

Claudia Amaral Santos ISCA - Universidade de Aveiro
Campus Universitário de Santiago
3810-193 Aveiro - Portugal (+351) 234 380 110
claudia.amaral@ua.pt

A apresentação intitulada “Abordagens Terminológicas na Construção de Ontologias” pretende reflectir sobre as metodologias utilizadas pela terminologia na construção de ontologias enquanto recursos para representação do conhecimento, bem como sobre as funções do terminólogo neste novo enquadramento.

Do ponto de vista das ciências da linguagem, temos duas grandes orientações metodológicas da terminologia : a orientação semasiológica, que parte do termo para o conceito, fundamentalmente através da análise de texto, e a orientação onomasiológica, que parte do conceito para o termo, contando essencialmente com a contribuição do especialista de domínio.

O conhecimento é um elemento cognitivo e não pode ser transportado de um lado para o outro. Dado que a língua natural é um dos veículos privilegiados da comunicação humana e que os especialistas de domínio transmitem o seu conhecimento normalmente através de o texto, decidimos observar, se seria possível, por um lado, extrair conhecimento através da análise de texto, e conseqüentemente, se seria possível extrair ontologias a partir de dados textuais de forma automática ou semi-automática ; e, por outro lado, saber qual o papel do especialista de domínio e do ponto de partida extralinguístico no processo de construção de ontologias, para que pudéssemos chegar a conclusões relativamente à relevância, amplitude e profundidade das representações de conhecimento das duas abordagens.

Uma ontologia é um recurso conceptual expresso através de um artefacto de engenharia, um produto de software, desenhado e construído de acordo com um determinado objectivo, para uma determinada aplicação e para um determinado cliente. É uma descrição formal de conceitos e relações entre conceitos que podem existir para um agente ou comunidade de agentes, pretendendo fornecer uma

descrição partilhada e consensual de um domínio de especialidade, expressa numa linguagem legível por máquinas.

Para a inteligência artificial, só existe o que pode ser representado numa linguagem formal. Para que um sistema possa realizar inferências, tem de basear-se em dados perfeitamente claros e isentos de ambiguidade. O conhecimento é definido de uma forma estritamente funcional e baseado no princípio da racionalidade, onde a verdade foi suplantada pela eficiência da computação. Assim, não temos acesso ao conhecimento, mas a representações do conhecimento que funcionam como substitutos, aproximações ao real. A informação essencial contida numa ontologia deriva em grande parte do conjunto de conceitos que reflectem uma forma de pensar sobre o mundo. A esse conjunto de conceitos chamamos compromisso ontológico, que se prende, por um lado, com a selecção de conceptualizações e, no âmbito computacional, com a especificação dessas conceptualizações. Existe pois uma relação entre a natureza do raciocínio e a natureza do mundo, entrando em linha de conta com variáveis económicas e sociais, o papel do conhecimento codificado e do conhecimento tácito para fins de reutilização e partilha, e o tipo de conhecimento em que vale a pena investir e capturar. Por isso a representação sistemática da realidade através de linguagens simbólicas ainda não é proporcional à amplitude e profundidade do conhecimento representado.

Na perspectiva semasiológica, o objecto da terminologia é fundamentalmente o termo. Um termo constitui-se como uma unidade lexical especializada, a manifestação linguística do conceito, e o seu estudo deve ter lugar dentro de situações comunicativas. É no ambiente discursivo que o terminólogo procede à sua observação e análise. O valor de um conceito vai sendo construído e circunscrito através da análise terminológica. A terminologia que parte da perspectiva semasiológica estuda o conhecimento na sua relação com a verbalização. A esta metodologia não é alheio o facto de termos vindo a dispor há algumas décadas de acesso facilitado a corpora electrónicos, passíveis de serem

trabalhados através de ferramentas de processamento de língua natural.

Por seu turno, a perspectiva onomasiológica centra-se no conceito. Baseada nos pressupostos de Wüster, o conceito é uma unidade de conhecimento extralinguística que precede a forma linguística. Os conceitos existem independentemente dos termos e de qualquer língua natural. O objectivo era essencialmente alcançar uma comunicação isenta de ambiguidade, onde se previa uma relação biunívoca entre conceito e termo. Podemos estabelecer aqui uma aproximação às orientações actuais da representação do conhecimento. Tendo como base os princípios aristotélicos, o conceito é definido pelo conjunto das suas diferenças e pelas relações que estabelece com outros conceitos. A relação de um para um entre termo e conceito torna-se fundamental para os sistemas informáticos, permitindo a representação de conhecimento estático, passível por isso de formalização.

O ponto de partida linguístico tem como pressuposto que é possível construir ontologias a partir de texto através de ferramentas de processamento automático da língua natural, com análise lexical, semântica e estatística. A 'aprendizagem de ontologias' teria como elemento catalisador o texto de especialidade, contendo, à partida, conhecimento organizado, onde um conjunto de termos corresponderia a um conjunto de conceitos, reflectindo o conhecimento de domínio. Porém, ao contrário do que à partida se poderia supor, no texto científico grande parte do conhecimento base não se encontra explícito, está implícito, com alterações permanentes na associação e instanciação dos conceitos, observando-se uma reconfiguração permanente do nosso conhecimento periférico, dado que o conhecimento central é implicitamente partilhado.

O ponto de partida extralinguístico baseia-se numa premissa contrária : não é possível construir ontologias directamente a partir de texto. O texto, pela sua natureza, é dinâmico, e o que se pretende é capturar conhecimento estático, em formato armazenável. Para alcançarmos esse objectivo, devemos basear-nos no

conhecimento transmitido directamente pelo especialista de domínio, onde o raciocínio prevalece sobre o discurso. O texto não contém conceitos, por isso não há relação directa entre texto e ontologia. A conceptualização só é clara se for descontextualizada, sendo necessário um distanciamento em relação à língua natural. Acontece que grande parte dos conceitos e relações entre conceitos são unicamente intelectualizáveis e verbalizáveis por via de enunciados linguísticos.

Decidimos testar as metodologias onomasiológica e semasiológica a um domínio de conhecimento para avaliar o seu grau de eficiência na representação de conceptualizações informais. Construámos um mapa conceptual e uma rede lexical sobre o mesmo domínio, com o objectivo de avaliar se devemos ou não optar por uma única abordagem terminológica. Um mapa conceptual não é uma ontologia, sendo que as relações entre conceitos não estão formalizadas, por isso o mapa conceptual que foi construído foi resultado de um trabalho directo com os especialistas de domínio, e resultado também de conhecimento de domínio que foi sendo adquirido pelo terminólogo, sem recurso a ferramentas de PLN.

Após a conclusão do mapa conceptual, analisámos o mesmo domínio de especialidade através de uma metodologia semasiológica, extraindo dados de texto com recurso a ferramentas de processamento de língua natural. Nesta altura, já éramos detentores de conhecimento de domínio, facilitando a tarefa de selecção de um corpus. Para que pudéssemos comparar resultados, optámos por utilizar a mesma representação gráfica, com a mesma ferramenta. A análise de concordâncias, de listas de candidatos a termo e de outros dados de natureza linguística permitiu seleccionar algumas sequências de texto centradas num candidato a termo seleccionado. Foi em torno dele que se organizou a rede lexical. A nossa metodologia baseou-se fundamentalmente na análise dos processos de reformulação presentes em discurso que podem fornecer dados de natureza linguística e, potencialmente, extralinguística e cognitiva. Em primeiro lugar, detectou-se e sinalizou-se o marcador linguístico, atribuindo-lhe uma relação semântica de tipo hierárquico ou não hierárquico. Posteriormente, construíram-se

redes lexicais com uma apresentação gráfica correspondente. Cada rede lexical correspondente a cada sequência de texto sofreu por isso um processo cognitivo de transposição entre os dados de natureza linguística e os dados de natureza conceptual.

Foram observadas muitas intersecções entre os termos extraídos da rede lexical e os conceitos extraídos do mapa conceptual, ou seja, entre a conceptualização de determinado processo e a verbalização do mesmo processo, embora, para fins comunicativos, recorram os dois à língua natural. A análise e observação do corpus pode fornecer dados importantes para a conceptualização do domínio, embora não deva sofrer processos automáticos de transposição.

Observou-se também que a ordem das metodologias não é arbitrária. A utilização da abordagem onomasiológica em primeiro lugar influenciou positivamente e facilitou em grande medida a implementação da abordagem semasiológica. De facto, quando confrontado com a aplicação e os resultados das duas metodologias, o especialista não teve dúvidas em escolher a metodologia onomasiológica. Contudo, embora o especialista privilegie a utilização da perspectiva onomasiológica, e apesar desta se basear fundamentalmente em dados extralinguísticos, sente sempre a necessidade de mediação da língua natural na especificação informal da conceptualização.

Por sua vez, a implementação da perspectiva linguística é enriquecida pelo conhecimento prévio do domínio de especialidade, facilitando a representação do conhecimento explícito e facilitando a inferência do conhecimento implícito. Mas salienta-se que : 1. a construção autónoma por parte do terminólogo da rede lexical só foi possível porque já tinha adquirido conhecimento de domínio ; 2. deparámo-nos com frequência com conhecimento implícito e omissos. Mas conseguimos inferir termos e relações entre termos, que foram propostas e adicionadas dado serem necessárias para a construção da rede lexical. Foram também suprimidas algumas sequências de texto. 3. Como em todos os textos de especialidade foram

observadas algumas imprecisões, decorrendo, muitas vezes, da tentativa de redigir de forma esteticamente agradável, que não comprometem a sua validade dentro da comunidade. NO entanto, podem existir situações em que as imprecisões discursivas em termos de marcas gráficas de pontuação foram decisivas para a incorrecta interpretação do conteúdo conceptual, só descodificado aquando da validação pelo especialista. Isso demonstra que nem sempre as estruturas semânticas encerram estruturas conceptuais.

Apesar destas observações, foi possível concluir que através da análise de sequências de texto podemos aceder ao processo de construção de conhecimento em discurso, indo do lexical para o conceptual, onde os marcadores linguísticos e as relações semânticas permitem clarificar a conceptualização, activando simultaneamente a dimensão linguística e conceptual do termo. Não existe uma metodologia correcta para construção de ontologias, nem tão pouco uma ontologia correcta. Depende em grande parte dos objectivos do que se pretende construir, do factor tempo, e claro, do factor custo e do factor humano. A formalização das representações obriga a que as especificações das conceptualizações sejam feitas através de uma linguagem, não da língua, resultando daí um conjunto de restrições. Por isso as ontologias não são extraíveis, dado que no texto não temos conceitos, e por outro lado, existe uma forte presença do conhecimento implícito. Não existe uma relação de um para um entre texto e ontologia nem entre especialista e ontologia. A construção de ontologias não pode nem deve subtrair-se à questão linguística, sendo o contrário igualmente verdadeiro : o lexical não deve substituir o conceptual. Assim, na construção de ontologias não existe exclusividade, mas complementaridade, entre as abordagens que partem do linguístico e do extralinguístico.

Na sociedade tecnológica em que vivemos seria importante reequacionar o estatuto do corpus, do terminólogo, do especialista e do engenheiro do conhecimento. Não existe conhecimento no texto. Não existem conceitos no texto. No texto temos manifestações linguísticas que podem apontar para uma eventual

organização conceptual de um domínio. O discurso verbal é uma plataforma de análise. Um texto por si só não fornece dados que viabilizem a construção automática de ontologias. O que obtemos são redes lexicais, pré-estágios semânticos de redes conceptuais. O texto constitui-se como um elemento operacional. O especialista é fundamental em todas as fases de construção da ontologia, uma vez que é ele o detentor do conhecimento : do conjunto de conceitos e das suas relações.

O recurso à língua natural é inevitável. É através dela que manifestamos verbalmente as definições dos conceitos por via dos termos. O acesso ao discurso do especialista é fundamental. Mas o processamento de corpus pode não ser necessário. Por isso o terminólogo deve saber posicionar-se neste novo território. Deve dominar as ferramentas de processamento de texto, deve aprender a interagir com o especialista, e deve também adquirir competências na área computacional para conseguir trabalhar eficazmente com o engenheiro do conhecimento. A metodologia testada e utilizada por nós evidencia benefícios claros para o terminólogo, uma vez que lhe proporciona uma aproximação muito mais eficaz às necessidades do especialista e do engenheiro. Não há uma diminuição das funções. Existe antes um reforço à luz das necessidades actuais. A metodologia deve servir os objectivos do grupo profissional a quem se dirige. Somos linguistas, e damos valor aos termos. Mas devemos considerar a abordagem ao conceito como prioritária na construção de ontologias.

Referências bibliográficas :

- Bachimont, B. (2006, 25 Novembro 2008). *Qu'est-ce c'est une ontologie ?* , from http://www.technolangue.net/imprimer.php3?id_article=280
- 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*

- BBrewster, C., & O'Hara, K. (2004). *Knowledge Representation with Ontologies : The Present and Future*. IEEE Intelligent Systems, 19(1), 72-81.
- 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., 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.
- Candel, D. (2004, 2004). *Wüster par lui-meme*. Cahier du CIEL, Des fondements théoriques de la terminologie, 15-31.
- Clancey, W. J. (2007). The knowledge level reinterpreted : Modeling socio-technical systems. *International Journal of Intelligent Systems*, 8(1), 33-49.
- Conceição, M. C. (2005). *Concepts, Termes et Reformulations* : Presses Universitaires de Lyon. Cornejo, M. (2003). *Unity, Value and Knowledge Communities*.
- 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. (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>
- 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.
- Gillam, L., Tariq, M., & Ahmad, K. (2005). *Terminology and the construction of ontology*. *Terminology*, 11, 55-81.

- 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.
- Guarino, N. (1997). *Understanding, building and using ontologies*. International Journal of Human Computer Studies, 46, 293-310.
- 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).
- 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.
- 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. (1982). *The Knowledge Level*. *Artificial Intelligence*, 87-127.
- 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. (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.
- Rastier, F. (2004). *Ontologie(s)*. *Revue des sciences et technologies de l'information - Révue d'intelligence artificielle*, 18, 15-40.
- 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. (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. (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.
 - Sintek, M., Buitelaar, P., & Olejnik, D. (2004). *A formalization of ontology learning from text*. Paper presented at the Workshop on Evaluation of Ontology-based Tools Germany.
 - 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.
- 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.