terme: entre ontologie et linguistique. *Le banque des mots*, 7, 35-65.
- Rastier, F. (1998). On signs and texts: cognitive science faces interpretation. *Applied Semiotics/Sémiotique appliquée* 2, 195-244.
- Rastier, F. (2004). Ontologie(s). *Revue des sciences et technologies de l'information - Révue d'intelligence artificielle*, 18, 15-40.
- Rees, R. v. (2003). *Clarity in the usage of the terms ontology, taxonomy and classification*. Paper presented at the CIB W78's 20th International Conference on Construction IT, Construction IT Bridging the Distance, Waiheke Island, New Zealand.
- Reinberger, M.-L., & Daelemans, W. (2005). Unsupervised text mining for ontology learning. Dagstuhl.
- Reppen, R., Fitzmaurice, S. M., & (eds.), D. B. (2002). *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Review, M. I. (2002). Ten Taxonomy Myths. Retrieved 22 November 2008, from <http://www.montague.com/review/myths.html>
- Roberto, M. T. C. G. (1987). Meaningful Learning. In C. I. d. F. d. Professores (Ed.) (pp. 17). Aveiro: Universidade de Aveiro.
- Roberto, M. T. C. G. (1999). *A linguistic analysis of the in-trade catalogue produced by industrial business concerns of the Aveiro district*. Aveiro University, Aveiro.
- Roche, C. (2003). *Ontology: a survey*. Paper presented at the 8th Symposium on Automated Systems Based on Human Skill and Knowledge - IFAC 2003, Göteborg, Sweden
- Roche, C. (2003a). *The differentia principle as a cornerstone of ontology*. Paper presented at the Knowledge Management and Philosophy - WM 2003 Conference, Luzern.
- Roche, C. (2005, Mars 2005). Terminologie & Ontologie. *Langages*, 11.
- Roche, C. (2006). *How words map concepts*. Paper presented at the VORTE 2006 - EDOC Conference, Hong Kong.

- Roche, C. (2006a). Lexical and Conceptual Structures in Ontology. In *Advances in Applied Artificial Intelligence - Lecture Notes in Computer Science* (pp. 1034-1041): Springer Berlin / Heidelberg.
- Roche, C. (2007). *Le terme et le concept: fondements d'une ontoterminologie*. Paper presented at the Terminologie & Ontologie: Théories et Applications, Annecy.
- Roche, C. (2007a). *Saying is not modelling*. Paper presented at the 9th International Conference on Enterprise Information Systems, Madeira.
- Roche, C. (2009). *Faut-il revisiter les Principes terminologiques ?* Paper presented at the TOTH 2008.
- Rogers, M. (2004). Multidimensionality in concept systems. In J. Benjamins (Ed.), *Terminology 10:2* (pp. 215-240): John Benjamins.
- Rondeau, G. (1981). *Introduction à la Terminologie*. Québec, Canada: Chicoutimi.
- Rousseaux, F. (2003). *Knowledge acquisition or manifestation of the thought? Cathedral builders and knowledge acquisition method designers* Paper presented at the CEUR - Knowledge Management and Philosophy, Luzern.
- Sager, J. C. (1990). A Practical Course in Terminology Processing. In (pp. 255). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Sampson, G., & (eds.), D. M. (2004). *Corpus Linguistics - Readings in a Widening Discipline*. London: Continuum International Publishing Group Ltd.
- Schmidt, R. A. (1994). *Terminological Logics and Conceptual Graphs: An Historical Perspective*. Paper presented at the KI-94 Workshops - Extended Abstracts, Saarbrücken, Germany.
- Shönefeld, D. (1999). Corpora, linguistics and cognition. *International Journal of Corpus Linguistics*, 4(1), 137-171.
- Simões, M. d. G. (2008). *Da abstracção à complexidade formal – relações conceptuais no tesouro*. Coimbra: Almedina, S.A.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation (Describing English Language)*: Oxford University Press
- Sintek, M., Buitelaar, P., & Olejnik, D. (2004). *A formalization of ontology learning fom text*. Paper presented at the Workshop on Evaluation of Ontology-based Tools Germany.
- Smith, B. (1998). The basic tools of formal ontology. In N. Guarino (Ed.), *Formal Ontology in Information Systems* (Vol. Frontiers in Artificial Intelligence and Applications, pp. 19-28). Amsterdam: IOS Press.

- Smith, B. (2001). Objects and Their Environments: From Aristotle to Ecological Psychology. In A. Frank, J. Raper & J.-P. Cheylan (Eds.), *The Life and Motion of Socio-Economic Units* (pp. 79-97). London: Taylor and Francis.
- Smith, B. (2003). Ontology: an introduction. Longer draft: Ontology and information systems [Electronic Version], 1-97, from <http://ontology.buffalo.edu/smith/articles/ontologies.htm>
- Smith, B. (2004). *Beyond concepts: ontology as reality representation*. Paper presented at the FOIS 2004 - International Conference on Formal Ontology and Information Systems, Turin, Italy.
- Smith, B., & Ceusters, W. (2007). Ontology as the Core Discipline of Biomedical Informatics: Legacies of the Past and Recommendations for the Future Direction of Research. In G. D. Crnkovic & S. Stuart (Eds.), *Computing, Information, Cognition* (pp. 104-122): Newcastle: Cambridge Scholars Press.
- Smith, B., Ceusters, W., & Temmerman, R. (2005). *Wüsteria*. Paper presented at the Medical Informatics Europe 2005, Geneva.
- Smith, B., & Klein, G. O. (2005). Concept Systems and Ontologies. Recommendations based on discussions between realist philosophers and ISO/CEN experts concerning the standards addressing "concepts" and related terms [Electronic Version], from <http://ontology.buffalo.edu/concepts/ConceptsandOntologies.pdf>
- Smith, B., & Mulligan, K. (1985). Franz Brentano on the Ontology of Mind. *Philosophy and Phenomenological Research*, 45, 627-644.
- Smith, B., & Rosse, C. (2004). The Role of Foundational Relations in the Alignment of Biomedical Ontologies. In M. Fieschi, E. Coiera & Y.-C. J. Li (Eds.), *Medinfo 2004 - Proceedings of the 11th World Congress on Medical Informatics* (pp. 444-448): IOS Press.
- Soares, A. L., & Pereira, C. (2008). Ontology development in collaborative networks as a process of social construction of meaning. *Lecture Notes in Computer Science - Proceedings of the OTM Confederated International Workshops and Posters on On the Move to Meaningful Internet Systems: 2008 Workshops: ADI, AWeSoMe, COMBEK, EI2N, IWSSA, MONET, OnToContent + QSI, ORM, PerSys, RDDS, SEMELS, and SWWS 5333*, 605-614.
- Soares, A. L., & Simões, D. (2006). The Socio-Technical Design of a SME Knowledge Community in the Construction Industry. In *Network-centric Collaboration and Supporting Frameworks* (Vol. 224/2006, pp. 361-370): Springer Boston.
- Soares, A. L., Simões, D., & Silva, M. (2006). *Selecting and structuring semantic resources to support SMEs knowledge communities*. Paper presented at the 12th IFAC Symposium on Information Control Problems in Manufacturing (INCOM'2006), Paphos, Cyprus.

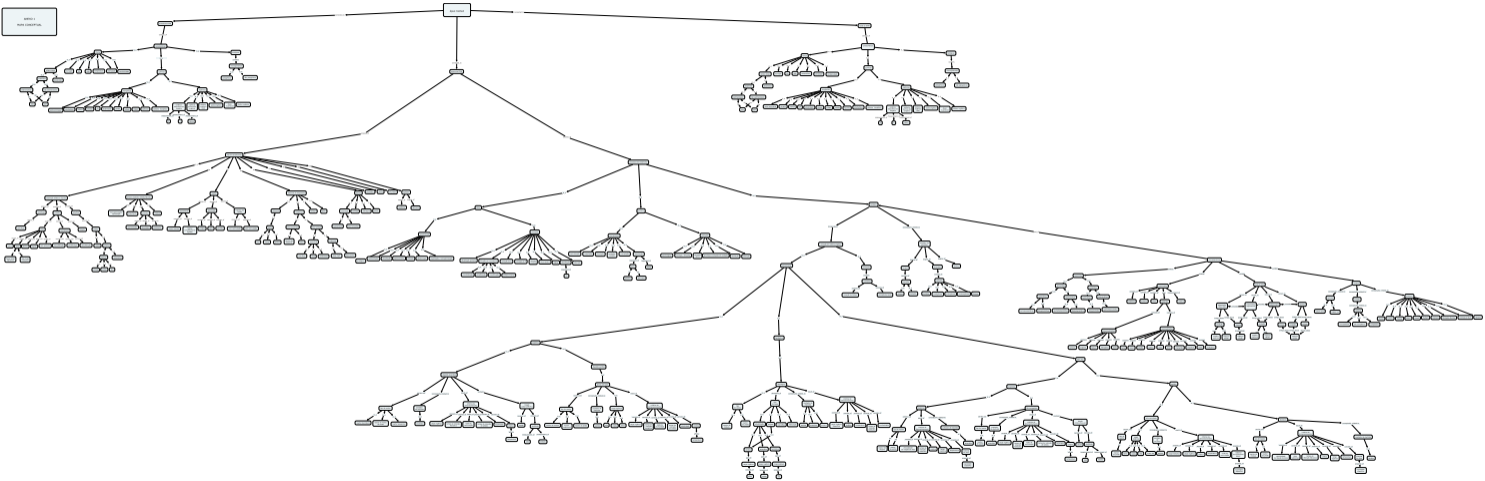
- Sowa, J. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks/Cole Publishing Co.
- Sowa, J. (2005). The Challenge of Knowledge Soup. In J. R. S. Chunawala (Ed.), *Research Trends in Science, Technology and Mathematics Education* (pp. 55-90). Mumbai: Homi Bhabha Centre.
- Sowa, J. (2006). Building, Sharing, and Merging Ontologies. Retrieved 20 June 2006, from <http://www.ifsowa.com/ontology/ontoshar.htm>
- Sowa, J. (2006a). A dynamic theory of ontology. In B. Bennett & C. Fellbaum (Eds.), *Formal Ontology in Information Systems* (pp. 204-213). Amsterdam: IOS Press.
- Sowa, J. (2006b). Categorization in cognitive computer science. In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science* (pp. 141-163): Elsevier.
- Tchobanoglous, G., & Burton, F. L. (2004). *Wastewater Engineering: Treatment and Reuse*: McGraw-Hill
- Temmerman, R. (2000). *Towards New Ways of Terminology Description*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Termnet. (2008). What is terminology? [Electronic Version]. Retrieved 12 November 2008, from http://www.termnet.org/downloads/english/about_us/what_is_terminology_2006_05.pdf
- Tricot, C., & Roche, C. (2006). *Visualisation of Ontology: a focus and context approach*. Paper presented at the International Conference on Multidisciplinary Information Sciences & Technologies, Merida, Spain.
- Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *knowledge Engineering Review*, 11, 96-136.
- Uschold, M., & King, M. (1995). *Towards a methodology for building ontologies*. Paper presented at the Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95, Quebec, Canada.
- Vossen, P. (2003). Ontologies. In R. Mitkov (Ed.), *Handbook Of Computational Linguistics*: Oxford University Press.
- Wilks, Y. (2002). *Ontotherapy: or, how to stop worrying about what there is* Paper presented at the Ontolex 2002 (Workshop held in conjunction with LREC 2002), Invited presentation Las Palmas, Canary Islands.
- Wilks, Y. (2008). The semantic web: apotheosis of annotation, but what are its semantics? *IEEE Intelligent Systems*, 23(3), 41-49.

- Wilson, R. (2004). *The role of ontologies in teaching and learning*. Paper presented at the TechWatch report.
- Wittenburg, P. (2006). *Perspectives for Ontologies in Linguistics*. Paper presented at the LREC2006, Genova.
- Wright, S. E. (2007). Coping with indeterminacy: terminology and knowledge representation. Resources in digital environments. In B. Antia (Ed.), *Indeterminacy in Terminology and LSP: Studies in honour of Heribert Picht*.
- Wright, S. E., & (eds.), G. B. (1997). *Handbook of Terminology Management. Volume I - Basic Aspects of Terminology Management* (Vol. I). Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Wright, S. E., & (eds.), G. B. (1997a). *Handbook of Terminology Management. Volume II - Applications-Oriented Terminology Management* (Vol. II). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Wüster, E. (1985). *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*: Infoterm.

Anexo 1

1. Mapa conceptual
2. Rede lexical

00001
NEW CONCEPT



ANEXO 2
REDE LEXICAL

