A Note on Ontology Localization

Philipp Cimiano\textsuperscript{a,∗}, Elena Montiel-Ponsoda\textsuperscript{b}, Paul Buitelaar\textsuperscript{c}, Mauricio Espinoza\textsuperscript{b} and Asunción Gómez-Pérez\textsuperscript{b},
\textsuperscript{a}Semantic Computing Group, Cognitive Interaction Technology Excellence Center (CITEC), Universität Bielefeld, Germany
E-mail: cimiano@cit-ec.uni-bielefeld.de
\textsuperscript{b}Ontology Engineering Group, Universidad Politécnica de Madrid, Spain
E-mail: \{emontiel, jespinoza, asun\}@fi.upm.es
\textsuperscript{c}NLP Unit, DERI - National University of Ireland, Galway
E-mail: paul.buitelaar@deri.org

Abstract. We revisit the notion of ontology localization, propose a new definition and clearly specify the layers of an ontology that can be affected by the process of localizing it. We also work out a number of dimensions that allow to characterize the type of ontology localization performed and to predict the layers that will be affected. Overall our aim is to contribute to a better understanding of the task of localizing an ontology.

Keywords: ontology localization, multilinguality, ontology engineering

1. Introduction

Ontology localization has been defined in [15] as “the adaptation of an ontology to a particular language and culture”. In this short note, we propose a more general definition which de-emphasizes the adaptation to a particular language as follows: “ontology localization is the process of adapting a given ontology to the needs of a certain community, which can be characterized by a common language, a common culture or a certain geopolitical environment.” The adaptation to the language spoken by the target community is thus one possible aspect of ontology localization. In this sense ontology localization is clearly a specific type of ontology reengineering activity where the requirements (for the reengineering) are provided by the needs of the community to which the ontology is adapted. Ontology localization is thus an activity with very pragmatic goals, i.e. fostering reuse of ontologies already available for the domain in question instead of building them from scratch. Therefore, there is also an economic aspect of ontology localization as it has the potential to reduce costs compared to building a completely new ontology for the target community.

Ontology localization is a transformation process that takes an ontology as input and produces an (adapted) ontology as output, whereby the output can be the same ontology, extended with labels in additional languages, or a new ontology. Ontology localization can affect two different layers: the surface, or lexical layer, of an ontology, or the conceptualization itself. We consider the lexical layer of an ontology to include all the labels, definitions and accompanying documentation in natural language that make the ontology human understandable. It can already be foreseen that the lexical layer will always undergo modifications, regardless of whether the target community speaks a different language or not. The underlying reason for this is that any discrepancies at the conceptualization layer - due to differences at the cultural or geopolitical background - will have an impact on the lexical layer, since language is the means we have to understand and experience reality.

In this contribution we aim at clarifying the notion of ontology localization as well as the different layers of an ontology it affects. We describe how the localization of the different layers (lexical and conceptual)
interact and introduce different dimensions that characterize the localization process and which by a large extent determine the ontological layers that are affected by the localization activity. Overall, our goal is to contribute to a better understanding of the ontology localization process.

This short note is structured as follows: in Section 2 we discuss in more detail the different ontology layers (lexical vs. conceptualization) that can be distinguished and how these different layers interact; in Section 3 we then present an overview of different dimensions that can be identified in determining the type of localization that is to be performed and the layers that are affected. Section 4 discusses several real-life cases in which ontology localization has been performed for different purposes. Finally, before concluding, in Section 5 we discuss the relation to other areas.

2. Layers in Ontology Localization and their Interaction

2.1. Lexical Layer

The lexical layer of an ontology comprises: i) the labels of the concepts, properties and individuals defined in the ontology, ii) natural language definitions of these entities, as well as iii) the documentation accompanying the ontology, which describes its scope and purpose, its usage etc. The inclusion of definitions of ontology entities in natural language is in fact part of the “clarity” criterion that good ontologies should fulfill according to Gruber [8].

Obviously, the lexical layer is language-specific and is thus clearly affected by any ontology localization process, even when the adaptation is done within the same linguistic system. This means that the changes motivated by the cultural environment in which the ontology is to be used - be it within the same linguistic system or not - will be reflected at the lexical layer.

A straightforward way to localize the lexical layer is to provide a 1:1 translation for each label, definition and the accompanying documentation. However, as we will see below, the fact whether a 1:1 translation is suitable ultimately depends on the purpose for which an ontology is localized.

2.2. Conceptualization Layer

While the translation of labels is the most important aspect of the ontology localization process, the conceptualization may also need to be adapted if so required, e.g., by a different cultural or geo-political context. Consider an ontology about political functions and charges. Most democratic systems distinguish for example the role of head of government in the sense of head of the executive power vs. the role of head of state with mainly representative function. An ontology designed to model political functions and charges in Germany would further distinguish between the Bundeskanzler (chancellor) playing the role of head of government and the Bundespräsident playing the role of the head of state.

If we want to use an ontology about political charges engineered for the German geo-political and cultural environment in applications that concern (also) other countries, e.g., the UK or Spain, we will need to adapt the conceptualization expressed by the ontology. In the case of the UK, we would introduce the class of prime minister as head of government and the queen as head of state. In the case of Spain, we would introduce the class of presidente (president) as head of government and the monarca (monarch) as head of state. While one could argue that this adaptation can also be achieved at the lexical level, e.g., by adding additional labels (prime minister, presidente) for the class Bundeskanzler or (queen, monarca) for the class Bundespräsident, this is clearly insufficient as these concepts will have different extensions and even intensions. In this case, the adaptation to a different geo-political and cultural reality may require more than a 1:1 translation, i.e. a change as well in the underlying conceptualization.

It is important to emphasize that the adaptation of the conceptualization layer will be primarily driven by the inexistence of conceptual equivalents (or concepts with the same granularity level) in the target community, whenever the final purpose of the ontology is to be equally valid in source and target culture. If the concept of Bundeskanzler serves the function of head of government in the German culture, and we
aim at reusing the ontology in the UK, we should not make the mistake to stick to the words and translate it as federal chancellor, just because the word exists in the English language, unless the purpose of the localization is to “paraphrase” in English how the German political structure is organized. We will come back to this in section 3.

2.3. Interaction between Layers

The different layers that we have sketched above do certainly interact in the sense that changes to one layer can not be performed completely independently of the other layers. This means that changes to the conceptualization will be inevitably reflected at the lexical layer, and changes to the lexical layer may also end up influencing the conceptual layer. The only case in which this interaction is not reciprocal is when the lexical layer undergoes modifications without affecting the conceptualization, but this is not feasible in the other direction.

- First of all, changes in the conceptualization will also require the adaptation of the lexical layer (see the example of Bundeskanzler and prime minister or presidente above). The dependency of changes to the lexicon on changes in the conceptualization is clearly unavoidable if the target ontology should have appropriate labels in the language spoken by the target community.

- Second, in some cases where only the lexical layer of the ontology is changed to document the ontology for another language, unintended meaning shifts may occur in case the term chosen as a 1:1 translation has different connotations in the target community than in the source community. This would be the case of the term designating marriage in English or in German (Ehe). In both cases, marriage is defined as the union of people of different sexes, whereas the Spanish term matrimonio, which is the direct translation of marriage and Ehe that we would find in any dictionary, has a wider scope since it embraces people of different or the same sex.

- Third, even the adaptation of the lexical layer might require changes in the conceptualization. Imagine a geographical ontology designed by speakers of French. Speakers of the French language might be more inclined to include the distinction between rivers flowing into the sea (‘fleuve’) and rivers flowing into other rivers (‘rivière’) into their ontology. This distinction is clearly not a fanciful one, but it simply shows how the French culture experiences the world. The fact that this distinction is directly lexicalized in the French language (in contrast to other languages such as English, Spanish or German), makes a French ontology designer prone to include this distinction into the ontology. When localizing this ontology into a different language (say English, Spanish or German), an ontology engineer has two basic choices:

1. keeping the distinction between rivers flowing into the sea and rivers flowing into other rivers in the conceptualization. This means that there will be no direct lexicalization in terms of one designation that can be used as label for each concept, but a paraphrase in the target languages.

2. remove this distinction and keep only the concept of a river without distinguishing further between rivers flowing into the sea and rivers flowing into other rivers. In this sense the ontology engineer is de-constructing the original ontology by removing distinctions that result from granularity layers that are not completely shared by the cultures involved. This decision may be imposed by interoperability reasons. However, there still exists the option of keeping some cultural specificities at the lexical layer by means of powerful linguistic models that have been developed lately to associate linguistic information to ontologies (see also [2], [3] or [14] on this issue).

The decision will be taken by considering whether the real-world distinction between rivers flowing into the sea and rivers flowing into other rivers is a relevant one considering the applications that the target ontology is assumed to support.

The fact that decisions at different layers are clearly dependent on each other makes ontology localization a challenging and non-trivial endeavour, not to mention the goal to support ontology localization semi-automatically.
3. Dimensions of Ontology Localization

Localization implies the existence of an input ontology that is ‘adapted’ to serve the purposes of a different linguistic and/or cultural community. From our viewpoint, localization has an eminently practical importance as it fosters the reuse of already conceptualized knowledge in different linguistic and cultural settings. As already outlined in Section 1, this adaptation may have different implications, that is, different layers of the ontology will be affected by the localization to different extents.

We have identified three crucial dimensions that determine the type of localization to be performed, already outlined in [6]:

– **international (standardized) domain vs. culturally influenced domain**: Some domains are clearly “internationalized” or “standardized” as a byproduct of globalization activities driven by the need to exchange data on a global level. This is often the case in very technical domains, e.g., in engineering and finance which have standards for processes (e.g. the ISO standards) or reporting standards (e.g. XBRL in the financial domain). Other domains are more culturally influenced, e.g., in the public administration of various countries on issues such as taxation, laws, political charges etc. The resulting models of the same domain in different communities are going to show an important divergence.

– **functional vs. documental localization**: Inspired by Functionalist theories to translation (see [13]) we state that an ontology might be localized with different goals in mind. On the one hand, the goal of the target ontology can be to have the same function in the target community as the original ontology in the source community. Take again the example of ontologies in use within public administration, e.g., ontologies modelling immigration procedures. If we want to port these models to a different geo-political reality, we will need to change the conceptualization to fit the requirements of the target community and to make sure that the ontology can have the same function in applications that the original model had in the source community. Functional localization thus typically implies the creation of a new ontology on the basis of the old one, adapted to the requirements of the target community. In the simplest case the ontology can be reused as is, but from a practical point of view this will rarely be the case. In documental localization on the other hand, the purpose is only to support the use of the original ontology by members of another (linguistic) community. Let us take again the example of the ontology modelling immigration procedures. In order to make these procedures accessible to an immigrant (a member of a different cultural and linguistic community), we need to document the meaning of these procedures in their language. This does not involve the creation of a completely new ontology, but only the documentation of the existing ontology in a different language.

– **interoperable vs. independent ontology**: One important aspect when reengineering an ontology to meet the needs of a certain target community is how interoperable the new ontology needs to be to the original one. There is clearly a trade-off here between meeting the special needs of the target community and maintaining a certain level of interoperability. If the target ontology should still be used to exchange data between the source and target community, the changes to the conceptualization should be restricted to those strictly needed to accommodate both cultures, and guarantee interoperability in this way. If the target ontology will be used as an independent ontology in an equivalent manner (see point above on the functional dimension), then significant changes to the conceptualization are acceptable to meet the needs and capture the specificities of the target community.

Given these different dimensions of the ontology localization task, we can now define which ontology layers will be affected depending on the type of localization that is to be to carried out. The following table summarizes this for the international vs. culturally influenced domain and functional vs. documental localization dimensions:

<table>
<thead>
<tr>
<th>Purpose / Type of domain</th>
<th>International</th>
<th>Culturally-influenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional</td>
<td>n.a.</td>
<td>conceptualization, lexical layer</td>
</tr>
<tr>
<td>Documental</td>
<td>lexical layer</td>
<td>lexical layer</td>
</tr>
</tbody>
</table>
The reason why the configuration corresponding to an internationalized domain and functional localization is not applicable according to the above table is that according to our definition the functional localization implies creating a new ontology. In the case of an internationalized domain, we would however not like to come up with different ontologies, but share one ontology across national borders. We might however document the meaning of the classes and relations defined in the ontology modeling an internationalized domain in different languages so that it is accessible by speakers of various languages. The localization will thus affect only the lexical layer of the ontology in this case.

In the case of a culturally influenced domain, the main distinguishing criterion is whether the ontology is supposed to be used in a different geo-political and cultural environment, in which case the conceptualization needs to be adapted, or the goal is to allow people with a different cultural and linguistic background to access and use the ontology.

The degree of interoperability is not a crisp dimension and only affects the case of the functional localization of a culturally influenced domain. Depending on the degree of interoperability desired (in particular the granularity at which the ontologies need to be interoperable), the conceptualization can change more or less. Imagine that each country has a different system for capturing censorship information. If we want to maintain interoperability at least at the level of counting the number of citizens with a certain confession and so on across Europe, then the corresponding concepts have to be kept intact when localizing the ontology (by reusing the concept or via appropriate mappings).

4. Examples of localization projects

Domain type and function of the localized ontology can be combined to result in different scenarios. In the following, we describe real-life use cases of localization projects, illustrating the interplay between the different layers and dimensions discussed.

Use Case 1: GenomaKB In the GenomaKB project, terminology experts of the Institute of Applied Linguistics at the Universitat Pompeu Fabra in Barcelona, Spain, created a biomedical knowledge base of the human genome in three languages (Spanish, English and Catalan) to assist terminologists, translators and scientific journalists working in this domain. The starting point was an ontology that models the domain with links to three further modules on terminological, textual and factographic information. Domain experts from the three linguistic communities worked together to come up with a common and consensual conceptualization of the domain. Once the ontology was stable, its concepts were linked to the terms in English, Spanish and Catalan stored in the terminological module. Here the conceptualization is a good example of what we understand as an internationalized domain, reflecting the common view of all the cultures represented in the project. This is a clear example of a localization project on an internationalized domain, affecting only the lexical layer in the ontology. Regarding the interoperability aspect, it is completely guaranteed by the existence of only one ontology, and the ontology can be used in applications that only require labels in one language, or in multilingual applications.

Use Case 2: New to Holland The New to Holland project website concerns an ontology-driven application developed in the Netherlands by the company BeInformed for the Dutch government on informing immigrants, e.g., on the process of applying for an immigration permit. The underlying conceptualization of the New to Holland ontology reflects certain specific characteristics of Dutch immigration procedures that need to be localized in other languages. In this scenario, the ontology is modeling what we have called a culturally-influenced domain and the purpose of localization is to document specifics of Dutch administration services into several other languages. This is therefore clearly a case of localization for documental purposes, i.e., for the purpose of explaining the meaning of concepts and procedures in the language of target users of applications that build on the adapted ontology.

1 http://genoma.iula.upf.edu:8080/genoma/index.jsp
2 http://www.newtoholland.nl
3 see http://www.beinformed.nl
Use Case 3: WordNet related projects (EuroWordNet⁴, Meaning⁵, GlobalWordNet⁶, Kyoto⁷) In
the different projects that have been running since the beginnings of the EuroWordNet project for link-
ing WordNets in different languages to the Princeton English WordNet [11], we come across different
strategies for the construction of the multilingual WordNets. Although WordNet cannot be considered an
ontology in a strict sense, we believe that these projects reflect the difficulties of having to perform a
functional localization of a general lexicon to different target languages. The objective of each lexicon is
to capture the specificities and particularities of each language, while maintaining a considerable degree
of interoperability with the remaining WordNets. Guaranteeing interoperability among WordNets repre-
senting culturally-influenced domains is not a trivial task, in most of the cases the creation of the local
WordNets consisted of reusing the English WordNet and adapting it to the specific needs of each culture
(the so-called Expand Model in [16], followed for instance in the case of the Spanish WordNet in the
EuroWordNet project, or the Japanese WordNet in the current Kyoto project (see [1], [17]). Taking into
account that the type of localization was functional, since each WordNet was going to be used in NLP
tasks in the target cultures, not only the lexical layer, but also the underlying conceptual structure required
modifications to accommodate the specifics of the target cultures. While each WordNet is in principle
independent, interoperability is achieved by creating a mediator ontology (the so-called Interlingua) with
mappings between them.

5. Related Areas

Ontology localization is related to a number of areas. We discuss the relation to the following fields:
software localization, thesauri translation and machine translation.

5.1. Software Localization

Localization is by now a core issue for the software industry where it is heavily related to the notion
of internationalization, i.e. of developing products which can be commercialized world-wide. Esselinck [7]
for example states that “localization revolves around combining language and technology to produce a
product that can cross cultural and language barriers. No more, no less.” In fact, the analogy to localization
in software engineering supports the understanding of ontology localization that we have put forth so
far in this article. Much as the localization of an ontology can affect two layers, i.e. the lexical layer or the
conceptualization itself, localization in software engineering can affect the “surface” of a software product
or the actual functionality and behaviour of the software. Selling software in a certain country (say Ger-
man) requires that the documentation, online help as well as graphical user interfaces are translated into
German. This is what we call ‘surface’ localization as it does not affect the core of the software in terms
of functionality or behaviour. This is similar to the “label translation” aspect of the ontology localization
process. However, it might well be the case that the functionality and behaviour of the software itself has
to be changed to comply with the different processes and rules in place in another country. In this case
we do actually change the functional core of the software as much as we change the conceptualization
to meet the requirements of a given geo-political and cultural environment. The most obvious difference
between software localization and ontology localization is the fact that in the case of software and hard-
ware products that are to be localized, they have previously undergone a process of “internationalization”.
This basically means that those features of products considered “specific to a certain locale” are adapted
to support changes or additions already at production time (e.g. support of international natural language
character sets or addition of functionalities specific to foreign markets). The main reason for this is that
software developers early saw the need to localize their products for international markets, a need that is
now arising in the Semantic Web.

⁴http://www.illc.uva.nl/EuroWordNet/
⁵http://www.lsi.upc.edu/nlp/meaning/
⁶http://www.globalwordnet.org/
⁷http://www.kyoto-project.eu/
5.2. Thesauri Translation

The issue of localization is also crucial in the field of thesauri development. Much as in ontology localization, the goal here is to reuse existing thesauri or create multilingual systems for the purpose of indexing documents across languages. The issue of developing automatic approaches that can reduce the costs in translating thesauri into other languages is also crucial (see [9]). In contrast to ontology localization, the goal in thesauri translation is to find a reasonable translation such that the thesaurus can support the same applications (e.g. indexing of documents) across languages.

A crucial difference is the fact that the issue of changing the conceptualization does not arise in thesaurus translation. In thesauri, the semantics of terms is typically neither formally defined in terms of axioms nor are the hierarchical relations between terms interpreted as strictly as in ontologies, where the subclass-of relation is formally interpreted in terms of (extensional) subsumption. Thesauri are thus typically loosely defined structures in comparison to ontologies, so that the question whether a certain translation has unintended meaning shifts is not as relevant.

5.3. Machine Translation

Machine translation deals with the translation of strings (sentences, documents etc.) from a source language into a target language. Different approaches to machine translation exist, including i) statistical, ii) transfer-based and iii) interlingua-based approaches. Interlingua-based approaches are based on a language-independent universal representation to which texts in the source language are mapped to and from which a translation in the target language is generated [12]. Interlinguas share with ontologies the fact that they are language-independent representations. However, ontologies are typically domain-specific and do not aim to represent a universal language that can be used to translate textual input from a source language into a target language. Prominent interlinguas include for example the Universal Networking Language (UNL) or the Mikrokosmos ontology [10].

Different machine learning techniques are applicable to the problem of translating the lexical layer into a target language. However, as argued above, changes in the lexical layer and in the conceptual layer depend on each other. Thus, off-the-shelf MT techniques have a restricted use here. The task of translating ontology labels actually needs to take into account the whole context and purpose of the ontology. A system that has been a pioneer in this sense is LabelTranslator [4, 5]. LabelTranslator, released as a plug-in of the ontology editor NeOn Toolkit8, is a system created with the aim of supporting a semi-automatic localization of ontologies by providing a functional translation of ontology labels. In its current version it supports the translation from and into English, Spanish and German. The ontological context of each of the labels to be translated will be crucial for finding the most appropriate translation candidate in the target language. In a first step, candidate translations will be obtained from multilingual lexical resources and/or machine translation web services. Then, translation candidate senses are retrieved from different ontology pools, i.e. the system accesses Semantic Web search engines such as Watson to retrieve the senses that correspond to the candidate ontology concepts from different ontologies. In doing this, it not only obtains natural language descriptions of the concept or synonyms, if available, but also its "local context", i.e., the hierarchical graph of hypernyms and hyponyms in which the searched concept is inserted. This permits to compare the ontological contexts of candidate translations with the one of the original label, and perform a ranking to offer the most appropriate translation in each case.

8http://neon-toolkit.org
6. Conclusion

In this note we have proposed a more general definition of ontology localization as “the process of adapting a given ontology to the needs of a certain community, which can be characterized by a common language, a common culture or a certain geopolitical environment”. This definition is more general than previous ones in the sense that it emphasizes that adaptation to a specific language is not the only goal and purpose of the localization activity. We have further characterized the task of ontology localization along three dimensions: degree of internationalization, purpose (functional vs. documental) and degree of interoperability. From these dimensions we have derived different types of localization activity which affect the two layers (conceptualization and lexicon) in different ways, having different inputs and outputs. We have also discussed several real-life scenarios corresponding to different configurations along the above mentioned dimensions. Finally, we have argued that changes to the conceptualization and to the lexicon are clearly not independent from each other but interact in a number of unforeseeable ways that need to be balanced by the person, agent or algorithm performing or supporting the localization.

Ontology localization is per se an important and practical activity with high economic impact as it allows to reuse ontologies engineered for a specific linguistic and cultural community to fit the needs of a different community. In this sense localization is a special type of re-engineering activity inheriting all the known difficulties involved in the task of engineering an ontology. Specific methodologies to adapt an ontology to a different community characterized by a common language, a common culture or a certain geopolitical environment are certainly needed if localization is expected to be performed at a reasonable cost and with high-quality output.

In this sense, the impending need of some international organizations for ontologies that support multilinguality has revealed the lack of methodological support for this activity. This has been the case of the european project NeOn, in which the localization of ontologies has been identified as a crucial activity in the Semantic Web of the future. NeOn has tried to palliate this lack by putting in a great deal of effort in the development of technological and methodological support for this aim. Ontology localization has been integrated for the first time in the NeOn Methodology as one of the activities to be performed in the ontology development process of ontology networks whenever multilingual ontologies are needed, and methodological guidelines have been given with this purpose (see [6]). However, the solutions proposed in NeOn come to solve the specific needs of certain use cases, and there still remains the challenge of extending not only the methodological but also the technological support to cover a wider range of localization scenarios and use cases.

Overall, the aim of this short note has been to shed light on the notion of ontology localization. Future work will have to be devoted to developing new or adapting existing ontology engineering methodologies to the specifics of the ontology localization task as well as to develop tools to support users in this task.

References


