

LOCALIZAÇÃO DE ONTOLOGIAS – ELEMENTOS PARA A CONSTRUÇÃO DE UMA ABORDAGEM

Manuel Moreira da Silva
Instituto Superior de Contabilidade e Administração do Porto
ISCAP
Portugal
mdasilva@iscap.ipp.pt

Resumo:

O desenvolvimento de recursos multilingues robustos para fazer face às exigências crescentes na complexidade dos processos intra e inter-organizacionais é um processo complexo que obriga a um aumento da qualidade nos modos de interação e partilha dos recursos das organizações, através, por exemplo, de um maior envolvimento dos diferentes interlocutores em formas eficazes e inovadoras de colaboração. É um processo em que se identificam vários problemas e dificuldades, como sendo, no caso da criação de bases de dados lexicais multilingues, o desenvolvimento de uma arquitectura capaz de dar resposta a um conjunto vasto de questões linguísticas, como a polissemia, os padrões lexicais ou os equivalentes de tradução. Estas questões colocam-se na construção quer dos recursos terminológicos, quer de ontologias multilingues. No caso da construção de uma ontologia em diferentes línguas, processo no qual focalizaremos a nossa atenção, as questões e a complexidade aumentam, dado o tipo e propósitos do artefacto semântico, os elementos a localizar (conceitos e relações conceptuais) e o contexto em que o processo de localização ocorre. Pretendemos, assim, com este artigo, analisar o conceito e o processo de localização no contexto dos sistemas de gestão do conhecimento baseados em ontologias, tendo em atenção o papel central

da terminologia no processo de localização, as diferentes abordagens e modelos propostos, bem como as ferramentas de base linguística que apoiam a implementação do processo. Procuraremos, finalmente, estabelecer alguns paralelismos entre o processo tradicional de localização e o processo de localização de ontologias, para melhor o situar e definir.

Abstract:

The development of robust multilingual resources to meet the growing demands of intra and inter-organizational complexity is a complex process that demands an increase of quality in the organizations' resource interaction and sharing modes, trough, for example, stronger involvement of different stakeholders in effective and innovative collaboration ways. It is a process where several problems and difficulties arise, such as, in the creation of multilingual lexical databases, the development of an architecture capable of addressing a wide range of language issues such as polysemy, lexicalization patterns or equivalence in translation. These questions appear also in the construction of terminological resources or multilingual ontologies. In the construction of a multilingual ontology, the questions and their complexity increases, given the nature and purposes of the semantic artifact, of the elements to localize (concepts and conceptual relations) and the context in which the localization process occurs. We intend in this article to analyze the concept and the localization process in the context of ontology-based knowledge management systems, taking into account the central role of terminology in the localization process, the different approaches and models proposed and the existing linguistic tools that support the implementation process. Finally we will try to establish some parallels between the traditional localization process and the ontologies localization process to better contextualize and define it.

Palavras-chave: Localização, ontologias, terminologia, gestão do conhecimento, multilinguismo

Keywords: localization, ontologies, terminology, knowledge management, multilingualism

Introdução

Num modelo de contextualismo universal, o diálogo intercultural que acontece através da comunicação multilingue é, para Cortese (2007: 203), um pré-requisito fundamental para a *self-reflexivity, thecritical self-awarenessof local identitieswhichprovidesthecornerstone for overcomingparticularizedbehaviorsand for thediffusionofequity in socioeconomicrelations*. Este diálogo tornou-se ainda mais premente com o advento da globalização, fenómeno que, apoiado pela rápida propagação da Internet, aumentou a necessidade das organizações de acederem, gerirem e disseminarem a informação num número crescente de línguas. Como descreve Bergamashi (2003)¹ o Mundo está a encolher à medida que o avanço da comunicação e da computação se desenrolam a um ritmo impressionante. No entanto, e numa afirmação na qual nos revemos, a mesma autora deixa claro não haver qualquer dúvida sobre o facto de que as pessoas continuarão ligadas às suas culturas e língua de origem, dado que as:

Individual languages and their peculiarities in expressing information and concepts are a treasure, a cultural knowledge base that must not be underestimated or, even worse, ignored as pressures to be able to use common languages increase.

¹ Sewasie Project – D 2.1 – Semantic enrichment

A globalização, entendida como o processo de desenvolvimento de um produto que pode ser usado com sucesso em diferentes contextos culturais sem sofrer modificações (Hoft, 1995), tornou uma realidade a necessidade de criar produtos e serviços transaccionáveis na língua dos mercados a que se destinam, normalmente designados de locais. O processo de globalização ocorreu não só ao nível das trocas comerciais, mas também ao nível das trocas científicas, técnicas e do desenvolvimento de redes de comunicação internacionais, o que resultou na necessidade de criação e uso de recursos terminológicos para apoiar este processo e a sua expansão.

No entanto, a comunicação multilingue em termos globais não avançou ao mesmo ritmo que as trocas comerciais. Tal acontece devido às dificuldades inerentes à criação e gestão de conteúdos multilingues. Tal como identificavam Hovy *et al.* (1999), o objectivo dos diferentes ramos que se dedicam ao Processamento da Linguagem Natural permanece

(...) tantalizingly out of reach. The principal difficulty lies in dealing with meaning. However well systems perform their basic steps, they are still not able to perform at high enough levels for real-world domains, because they are unable to sufficiently understand what the user is trying to say or do.

Outros autores, como Benjamins *et al.* (2002), definem mesmo o multilinguismo como um dos seis desafios à realização da Web Semântica, sendo necessário ter em conta, segundo os autores, que:

1. At the ontology level, ontology builders may want to use their native language for the development of the ontologies in which annotations will be based.

2. At the annotation level, annotation of content can be performed in various languages.
3. Finally, at the user interface level, millions of people would like to access relevant content in their native language irrespective of the source language in which annotations are presented.

Para responder a estes desafios surgiram, na última década, novos paradigmas com origem na análise dos ambientes em linha e centrados na comunicação, colaboração e co-criação de recursos linguísticos. No entanto, um conjunto significativo de barreiras linguísticas bem identificadas permanece ainda sem resolução, sobretudo quando o nosso enfoque recai sobre a comunicação multilingue em domínios específicos do saber. Uma dessas lacunas é assinalada por Minako O'Hagan (2002:7), que refere que uma das fontes mais significativas de falhas na comunicação interlinguística pode provir de *incorrect assumptions of common beliefs and experience that actually differ according to cultural background, knowledge, preferences and pragmatics*. Estas barreiras poderão ser mais facilmente transpostas se, como apontam Cencioni e Rossi (2008), se promover uma nova abordagem, multidisciplinar, com recurso a modelos de conhecimento do mundo inspirados nas redes sociais, cujos princípios interessa compreender e integrar.

De facto, a interacção entre indivíduos que prosseguem objectivos comuns, sobretudo quando estamos perante organizações multinacionais e multiculturais, tende a realizar-se, cada vez mais, em ambientes colaborativos e cooperativos disponíveis em linha, i.e., em espaços virtuais mediados pela internet, que podem ser vistos como espaços de comunicação computacional, ainda que altamente humanizados, que não podem ser separados do contexto, da experiência, da cultura e da língua e do potencial criativo e definatório desta.

O estabelecimento de espaços de informação multinacionais e multilingues, que se caracterizam pela sua heterogeneidade e multiculturalidade, criam a necessidade de existência de novos tipos de sistemas de organização do

conhecimento que facilitem e tornem mais eficiente o acesso aos recursos de informação existentes e a partilha de conhecimento. O processo de desenvolvimento tecnológico, por um lado, e o aumento da cooperação a nível nacional e internacional levam a que recursos de informação estruturados de modo heterogéneo possam e devam ser reutilizados, integrados e disponibilizados de modo a facilitar o acesso à informação em diferentes línguas.

Nestes ambientes, a língua, meio por excelência da comunicação humana, e a terminologia, enquanto elemento de especificação do conhecimento, representam um papel de mediação cada vez mais importante na comunicação homem-máquina e na comunicação máquina-máquina. Percepcionadas como fenómenos complexos, que extravasam a ideia de simples meios de transmissão da informação, surgem como o elo fundamental para a descoberta e a criação do conhecimento, bem como para a sua gestão e disseminação.

A análise da informação existente em cada organização e da informação específica disseminada no seu contexto de acção, de forma a transformá-las em conhecimento, torna-se essencial ao desenvolvimento de novos processos de gestão, projectos, serviços e produtos.

Para fazer face a estas exigências crescentes na complexidade dos processos intra e inter-organizacionais, tem-se vindo a verificar, por um lado, um aumento de qualidade nos modos de interacção e partilha dos recursos das organizações, através de um maior envolvimento dos diferentes interlocutores em formas eficazes e inovadoras de colaboração, como é o caso das redes virtuais de cooperação empresarial e, por outro, ao desenvolvimento de uma variedade de recursos multilingues.

O desenvolvimento destes recursos é um processo complexo, em que se identificam vários problemas e dificuldades, como sendo, por exemplo, no caso da criação de bases de dados lexicais multilingues, o desenvolvimento de uma arquitectura capaz de dar resposta a um conjunto vasto de questões linguísticas, como a polissemia, os padrões lexicais ou os equivalentes de tradução.

Estas questões colocam-se na construção quer dos recursos terminológicos quer de ontologias multilingues. No caso da construção de uma ontologia multilingue, as questões e a complexidade aumentam, dado o tipo e propósitos do artefacto semântico, os elementos a localizar (conceitos e relações conceptuais) e o contexto em que o processo de localização ocorre. Tendo por base esta percepção, analisaremos em seguida o conceito e o processo de localização no contexto dos sistemas de gestão do conhecimento baseados em ontologias, tendo em atenção o papel central da terminologia no processo de localização, bem como as diferentes abordagens e modelos propostos, procurando assim situar e definir o processo de localização de ontologias.

Localização – conceito em evolução

A indústria da localização, emergida na década de 80, sofreu um avanço intenso na década seguinte, fomentado pelo mercado dos computadores pessoais e pela necessidade de localizar o grande número de aplicações a que aqueles recorriam para o seu funcionamento. Desde então, o paradigma da localização tem estado a amadurecer e a abranger novos domínios, sobretudo graças ao avanço técnico e tecnológico promovido e procurado pela indústria da língua, à medida que as práticas de localização vão evoluindo em resposta às necessidades de globalização de uma variedade de conteúdos digitais. No entanto, esta dinâmica não tem encontrado grande paralelismo na análise e estudo do fenómeno junto da academia, apesar da sua importância e das perspectivas que abre sobre novas dimensões e caminhos para o futuro do fenómeno da tradução. (O'Hagan, 2004:02)

Com a evolução deste sector, foram-se definindo um conjunto de conceitos nucleares, reunidos em torno de um novo acrónimo – GILT - que incorpora a Globalização, a Internacionalização, a Localização e a Tradução, acrónimo que reflecte a complexidade inerente à preparação de um produto ou conteúdo para o mercado global. Segundo O'Hagan (2004:02) GILT, para além de destacar os

processos específicos necessários ao tratamento de conteúdo electrónico, coloca a transferência dos conteúdos linguísticos na *wider picture of globalization*. Interessa, por isso, explicitar melhor estes conceitos, as suas interligações e o modo como se reflectem nos processos de localização e de tradução e de transferência de informação e conhecimento entre culturas.

A IBM define, no seu Glossário sobre internacionalização, o conceito de *globalização* como sendo o processo de *developing, manufacturing, and marketing software products that are intended for world wide distribution*. Este conceito combina dois aspectos do trabalho: a internacionalização (criar um produto que possa ser usado livre de barreiras linguísticas e culturais) e a localização (tradução e adaptação de um produto a um *locale* específico)². A globalização pode, assim, ser descrita como sendo o processo de criação de um produtopassível de ser usado com sucesso em diversos contextos culturais sem modificação. Autores como Esselink (2000) ligam ainda esta perspectiva à dos processos e estratégias de negócio associados à criação de um produto global.

A internacionalização surge associada ao processo de globalização e é descrita como o processo de reengenharia de um produto de informação de modo a que possa ser *easily localized for export to any country in the world. An internationalized product consists of two components: core information and international variables*. Hoft (1995). Esselink (2000) descreve este como sendo o processo de generalização de um produto, de forma a que este possa lidar com “*multiple languages and cultural conventions without the need for re-design*”. Corresponde, assim, a um processo de implementação de um produto de modo a que este não seja concebido especificamente para um local, i.e. que possa suportar diferentes línguas e convenções, independentemente da sua origem, sem modificar qualquer componente do seu design inicial.

² <http://www.ibm.com/developerworks/library/glossaries/unicode.html>

Este processo antecede o da localização. Na origem do conceito de localização está o termo conhecido em Inglês por *locale*. Este termo estendeu o seu significado a partir da aceção da língua geral do substantivo *locale*. Quando se fala em *locale*, no contexto de um sistema operativo ou de uma aplicação informática, faz-se referência a uma área geopolítica e a uma língua determinadas, bem como a todas as particularidades que é preciso ter em consideração – no caso de uma aplicação, o modo de expressar as horas e as datas, as letras do seu alfabeto, etc. (Pagans, 2000). Autores como Texin (2004) e Guidère (2003) vêem um *locale* como um grupo de preferências culturais. Um *locale*, segundo Texin (2003:26), é composto por e pode ser descrito tendo em conta um conjunto de características:

- world – representing the cultural preferences of the users in the whole world;
- distinctive – permitting differentiation of the various groups of users;
- correlative – connecting a culture to a locale;
- determinative – correctly describing the types of preference linked to each locale;
- specific – defining accurately the specific preferences of all locales.

A localização, elemento nuclear e essencial aos dois processos anteriormente descritos, é definida como um processo de adaptação linguística e cultural de informação digital para mercados locais (Esselink, 2000; Dunne, 2006). O facto de estes autores se referirem a informação e a conteúdos digitais, afasta-os da perspectiva mais restritiva de Hoft (1995), supracitada, para a qual a localização se dedicava ao processamento de *produtos de informação*, enquadrando estes, quase só, as aplicações típicas de software. A localização pode, então, ser caracterizada como um processo de adaptação (1) de carácter linguístico, com recurso à tradução, à terminologia e à edição, e (2) de carácter não-linguístico, pelos processos de programação e ergonomia envolvidos, tendo como objectivo último o

desenvolvimento de um produto aceite por uma determinada comunidade linguística e cultural.

Este conceito tem vindo a evoluir à medida que a própria actividade e a indústria evoluem e Schäler (2007:157) aponta agora uma nova dimensão, na qual nos revemos. Para este autor, a localização representa não só a adaptação linguística e cultural de conteúdos digitais aos requisitos de um mercado internacional, mas também a *provision of services and technologies for the management of multilingualism across the digital global information flow*. A localização necessita, assim, de ter em consideração a comunicação intercultural e factores como o estatuto de uma língua, a política linguística, as preferências culturais, os usos estabelecidos e, até mesmo, as necessidades dos utilizadores, centrais ao processo de aceitação e uso e sucesso do produto.

Terminologia e localização

As terminologias são conjuntos de termos específicos a um domínio, estruturados, inter-relacionados, que se constituem como elementos de organização e designação do conhecimento, funcionando como pré-requisitos multifuncionais para e como factores de comunicação intercultural (Budin, 2005). A terminologia constitui um elemento crucial sempre que a informação e o conhecimento relativo a um domínio de especialidade são:

- gerados (ex.: na pesquisa e desenvolvimento);
- utilizados (ex.: em textos de especialidade);
- armazenados e processados (ex.: em bases de dados);
- transmitidos (ensino/formação);
- implementados (ex.: na transferência de tecnologia e conhecimento);
- traduzidos, localizados ou interpretados.

Daqui decorre, também, a sua importância para o desenvolvimento do processo de localização e para a garantia da sua qualidade e consistência, evidenciada por diferentes autores, como Schmitz (2007), para quem a terminologia constitui o elemento primário de transferência de conhecimento entre o produtor e o utilizador final ou, Quirion (2003: 548), que vê a terminologia como o elemento nuclear da localização, dado que a terminologia relacionada com o produto tem que ser estabelecida tendo em conta os:

"linguistic usages of the local specialists" (horizontal axis) and ensuring "rigorous terminological consistency between all the versions [of the] product and with related articles" (vertical axis)."

Gouadec (2003: 528), por seu lado, associa o processo de gestão da terminologia ao da internacionalização, e afirma que, idealmente, os produtos³ deveriam ser internacionalizados desde o momento da sua concepção de modo a facilitar o trabalho terminológico. Esta perspectiva aproxima o processo de localização do terminológico, dado que uma terminologia procura, através de uma análise sistemática, uma estruturação e representação completa de uma área de saber, e a localização deve, à partida, procurar aglutinar todos os elementos respeitantes ao produto que se enquadrem na *scope of cultural, linguistic, technical, religious, philosophical, value system, argumentation and presentation method or other specificities*. (Gouadec, 2003: 528).

Esta proximidade é reforçada pela existência de um conjunto de recursos, como bases de dados terminológicas, dicionários técnicos, etc., que apoiam o localizador no seu processo de consulta e decisão. Finalmente, a compilação de recursos lexicais codificados a partir de corpus, sobretudo a extraída de corpora paralelos, resulta em dados que contribuem para um acesso a informação mais rica

³Uma ontologia pode, por analogia, ser entendida como um produto.

e, normalmente, de carácter mais actual. Da conjugação destes elementos resulta uma relação simbiótica entre o estudo dos termos e das terminologias e o estudo da localização que coloca a terminologia no centro do processo de localização, sendo aquela descrita como o ADN da globalização e da localização⁴ (Warburton, 2010).

De facto, para o bom desenvolvimento, disseminação e uso de um produto, a terminologia tem que ser estabelecida tendo em conta os usos linguísticos dos especialistas (comunidades) locais e a necessidade de assegurar consistência entre todas as versões (de um conteúdo digital). A boa gestão terminológica pode, assim, ser vista como um contributo essencial para a boa gestão do processo de localização, constituindo-se o trabalho terminológico como um elemento nuclear do processo de localização⁵.

Nas palavras de Karsch (2009:144), cenários mais ou menos complexos de localização estão profundamente dependentes de uma boa gestão da terminologia. Segundo a autora, de modo a maximizar o retorno no investimento,

systematic terminology management must be based on sound concept analysis and can have as a byproduct explicit and formalized ontologies. The prerequisites for success of such a terminology and potentially ontology system are not easily satisfied – processes, tools and people must come together in a highly complex interplay.

A localização não é, assim, apenas uma questão de língua, mas também de conhecimentos. Preocupa-se também com o conhecimento sobre o discurso (termos complexos, colocações terminológicas) e permite, se bem executada, gerir a qualidade do discurso e, por consequência, a qualidade da comunicação multilingue e, logo, da gestão e transferência da informação e do conhecimento,

⁴<http://www.lisa.org/Touchpoint-Advisory-Webinars.586.0.html#c1087>

⁵Como explicita Karsch (2009:123), *Standardized terminology is the backbone of a quality-driven translation process.*

nomeadamente enquanto etapa integrante da gestão do conhecimento baseado em ontologias.

Uma ontologia, definida por Gruber (1993) como a especificação de uma conceptualização de um domínio, recorre, tal como os restantes sistemas de representação do conhecimento, aos *“thoughts of reference”, also known as concepts, to refer to the real world* (Montiel-Ponsoda, 2008). Ao comparar ontologias e sistemas linguísticos, a autora estabelece um paralelismo e afirma que ambos possuem três componentes principais: *signs or symbols used to designate concepts or thoughts of mind, which refer to phenomena in the real world*, componentes que, na linguística, são normalmente representados pelo triângulo de Ogden and Richards.

Já no que se refere à composição de uma ontologia, esta é normalmente composta por 6 camadas distintas, nomeadamente:

- Lexical layer: characters and symbols that make up the syntax
- Syntactic layer: structure of characters and symbols, i.e., the grammar. It embraces different representation languages (e.g. RDF(S), OWL, etc.)
- Representation paradigm layer: paradigm followed in the representation of the ontology (frames, semantic networks, DL, etc.) that allows a certain way of expressing and structuring knowledge
- Terminological layer: terms or labels selected to name ontology elements
- Conceptual layer: related to conceptualization decisions, such as granularity, expressiveness, perspective, etc.
- Pragmatic layer: final layout of the model according to user’s needs (Montiel-Ponsoda, 2008)

Partindo deste ponto de vista, a autora afirma que as camadas terminológica, conceptual e pragmática são as directamente ligadas ao processo de localização. No que respeita à camada conceptual, algumas ontologias podem necessitar de adaptar a sua estrutura conceptual “in order to fit in the thoughts of reference of a specific linguistic and cultural community” (*idem, ibidem*). No que se refere à camada pragmática, o modo como a informação multilingue é apresentada ao utilizador pode influenciar a aceitação e uso do modelo pelo utilizador final.

Já no que respeita à terminologia, a autora reconhece o papel decisivo que esta desempenha no processo de localização, uma vez que está relacionada com as designações atribuídas aos diferentes elementos da ontologia e é o veículo que permite que os termos da ontologia possam ser expressas em mais do que uma língua natural. É, assim, um elemento essencial, uma vez que contribui, de modo decisivo para a análise dos conceitos do domínio (e dos termos que os designam), contribuindo, através do recurso a um conjunto de métodos, para a recolha dos recursos lexicais, terminológicos e de representação conceptual, para assegurar a correcção, consistência e desambiguação no processo de escolha dos termos e permitindo o acesso à informação de especialidade e às comunidades de especialistas, o que contribui para potenciar o processo de conceptualização na comunidade multilingue.

No contexto de trabalho bilingue e multilingue em que se insere a localização, e tendo em conta a disponibilidade e o volume sempre crescente de corpora e outros recursos linguísticos e conceptuais, muitos dos quais em formato multilingue, a terminologia desempenha ainda um papel importante ao assistir os terminólogos, tradutores e especialistas – a comunidade – nas tarefas de aquisição de conhecimento, descrição conceptual, criação do sistema de conceitos, formulação de definições e, finalmente, estabelecimento de equivalentes entre os termos, para as diferentes línguas de trabalho.

Localização de ontologias

As dimensões do trabalho e dos processos envolvidos na localização, bem como a natureza dos conteúdos a localizar, têm vindo a evoluir, como ficou dito, à medida que surgem novas dimensões e novos desafios, resultantes, em grande parte, da rápida evolução da internet. A localização, apesar de ser um processo bem desenvolvido, cujas metodologias são utilizadas com sucesso pela indústria da língua no desenvolvimento e adequação de conteúdos multilingues, não foi ainda devidamente explorada como elemento de apoio à construção de ontologias multilingues.

Os sistemas de gestão do conhecimento baseados em ontologias permitem a pesquisa, a partilha e a reutilização de fontes de dados estruturados na Web semântica. A urgência em aceder ao conhecimento, ultrapassando as barreiras impostas pelas diferentes línguas naturais, tornou-se uma necessidade premente para a Web semântica o que resultou no surgimento das primeiras ontologias multilingues.

Partindo de uma observação da literatura existente, apercebemo-nos que a evolução da preocupação com os fenómenos multilingues e a sua integração no desenvolvimento de ontologias esteve entre as primeiras dificuldades detectadas para a correcta evolução da engenharia de ontologias. No entanto, a integração e o enriquecimento das ontologias com elementos multilingues foi sendo investigada de modo menos sistemático, dada a dificuldade que representa. Só mais recentemente se notou uma evolução nos sistemas técnicos e abordagens teóricas disponíveis, bem como na evolução de alguns princípios epistemológicos que possibilitaram a abertura de novos campos de reflexão e análise nesta área.

Tal como explicita Budin (2005:114), o objectivo mais genérico do desenvolvimento de processos de localização de ontologias é permitir a interoperabilidade semântica interlinguística em ambientes de informação de grande volume que contém, normalmente, um grande número de recursos de

conhecimento heterogéneos e distribuídos. A especificação de uma abordagem que potencie a interoperabilidade semântica entre estes recursos requer, segundo o autor, uma modelação detalhada das situações, *a description Framework and a methodology that allows us to focus on capturing semantic complexity, richness, and diversity in real-world resources.*

Uma vez que o conhecimento e as representações do conhecimento não se restringem ao uso de uma língua natural em particular, o recurso ao multilinguismo tornou-se, assim, uma necessidade evidente na construção de ontologias. A construção de ontologias sofre desta lacuna, que urge ultrapassar. De facto, quando acedemos a repositórios de ontologias como o DAML⁶ ou o Open OntologyRepository⁷, verificamos que apenas um número diminuto é bilingue ou multilingue, sendo grande parte desenvolvidas e implementadas apenas em Inglês. Este facto resulta, para os utilizadores com um domínio menos avançado da língua inglesa, em barreiras linguísticas, por vezes intransponíveis, e em dificuldades no acesso ao e uso do conhecimento ontológico. Este cenário agrava-se quando a pesquisa é dirigida a ontologias em língua portuguesa ou que a contenham. Daqui resulta uma dificuldade acrescida para os utilizadores, as organizações e redes internacionais que lidam diariamente com informação multilingue, onde se percebe uma cada vez maior necessidade de representar o multilinguismo nas ontologias que operam ou venham a operar nos seus sistemas de gestão da informação.

Dimensões e dificuldades

A língua de ontologias formalizadas em mais do que uma língua natural provém de vários factores, a começar pela dificuldade inerente à escolha das metodologias a adoptar para o processo de conceptualização e representação do

⁶<http://www.ai.sri.com/daml/ontologies/>

⁷<http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository>

conhecimento num ambiente de desenvolvimento e localização de ontologias multilingues e pela falta de ferramentas específicas que apoiem este processo.

Outras dificuldades mais específicas advêm, sobretudo, de problemas de ordem linguística, que surgem, por exemplo, na associação dos significados de termos de diferentes línguas a conceitos de ontologias, uma vez não se poder dizer que os significados dos conceitos coincidam, dado que os primeiros estão estreitamente relacionados com uma visão particular de uma língua e de uma cultura, ao passo que os conceitos de uma ontologia procuram capturar objectos do mundo real, sendo definidos e organizados de acordo com critérios acordados por consenso entre especialistas (Peters *et al.*, 2008: 69).

Assumindo que as ontologias descrevem conceitos e não o modo como estes conceitos são expressos em palavras numa língua natural, Weigand (2008) considera natural que se assuma que uma ontologia é independente da língua em que é criada. No entanto, aponta problemas a esta assunção, a começar pelo facto de que *there is the philosophical point that it is not possible to step outside our linguistic make-up. Concepts are shaped in the communication between members of a linguistic community*. Esta perspectiva pode ser ilustrada por inúmeros exemplos, segundo o autor, uma vez que existem conceitos que numa língua e não em outras (dando origem à neologia), e línguas que classificam de modo distinto o léxico, e, por consequência, os conceitos.

Para além destas dificuldades, o processo localização de uma ontologia, artefacto específico para a representação do conhecimento de um domínio e construído num contexto e com um propósito particular, coloca várias questões, de que se destacam as relacionadas com:

1. a definição e delimitação do domínio ou subdomínio(s) a conceptualizar;
2. a tipologia, classificação e conversão de recursos semânticos a adaptar para tornar o processo célere e enriquecer a ontologia;

3. a possível simultaneidade entre o processo de conceptualização e de localização e o desenvolvimento de metodologias necessárias à sua integração;
4. a limitação temporal normalmente imposta aos processos de conceptualização e de localização;
5. o modo de integração e (re)utilização de recursos e ferramentas linguísticas já disponíveis.

No que respeita a este último ponto, partilhamos, da perspectiva de Lenci (2008), que defende a existência de uma ligação bidireccional entre ferramentas de Processamento da Língua Natural (PLN) e recursos linguísticos. Segundo o autor, as ferramentas e aplicações de PLN são também criadoras de conhecimento, i.e., podem ser usadas para criar e modificar recursos⁸. Para Lenci, apesar das suas diferenças *prima facie*, os dois papéis dos sistemas de PLN como utilizadores e criadores de conhecimento têm que ser percepcionados como uma interrelação, representando, cada um, um dos lados de uma mesma moeda (Lenci, 2008:2).

Por outro lado, e tal como explicitam Fu *et al.* (2008: 02), num cenário de integração de duas ontologias de línguas diferentes, uma abordagem possível passa por traduzir uma delas para a língua da outra ontologia recorrendo, por exemplo, a técnicas de tradução automática, num momento anterior à aplicação de técnicas de mapeamento entre ontologias monolíngues. *In such a multilingual ontology mapping approach, challenges are mainly found in the ontology translation phase and the monolingual ontology matching phase.*

Mas, como destacam os autores, a identificação dos resultados mais apropriados da tradução é crucial para a fase de localização das ontologias, dependendo a qualidade destas, em grande medida, das escolhas feitas nesta fase. Esta perspectiva reforça a necessidade, já expressa, de produção de aplicações

⁸Lenci designa estes recursos de ontolexicais.

orientadas para a localização automática ou semi-automática de ontologias. Para Espinoza *et al*, (2009:34) é necessário ter em conta outras dimensões no processo de localização de uma ontologia, nomeadamente, a da tradução, a da gestão dos elementos multilingues ao longo do ciclo de vida da ontologia e a do modo de representação do multilinguismo na ontologia.

Assim, no que respeita ao que designam como processo de tradução, e uma vez que cada cultura percepção o mundo de modo distinto, ao traduzir uma ontologia deparamo-nos, segundo os autores, com diferentes situações:

1. *Existência de um equivalente exacto*: esta equivalência exacta é, sobretudo, específica de domínios técnicos altamente especializados, onde se pode encontrar uma equivalência directa entre termos de línguas diferentes que designam certos objectos ou processos.
2. *Existência de equivalentes dependentes do contexto*: quando um termo numa língua pode ser traduzido por mais do que um equivalente na língua alvo, cabendo ao utilizador escolher o mais adequado tendo em conta o contexto da ontologia, o registo sociolinguístico em que a ontologia será usada, etc.
3. *Existência de incompatibilidade entre as conceptualizações*: quando uma cultura categoriza a realidade com um grau de granularidade que não encontra correspondência na cultura alvo. Esta situação pode resultar numa lacuna lexical ou na falta de equivalentes na língua alvo.

Uma outra dimensão identificada é relativa aos problemas de gestão da ontologia. De facto, para além das dificuldades inerentes à localização dos termos da ontologia *per se*, a manutenção e actualização da ontologia localizada, ao longo do seu ciclo de vida, requer, de igual modo, uma atenção especial.

As dificuldades colocadas por este processão foram ainda totalmente consideradas, não existindo metodologias específicas para a gestão das possíveis

mudanças nos termos das ontologias e das correspondentestradições. De facto, na análise feita aos trabalhos sobre gestão de ontologias, nenhum aborda directamente a problemática da gestão da informação multilingue. A gestão dos elementos multilingues de uma ontologia coloca-nos perante diferentes cenários (*idem, ibidem*), nomeadamente:

- i. An ontology term is added, then the ontology label should be translated to all supported languages.
- ii. An ontology term disappears, then all its translations should be removed.
- iii. An ontology term is renamed, then all multilingual labels should be reviewed.

Estas alterações implicam uma atenção específica a este processo, bem como a disponibilidade de recursos e ferramentas linguísticas, por um lado, e de especialistas, por outro, para acompanhar, validar e apoiar a sua execução. A metodologia de representação do multilinguismo na ontologia coloca, também ela, algumas dificuldades. O processo de localização dá origem a uma ontologia com termos em diferentes línguas naturais, sendo três os modelos correntemente identificados como os mais desenvolvidos para o desenvolvimento deste processo (Montiel-Ponsoda *et al.*, 2008):

1. Including multilingual data in the ontology meta-model: this implies localization at the terminological layer since the ontology conceptualization remains unmodified

Na abordagem proposta neste modelo, a inclusão da informação multilingue na ontologia é feita através das propriedades *rdfs:label* e *rdfs:comment*. Este é o modelo mais utilizado pela comunidade da engenharia de ontologias e permite a associação

dos termos da ontologia a etiquetas multilíngues, permitindo a localização do modelo ontológico *at the terminological layer, as labels for ontology classes can be expressed in various natural languages* (Montiel-Ponsoda et al., 2008).

O uso desta abordagem permite que os termos sejam incluídas na ontologia em tantas línguas quantas as desejadas pelo utilizador. O primeiro modelo oferece, na perspectiva da autora, a possibilidade de uma representação adequada para ontologias de domínio altamente especializadas, uma vez que o tipo de conhecimento partilhado entre as diferentes comunidades linguísticas e culturais e a relação de equivalência entre os termos das diferentes línguas é adequada.

No entanto, o uso deste modelo impõe restrições à quantidade de informação linguística a disponibilizar, estando esta limitada a *strings* que não contêm qualquer informação quanto ao sentido do termo na língua de chegada, nem sobre a origem da informação, o que dificulta o processo de localização para as diferentes línguas naturais, assumindo-se, por outro lado, a equivalência a 100% entre os termos nas diferentes línguas, o que nem sempre acontece na realidade.

2. Combining the ontology meta-model with a mapping model: this allows localization at the conceptual layer since conceptualizations in different languages are mapped to each other.

Esta opção de modelação do conhecimento é organizada de acordo com as estruturas específicas a uma dada comunidade cultural de utilizadores, sendo possível o uso de mais do que uma abordagem à modelação do multilinguismo nas ontologias, nomeadamente através do:

- a) mapeamento directo entre duas ontologias monolíngues, em que cada organiza o conhecimento de uma determinada cultura

- b) mapeamento entre duas ontologias monolíngues, através de uma interlíngua. Esta consiste num conjunto de conceitos comuns que permitem o estabelecimento das equivalências⁹.

Sendo um modelo que permite a criação de uma conceptualização por língua e cultura envolvidas e estabelecimento de mapeamentos entre as diferentes conceptualizações, em que cada conceptualização diz respeito à categorização da realidade tal como é percebida por uma dada cultura, o que pode enriquecer a ontologia em determinados domínios, este modelo obriga, no entanto, a um grande esforço em termos do desenvolvimento das várias conceptualizações e da sua interligação/mapeamento, não disponibilizando, por outro lado, uma quantidade de informação linguística muito superior ao modelo anterior.

3. Associating the ontology meta-model to a multilingual linguistic model: localization is performed at the terminological layer, although conceptual layer adaptations are also allowed

Neste modelo a associação de informação multilingue à ontologia acontece através da ligação a dados linguísticos armazenados fora da ontologia. Na abordagem seguida, as camadas terminológica e conceptual são mantidas separadas e a actividade de localização desenvolve-se sobretudo *at the terminological layer*. *However, the ontology conceptualization layer can also undergo modifications, as the creation of language specific ontology modules, in order to meet localization needs.* (Montiel-Ponsoda *et al.*, 2008)

A decisão quanto ao uso de cada um dos três modelos acima descritos depende de dois factores: 1) o tipo de conhecimento de domínio representado na

⁹ O exemplo mais conhecido é o do EuroWordNet

ontologia e 2) a quantidade de informação linguística necessária ao funcionamento da aplicação final.

As ontologias são modelos, obtidos através de negociação e consenso, que representam e organizam explicitamente o conhecimento, mas, como alertam Espinoza *et al.* (2009), certas conceptualizações possuem uma tendência para reflectir particularidades culturais. O tipo de domínio a categorizar torna-se, assim, um elemento crítico para o desenrolar do processo de localização. Tal como afirmam os autores, se a conceptualização é partilhada por todas as culturas implicadas na actividade de localização (*LocalizationActivity*), a localização afectará apenas a camada terminológica¹⁰, i.e., as etiquetas que designam os termos da ontologia. No entanto, se o enfoque recair sobre domínios culturalmente dependentes,

(e.g., the judicature) in which categorizations tend to reflect the particularities of a certain culture, the localization may affect the conceptual layer, i.e. the conceptualization. Optionally, we may be able to account for cultural differences at the terminological layer.

Tendo em conta, por um lado, esta perspectiva e, por outro, a tipologia do domínio e a quantidade de informação linguística necessária ao processo de localização, propõem dois modelos para a representação do multilinguismo nas ontologias:

1. Se a conceptualização representar um domínio consensual, apresentam duas opções: optar pela inclusão da informação

¹⁰*Terms or labels that name ontology classes (Montiel-Ponsoda et al., 2008:67)*

multilingue na ontologia, ou pela associação de um modelo de base linguística (ex.: LingInfo) ao processo de localização¹¹.

2. Se a conceptualização representar um domínio dependente da(s) cultura(s) e se perceberem como necessária a representação das diferenças culturais, apresentam, de novo, duas opções: a criação de uma conceptualização por língua e cultura envolvidas, ou a associação de um modelo externo (ex.: LIR) que permita dar conta das divergências culturais ao nível da camada terminológica. (*idem, ibidem*)

Abordagens ao processo de localização de ontologias

Os problemas e dificuldades supramencionados, que retractam e oferecem uma visão da especificidade e complexidade do objecto e dos conteúdos a localizar, bem como da riqueza de possibilidades abertas ao desenvolvimento de aplicações e abordagens, conduziram ao surgimento de propostas de características distintas que propõem modelos, metodologias e ferramentas com o intuito de contribuir para estabelecer a interligação da informação multilingue, terminológica e linguística, com a ontológica. Destas propostas destacam-se projectos como o Eurowordnet (Vossen, 1997), o GENOMA-KB (Cabré *et al.*, 2004) ou a Termonography (Kerremans e Temmerman, 2004), ferramentas como o LingInfo (Buitelaar *et al.*, 2006) e o LexOnto (Cimiano *et al.*, 2007), ou modelos como o LIR - LinguisticInformationRepository (Peters *et al.*, 2008), cujas características e abordagens propostas analisaremos em seguida¹².

¹¹A decisão entre estas duas opções dependerá das necessidades linguísticas da aplicação a desenvolver. *If morphosyntactic data is needed for the purpose of Information Retrieval or Information Extraction, for example, the most suitable option will be the association of an external model such as LingInfo, which enriches the ontology with a great amount of morphosyntactic information.*

¹² Foram também identificados outros modelos, projectos e abordagens dedicados ao enriquecimento multilingue de recursos de conhecimento, como as propostas de localização de glossários (ex.: FAOTERM), de localização de bases de dados terminológicas (ex.: Eurodicautom), de localização de

Um dos projectos e metodologias que mais destaque alcançou neste domínio é o Eurowordnet (Vossen, 1998). O Eurowordnet é uma base dados lexical multilingue, que não é específica a qualquer domínio, construída com base nos Wordnets existentes, representando cada um deles uma língua europeia diferente, que propõe um modelo constituído por ontologias monolingues autónomas, cada uma das quais reflectindo as especificidades linguísticas e culturais de uma determinada língua, ligadas entre si por um *Inter-Lingual-Index* (Peters *et al.*, 1998:150), que permite o acesso às traduções e o estabelecimento de um mapeamento através de equivalências entre as diferentes ontologias.

Kerremans e Temmerman (2003), por seu lado, propõem uma metodologia funcional, a *Termontography*, cujo método sumarizam do seguinte modo:

first of all, in close collaboration with specialists of the domain of interest, a framework of domain specific categories and intercategory relationships is developed. This framework facilitates the manual and semi-automatic extraction of knowledge from a textual corpus. It will gradually evolve towards an enriched and more fine-grained network of semantic relations, reflecting culture-specific conceptualizations. (Kerremans *et al.*, 2003:4)

Propõem, assim, uma abordagem multidisciplinar, em que combina as teorias e métodos da análise terminológica multilingue, partindo da perspectiva sociocognitiva (Temmerman, 2000), com os métodos e princípios para a análise ontológica propostos por Gómez-Pérez *et al.* (1996); Fernández *et al.* (1997) e Sure e Studer (2003), cujo ponto de partida é “a pre-defined, language-independent framework of

thesaurus (ex.: Agrovoc;Eurovoc) que, apesar da sua importância e das metodologias propostas, não serão alvo, neste artigo, de uma análise mais aprofundada, uma vez que não lidam directamente com o processo de localização de ontologias.

domain-specific UoUs¹³ and intercategory relationships, set up in collaboration with field specialists, to which lexicalisations (from a domain-specific corpus) are mapped.” Definem esta abordagem como sendo terminológica, sendo o conhecimento terminológico multilingue extraído a partir de um corpus textual e estruturado tendo em conta uma *culture independent and task-oriented Framework of domain-specific knowledge which can be further refined in a culture-specific layer.*

Modelos de base linguística

Outros modelos adoptam uma abordagem distinta, que consiste na associação do meta-modelo da ontologia a um modelo linguístico, mantendo ambos separados. O modelo usado para representar a informação linguística pode ser uma base de dados, como nos casos dos projectos GENOMA-KB (Cabrét *et al.* 2004) e OncoTerm (Faberet *et al.*, 2002), ou uma ontologia, como no caso dos projectos LingInfo (Buitelaar *et al.*, 2006) e LexOnto (Cimiano *et al.*, 2007).

Esta abordagem, que separa os diferentes componentes, pretende permitir a integração de toda a informação terminológica necessária, bem como o estabelecimento de relações entre os elementos terminológicos, o que resultaria num enriquecimento dos elementos conceptuais. Para além disso, estes sistemas são considerados como independentes do domínio, podendo ser interligados a qualquer ontologia de domínio.

Um modelo que recebeu grande destaque foi o LingInfo, sistema desenvolvido como parte do projecto SmartWeb2. A abordagem conceptual assenta na ideia de construção de uma camada linguística para uma Ontologia da Web Semântica ou, mais especificamente, um "modelo lexical multilingue/multimédia para ontologias" (Buitelaar *et al.* 2006).

¹³UoU – *Unit of Understanding* – a noção de Unidade de Compreensão foi introduzida pela abordagem sociocognitiva à Terminologia, proposta por Temmerman (2000).

No LingInfo o conteúdo e o conhecimento são organizados em quatro camadas, estando a camada respeitante à ontologia localizada na camada central, aparecendo as características linguísticas, bem como as suas associações subsequentes com a camada central, localizadas nas camadas externas ao centro, ficando o conteúdo textual associado à camada exterior. Este modelo, apesar de atender aos aspectos de lexicalização da ontologia multilingue, é, como modelo, algo complexo para ser (re)utilizado por um público não-linguista, podendo dificultar a tarefa de conceptualização e implementação da ontologia ao exigir o conhecimento de formalismos linguísticos relativamente complexos.

Um modelo anterior, proposto por Cabré *et al.*, 2004, o GENOMA-KB, desenvolvido para recolher e representar o conhecimento no domínio da genética, pode ser descrito sumariamente como uma base de dados de conhecimento multimódulo, composta por quatro módulos interdependentes – o módulo da ontologia, o módulo da base de dados terminológica, o corpus (multilingue) e o módulo das entidades. O modelo propõe uma abordagem de cariz marcadamente terminológico, em que as *specialized knowledge units* (Feliu *et al.*, 2004) extraídas dos documentos que compõem o corpus multilingue são depositadas numa base de dados terminológica e ligadas a conceitos na ontologia. No módulo da ontologia, que aparece directamente ligado ao da base de dados, estes conceitos são depois alvo de uma “*knowledge organization based on a set of both hierarchical and non-hierarchical conceptual relations*.” (*idem, ibidem*).

Esta abordagem obriga à existência prévia da ontologia, construída recorrendo a um conjunto de base de conceitos de mais alto nível com base na perspectiva de especialistas, para que a ligação possa acontecer, o que implica, por exemplo, em termos teóricos, a biunivocidade entre termos e conceitos e, em termos práticos, que a ontologia seja construída num momento anterior ao da base de dados.

Um outro modelo, mais recente e que tem vindo a ser reutilizado no âmbito de diferentes projectos é o LIR - *LinguisticInformationRepository* (Peters *et al.*, 2008).

Este modelo foi recentemente integrado e desenvolvido no âmbito do projecto NeOn¹⁴, com o objectivo de localizar uma determinada categoria da ontologia de acordo com o universo linguístico e cultural de uma língua natural e capturar as especificidades da tradução entre línguas. Os autores descrevem o modelo como sendo holístico, uma vez que, defendem, contém um conjunto *of complete and complementary classes that allows the localization of a certain ontology element to a specific linguistic and cultural context (idem, ibidem)*, permitindo, paralelamente, que se aceda e integre a informação multilingue contida em recursos lexicais distribuídos e heterogéneos ao garantir um acesso homogéneo à informação. Este modelo consiste em:

Lexical Entries that in its turn consist of Lexicalizations (or word forms), Senses expressed by means of the Definition in a certain Language Context, Usage Context, Source (of any linguistic element), Notes (that can be related to any linguistic element), and relations established between Lexical Entries as, for example, "has Synonym", "has Translation" or "has Scientific Name".

O modelo LIR é um elemento independente da ontologia e procura permitir aos linguistas trabalhar em ambientes distribuídos de modo a gerir e editar a informação linguística associada à ontologia. A aplicação deste modelo surge associada ao uso da ferramenta LabelTranslator, um plug-in destinado à tradução automática dos termos da ontologia, que analisaremos adiante.

Os aspectos que diferenciam os vários sistemas supradescritos são determinados pelo tipo de elementos linguísticos que compõem cada modelo. Dependendo das necessidades linguísticas do utilizador final, alguns modelos podem considerar-se mais apropriados do que outros. De facto, o modelo a implementar depende, em grande parte, das necessidades predefinidas dos

¹⁴<http://www.neon-project.org/>

utilizadores e das aplicações e, no caso de uma rede colaborativa, dos propósitos da rede e, em última análise, do próprio modelo adoptado para potenciar a partilha de conhecimento.

Ferramentas de base linguística

Para além dos modelos e metodologias já analisados, foram também desenvolvidas aplicações para facilitar o processo de localização de ontologias. Dessas focalizaremos a nossa atenção naquela que consideramos a mais desenvolvida, o LabelTranslator, aplicação de localização de ontologias desenvolvida por Espinoza *et al.* (2007), que pretende ser um auxiliar à localização automática de ontologias recorrendo a uma metodologia que os autores denominaram de *OntologyLocalizationActivity* (OLA).

O sistema desenvolvido é composto por: 1) *the extraction of possible single-lemma translations from semantic and translation resources*, 2) *the disambiguation of the translation senses*, and 3) *the ranking of the translations* (Espinoza *et al.*, 2007: 01). Esta aplicação consulta bases de dados lexicais, dicionários e terminologias e propõe uma metodologia para medir a similaridade semântica entre o conceito da ontologia e o seu contexto lexical e semântico. O sistema, tal como é descrito por Espinoza *et al.* (2008:02), é composto por seis passos, fornecendo uma tradução automática acompanhada de mecanismos de desambiguação semântica sofisticados.

De acordo com os autores, o serviço de tradução tem como input uma *ontology label l described in a source language and returns a set of possible translations* $T = \{t_1; t_2; \dots; t_n\}$ *in a target language*. De modo a descobrir e propor as traduções para cada termo, o sistema acede a diferentes recursos lexicais, nomeadamente: 1) *remote lexical databases as EuroWordNet*, 2) *multilingual dictionaries as GoogleTranslate, Wiktionary, Babelfish, and FreeTranslation*, and 3) *other lexical resources as LATE*. (Espinoza *et al.*, 2008, 04). De modo a obterem uma validação interlinguística das traduções, os recursos externos do LabelTranslator foram limitados àqueles que possuem

informação multilíngue, como o Eurowordnet, acrescidos de outros recursos específicos do domínio, como o Agrovoc, para alargar a cobertura do vocabulário.

No entanto, esta ferramenta e a metodologia associada têm algumas limitações como reconhecem os autores (2008,13), nomeadamente, a elevada dependência dos serviços de tradução da tipologia de recursos disponíveis e da sua abrangência do domínio, e a impossibilidade de *learning of new lexical patterns*, que limita a escalabilidade da ferramenta. Para além destas limitações, o nome LabelTranslator é auto-explicativo. O processo de localização ocorre sempre depois de a conceptualização estar terminada e abrange apenas os termos disponíveis no sistema, permitindo *a posteriori* um acesso mais completo à informação disponível sobre o domínio e condicionando, desse modo, a capacidade dos diferentes agentes de melhorarem as estruturas de representação desse conhecimento e de fazerem a manutenção da ontologia. O LabelTranslator procura, no entanto, ao atribuir aos utilizadores a capacidade de decisão sobre as traduções, que a capacidade expressiva das ontologias sobre o conhecimento do domínio seja a mesma nas diferentes línguas de chegada.

Localização no contexto dos sistemas de gestão do conhecimento

Ressaltam, da análise desenvolvida até este ponto, um conjunto de paralelismos e de diferenças a vários níveis entre o processo de localização e o processo de localização de ontologias que interessa identificar, para melhor entender como este se desenrola e, sobretudo, para melhor compreender como situar um sistema de representação do conhecimento enquanto objecto desse mesmo processo.

Um primeiro paralelismo, que podemos considerar de mais alto nível e anterior ao processo de localização, aproxima o processo de conceptualização de uma ontologia e o processo de internacionalização. Assim, o processo de internacionalização, tal como já havia sido descrito, dimensiona um produto para

que seja funcional em vários mercados, o que implica uma certa depuração das convenções culturais e linguísticas que possam perturbar este processo. Transferindo este processo para o domínio da gestão e da representação do conhecimento, mais especificamente, para a construção¹⁵, no seio de uma comunidade multilingue e multicultural, de um sistema conceptual que servirá de base à representação do conhecimento em termos ontológicos, a internacionalização pode ser percebida como um processo de generalização de um sistema de organização do conhecimento (KOS¹⁶)

with reference to the representation of concepts and their semantic relations so that it can handle multiple languages and viewpoints in a networked environment. (...) It implies the identification of the most appropriate structural form for a core KOS and the decision for a concrete KOS that is taken as initial point for developing delocalized concepts and structures. (Boteram *et al.* 2009)

Este processo ocupa, como afirma Pym (2008), um papel central, *which builds on a concept of artificial equivalence¹⁷ in order to centralize decision-making power in a professional interculture, understood as a place created in the overlaps of primary cultures*. Já no que respeita à localização propriamente dita, e segundo Hubrich *et al.* (2009), a localização pode ser descrita, quase por contraste, como uma elaboração de *local/national specific concepts and semantic relations of KOS focused on the particular language, culture, historical development, politics, social structures, etc.*

¹⁵Esta construção é feita com base em concepções já existentes (convencionalizadas) sobre o artefacto a construir.

¹⁶ KOS – Knowledge Organization System

¹⁷When we no longer have a source, however, equivalents are artificially determined in the place of internationalisation. The central company determines what terms will be used, and can do so for all the languages involved, quite independently of previous usages in the end-locales. Localisation thus makes artificial equivalence far more visible. (Pym, 2008)

Outros paralelismos verificam-se ao nível do objecto a localizar. De facto, um sistema de representação do conhecimento – no nosso caso, um sistema conceptual – tem, no momento da sua conceptualização, características de grande fluidez, uma vez que é um sistema aberto à negociação e construído por consenso, sendo os seus autores, tal como num projecto típico de localização, não um indivíduo, mas um sujeito corporativo, o mesmo que inicia e, em última instância, valida a versão final¹⁸. Do mesmo modo, quando se localiza um sistema conceptual os condicionalismos do localizador/tradutor provêm dos objectivos do iniciador da localização e das limitações ou orientações explícitas dadas por este, nomeadamente, em termos da delimitação do domínio e do conhecimento a representar, e não de um texto original, que não existe, ainda que possam existir outro tipo de recursos. A localização de sistemas conceptuais excede assim, também ela, o processo de tradução, representando este apenas uma das fases.

Um outro condicionalismo advém do contexto e da situação comunicativa em que ocorrem os processos de conceptualização e a localização, que são desenvolvidos num ambiente virtual, onde a comunicação, a partilha, a representação do conhecimento e a sua localização são apoiadas, muitas vezes, em ambientes digitais de comunicação, nem sempre totalmente desenvolvidos para integrar todos os processos e as necessidades inerentes a cada. Outra aproximação entre o processo de localização e o de localização de ontologias existe quando se considera o “leitor alvo”, que, em ambos os casos, é, não tanto um leitor, assumindo antes o papel de utilizador, o que obriga a que a construção do sistema e a sua localização tenham em conta (1) o uso potencial que o utilizador fará do produto localizado e (2) as restrições impostas pela aplicação e orientações fornecidas pelos especialistas/cliente.

Finalmente, também no caso da localização de uma ontologia, sendo esta resultado de um esforço colectivo, os actores envolvidos no processo de

¹⁸Esta característica é essencial para potenciar o processo partilhado de construção de ontologias multilíngues e de localização no seio de uma rede colaborativa.

localização são remetidos ao anonimato, apesar da participação activa que têm na sua construção, uma vez que, após a sua formalização, a ontologia será um elemento da aplicação que apoiará a gestão do conhecimento, não mantendo qualquer referência aos participantes no processo de conceptualização e localização.

Conclusão

A abordagem à localização de ontologias desenrola-se, necessariamente, num ambiente multidisciplinar, envolvendo terminólogos, gestores de conhecimento, especialistas de domínio e engenheiros de ontologias, que procuram, num contexto e com um propósito específico, estruturar e representar o conhecimento de um domínio. Sendo a localização um conceito bem estabelecido e bem aceite na comunidade de engenharia de software, não nos parece, no entanto, que a definição do conceito seja suficientemente abrangente para os propósitos da localização de ontologias.

A construção de uma ontologia procura criar um artefacto que, na sua génese, é independente, em termos conceptuais, da língua em que é criado e que corresponde aos anseios de uma comunidade internacional, i.e., funciona num ambiente de pesquisa e gestão do conhecimento de um domínio específico e que, tal como no processo de internacionalização, seja facilmente adaptável às necessidades e especificidades linguísticas e culturais locais. Daí que a localização de ontologias deva ser entendida não só como a tradução ou a adaptação de um conceito de acordo com as especificidades linguísticas e culturais do país ou região a que se destina, mas sobretudo como a adequação da representação dos conceitos e das suas relações semânticas ao seu ambiente nativo (de destino), no respeito pela sua identidade linguística e cultural, bem como pelas suas estruturas sociais, profissionais e políticas.

Esta localização envolve, assim, por um lado, a necessidade de traduzir adequadamente os conceitos internacionalizados para cada língua alvo e, por outro,

a necessidade de adequar o sistema de organização de conhecimento à língua alvo, tendo em conta o modo como os conceitos e as relações semânticas se interrelacionam na representação do conhecimento em termos locais.

Localizar uma ontologia consiste, assim, em adaptá-la a uma língua e a uma cultura meta, de modo a que responda às necessidades dos mercados específicos e à preferência dos especialistas e dos potenciais utilizadores.

Referências bibliográficas

AGROVOC Thesaurus maintenance and refinement tools:
http://www.fao.org/aims/tools_thes.jsp

BOTERAM, Felix, Hubrich, Jessica (2008). Towards a comprehensive international Knowledge Organization System Networked Knowledge Organization Systems NKOS. ECDL-Conference, Århus.

BUTTELAAR, Paul, Sintek, Michael, Kiesel, Malte (2006). A Lexicon Model for Multilingual/Multimedia Ontologies In: Proceedings of the 3rd European Semantic Web Conference (ESWC06), Budva, Montenegro.

BUTTELAAR *et al.*, 2009. Towards linguistic grounded ontologies. In *The Semantic Web: Research and applications. Lecture Notes in Computer Science*, 2009, Volume 5554/2009, 111-125, DOI: 10.1007/978-3-642-02121-3_12

CABRÉ, M. Teresa; Bach, C.; Estopà, R.; Feliu, J.; Martínez, G.; Vivaldi, J. (2004a). "The GENOMAKB project: towards the integration of concepts, terms, textual corpora and entities". LREC 2004 Fourth International Conference on Language Resources and Evaluation . Lisboa: European Languages Resources Association. pp. 87-90.

C.K. Ogden and I.A. Richards. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Routledge&Kegan Paul. 1923.

CORTESE, Giuseppina. (2007). *LSP: Multilingual deficiency, multicultural ambiguity*. in *Evidenced-based LSP – Translation, text and Terminology*. Ahmad & Rogers (Eds)

DECLERCK, Thierry, Hans-U. Krieger, Thomas, Susan M., Buitelaar, Paul & Sean O'Riain, Tobias Wunner, Gilles Maguet, John McCrae, Dennis Spohr, Elena Montiel-Ponsoda. *Ontology-based Multilingual Access to Financial Reports for Sharing Business Knowledge across Europe*. 2010.

Directorate-General for Translation of the European Commission, July 2005, *Translation Tools and Workflow*, in http://ec.europa.eu/dgs/translation/index_en.htm

FELIU, J; Vivaldi, J.; Cabré, M.T. (2002) "Towards an Ontology for a Human Genome Knowledge Base". LREC2002. Third International Conference on Language Resources and Evaluation. Proceedings. Las Palmas de Gran Canaria, pp. 1885-1890. ISBN: 295-1740-808.

FAO's Role in Information Management and Dissemination-Challenges, Innovation, Success, Lessons Learned: <http://www.fao.org/docrep/008/af238e/af238e04.htm>

FRY, Deborah (2001): *The Localization Industry Primer*. LISA. [<http://www.lisa.org>]

FU, B.; Brennan R. and O'Sullivan, D. (2010). *Cross-Lingual Ontology Mapping and Its Use on the Multilingual Semantic Web*. Proceedings of the WWW2010, April 26-30, 2010, Raleigh, North Carolina, USA

GUIDÈRE, Mathieu (2004). *From adaptation to advertizing localization*", *Localization, the problem of training*, Paper presented at the "Localization: the problem of training in industry, government departments and training establishments" Symposium, Dir. James Archibald, Brossard (Québec), Linguatechéditeurinc , pp. 69-95.

GONZÁLEZ, L and P. Hernández: *La terminología en la Comisión Europea*
Link:

<http://www.termilat.info/public/env100.doc>

HOFT, Nancy (1995). *International Technical Communication – How to export Information about High Technology*. New York: John Wiley & Sons.

HOVY, Eduard *et al.* (2008). *Multilingual Information Management: Current Levels and Future Abilities*. US National Science Foundation.

<http://www.cs.cmu.edu/~ref/mlim/chapter1.html>

KARSCH, Barbara Inge (2009). Profile of a terminologist in localization environments. In *The Journal of Internationalization and Localization*, vol. I.

KERREMANS, Koen, Temmerman, Rita, Tummers, Jose. Representing Multilingual and Culture-Specific Knowledge in a VAT Regulatory Ontology: Support from the Termontography Method. 2003.

KERREMANS, K. and R., Temmerman. (2004a). "Towards Multilingual, Termontological Support in Ontology Engineering", in Proceedings Workshop on Terminology, Ontology and Knowledge Representation - 22 & 23/01/2004, Lyon, France.

KERREMANS, K., R. Temmerman and J. Tummers. (2004b). "Discussion on the Requirements for a Workbench supporting Termontography", in Proceedings Euralex 2004, Lorient, France.

MONTIEL-PONSODA *et al.* 2008, 12 NeOn: Lifecycle Support for Networked Ontologies D2.4.2 Multilingual and Localization Support for Ontologies.

MORENO ORTIZ A. Y Pérez Hernández (2000). "Reusing the Mikrokosmos Ontology for Concept-Based Multilingual Terminology Databases". Proceedings of 2nd International Conference on Language Resources and Evaluation (LREC-2000). Athens, pp. 1061-1067.

PAZIENTA M., Stellato A.. Linguistically motivated ontology mapping for the Semantic Web. In Proceedings of the 2nd Italian Semantic Web Workshop, 14-16, 2005

PAZIENZA M. T., Stellato A.. Exploiting linguistic resources for building linguistically motivated ontologies in the Semantic Web. In Proceedings of OntoLex Workshop, 2006

PYM, Anthony, 2008. Localization from the Perspective of Translation Studies: Overlaps in the Digital Divide? <http://www.elda.org/en/proj/scalla/SCALLA2004/Pym-2.pdf>

SCHÄLER, R., (2007), "Localization", in: *Encyclopedia of Translation Studies*, Baker, M. and Saldanha, G. (Eds.), second edition, 157-161.

R. PALMA, P. Haase, O. Corcho, A. Gómez-Pérez, and Q. Ji. An editorial workflow approach for collaborative ontology development. In 3rd Asian Semantic Web Conference. ASWC 08. Bangkok, Thailand, 2008.

T. TUDORACHE, N. Noy, S. Tu, and M. Musen. Supporting collaborative ontology development in Protege. In International Semantic Web Conference, 2008.

INGRID FALK, Samuel Cruz-Lara, Nadia Bellalem, Lotfi Bellalem (2009). LINGUISTIC INFORMATION FOR MULTILINGUALITY IN THE SEMBYSEM PROJECT

PAUL BUTTELAAR, Michael Sintek, Malte Kiesel. A Lexicon Model for Multilingual/Multimedia Ontologies. In: Proceedings of the 3rd European Semantic Web Conference (ESWC06), Budva, Montenegro, June 2006.

TEXIN, Tex "The Challenges of Defining locale Structures : Language and culture identifiers need to be adaptable", *Multilingual Computing & Technology*, Vol. 14, No 55, Issue 3, Sandpoint (Idaho), Multilingua, Computing Inc, pp. 24-28.

VOSEN, P. (1998) "Introduction to EuroWordNet". In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. *Computers and the Humanities*, Volume 32, Nos. 2-3 1998. 73-89.

WIM Peters and Piek Vossen and Pedro Díez-Orzas and Geert Adriaens, "Cross-linguistic Alignment of Wordnets with an Inter-Lingual-Index," pp. 149-179, 1998.