Ontology Localization: Bangla Language

Syed Shahrier Rahman
Associate Professor, Department of Linguistics
University of Dhaka

Abstract

Ontology localization in Bangla is clearly a particular kind of ontology re-engineering movement where the prerequisites (for the re-engineering) are given by the requirements of the Bengali community to which the ontology is adapted. The focal point of this study is that ontology localization can affect two separate layers: the surface and the conceptualization itself. In this paper, I have proposed a more general meaning of ontology localization as the procedure of adapting an offered ontology to the needs of a certain community, which might be described by a normal language, a basic culture or a certain nature's domain. Regarding Bangla language, particular strategies are necessary to adapt for developing ontology localization which is relied upon to be performed at a sensible cost and with great yield.

Keywords: Ontology, Ontology localization, Lexical layer, Conceptualization layer

1. Introduction

The area of ontology localization has been increased throughout the last few decades by contributions from various undertakings. Ontology has been characterized as a formal, precise depiction of thoughts, where ontology localization indicates the accommodation of ontologies to a specific language and culture (Miller, 1990). In this sense ontology localization in Bangla is clearly a particular kind of ontology re-engineering movement where the prerequisites (for the re-engineering) are given by the requirements of the Bengali community to which the ontology is adapted. Bangla ontology localization is along these lines a movement with extremely pragmatic objectives. Therefore, the aim of this paper is to illuminate the thought of ontology localization with a special focus on Bangla. I illustrate how the localization of the diverse layers (lexical and conceptual) communicate and introduce distinctive dimensions that portray the localization procedure. Largely, my objective is to help an improved comprehension of the Bangla ontology localization process.

Ontology specifies concepts and plays important role in inter-cultural knowledge sharing. In fact, ontology is used in the domain of natural language processing as a set of definitions of formal vocabulary. An ontological framework is established on the basis of a community-driven agreement to use a vocabulary. In general, a big quantity of ontology precisely accommodates large number of concepts (e.g., river, mountain chain) of the real world entities (e.g., Padma River, Himalayas). It includes different domains of concepts, where each of them corresponds to an area of knowledge (Gruber,1995). Each domain is organized in facets, where a facet can be defined as a hierarchy of homogeneous concepts describing the different aspects of meaning (Giunchiglia et al., 2009). A partial illustration of a facet is shown in Figure1 (Ganbold et al., 2013).
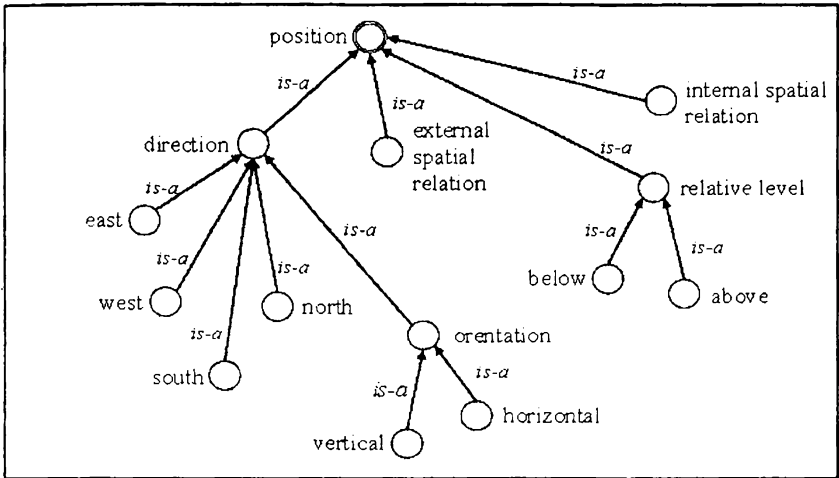
Figure 1: A relational schema of a facet

The figure illustrates the spatial way in which something is located. Leaf nodes of this facet represent relations between entities. For instance, the Bay of Bengal is on the south of Bangladesh. The double circled node is distinguished as the root concept from the rest of the concepts that are represented with the single circle.

The affectivity and importance of such layred patterns have been investigated in the present discussion. Therefore, the proposed paper has been organized as takes after: in Section 2, I have discussed distinctive ontology layers (lexical vs. conceptualization) in detail and how these diverse layers communicate. In Section 3, I have tried present an outline of distinctive magnitudes that could be distinguished in figuring out the sort of localization that is to be performed and the layers that are influenced. Section 4 discusses a number of effective tools which are used to meet the challenge of localization process. Finally, I have made a conclusion of this study.

2. Localizing different layers

Ontology localization maintains a transformational relationship through using ontologies as the input and the output; whereby the

output might be the same ontology as the input, or it can be reached out with extra linguistic characteristics, or it can be yielded as another ontology (Mauricio, 2009). The focal point of this study is that ontology localization can affect two separate layers: the surface, or lexical layer of ontology, and the conceptualization itself. I consider the lexical layer of ontology to incorporate all the levels, definitions and going hand in hand with documentation in natural language that make the ontology human-justifiable. It can recently be anticipated that the lexical layer experience changes, paying little respect to if the target community uses an alternate language or not. The underlying purpose behind this is that any deviation at the conceptualization layer, because of contrasts at the social or geo-political foundation have an effect on the lexical layer.

2.1. Lexical Layer

The surface layer of ontology bears lexical attributes. This lexically marked layer of an ontology embodies: i) the names of the ideas, properties and people characterized in the ontology, ii) natural linguistic descriptions of these entities, and additionally iii) the documentation going with the ontology, which depicts its area and purpose, its use and so forth. The consideration of definitions of ontology entities in natural language require "clarity" paradigm that well defined ontologies need to satisfy (Gruber,1995). It is worth to mention here that the lexical layer is a language-dependent phenomenon and is consequently influenced by any ontology localization process; especially when the adaptation is carried out inside the same linguistic framework. This implies that the differences persuaded by the cultural environment in which the ontology is to be used– be it inside the same linguistic framework or not– is reflected at the lexical layer.

2.2. Conceptualization Layer

While the interpretation of labels is the most critical part of the ontology localization initiative, the conceptualization might

additionally need to be revised, e.g., by an alternate social or geo-political as well as cultural situation. First take an example. An ontology intended to model political capacities and charges in Bangladesh might further recognize the "prodhanmontri (প্রধানমন্ত্রী)" (prime minister) assuming the part of the head of government and the "rashtropoti (রাষ্ট্রপতি)" (president) assuming the part of the head of state. Assuming that I need to utilize ontology about political charges designed for the Bangladeshi geo-political and social environment in provisions that concern also different nations, e.g., the UK or Spain and therefore I require adapting the conceptualization communicated by the ontology. On account of the UK, I might present the class of leader as head of government and the queen as head of state. On account of Spain, I might present the class of "presidente" (president) as head of government and the "monarca" (monarch) as head of state. While one could contend that this adjustment can likewise be accomplished at the lexical level, e.g., by including extra names (prime minister, presidente) for the class "prodhanmontri" or (queen, monarca) for the class "rashtropoti", this is clearly insufficient as these concepts have different extensions and even intensions. Hence, the adaptation to an alternate geo-political and social fact may require more than a 1:1 interpretation, i.e. a change as well in the underlying conceptualization.

It is vital to underscore that the adaptation of the conceptualization layer is basically determined by the inexistence of reasonable equivalents (or ideas with the same granularity level) in the target community, whenever the final purpose of the ontology is to be legitimated in source and target culture. If the concept of "prodhanmontri" serves the capacity of head of the government in Bangladeshi culture, and I aim at reusing the ontology in the UK, I should not make the mistake to stick to the words and translate it as "head minister", only in light of the fact that the word exists in the English language, unless the motivation behind the localization is to paraphrase in English how the Bangladeshi political structure is

composed. A few more examples of such conceptualization of ontologies are given bellow:

Word	Description	Remark
County (প্রদেশভুক্ত বৃহত্তম প্রশাসনিক জেলা বিশেষ)	The largest administrative district within a state; "the county plans to build a new road" (Interglot; 2014)	Not available in Bangladeshi society. Therefore a literal translation or lexical borrowing is needed.
Border post (সীমান্ত চৌকি)	A post or station at an international boundary for the regulation of movement of people and goods (Interglot; 2014)	This ontology is available in Bangladeshi culture.
Mall (বিপণীকেন্দ্র)	Mercantile establishment consisting of a carefully landscaped complex of shops representing leading merchandisers; usually includes restaurants and a convenient parking area; a modern version of the traditional marketplace; "a good plaza should have a movie house"; "they spent their weekends at the local malls" (Interglot; 2014)	This ontology has recently (for last fifteen years) been added to Bangladeshi culture.

Localization infers the presence of an information ontology that is "adapted" to serve the purposes of a different linguistic and/or cultural community. I shall come once again to this in section 3.

2.3. Interface between Lexical and Conceptual Layers

The diverse layers that I have portrayed above do positively collaborate in the sense that changes to one layer cannot be performed independently of the other layers. This implies that changes to the

conceptualization are inescapably reflected at the lexical layer, and changes to the lexical layer might likewise wind up impacting the conceptual layer. The main case in which this communication is not proportional is the point at which the lexical layer experiences alterations without influencing the conceptualization; however this is not feasible in the other direction. Primarily, changes in the conceptualization likewise require the adaptation of the lexical layer. The dependency of modifications to the lexicon on changes in the conceptualization is evidently necessary if the target ontology should have appropriate levels in the target speakers' utterance. Therefore, four effective levels are possible to point out to understand the process. These levels are evaluated by deciding multifaceted design and need of extra examination to set up a complete deciphered ontology with customized characteristics (Sturm, 2002, MSDN, 2012). The following chart (figure 2) is adapted from Sturm (2002) and describe these levels at a glance:
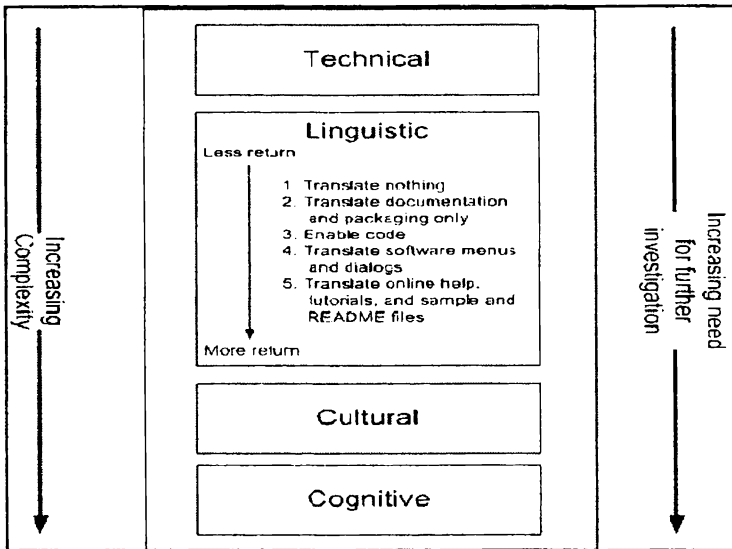

Figure 2: Localization levels

Technical Level: I can start out with the technical level. This level covers all specialized parts of an ontology product. It incorporates the specialized framework and specialized principles utilized in the foreign country the product has to be adapted to (e.g., ISO-standards for character sets, for example, Unicode, ANSI, and so forth.). The adjustment of these issues guarantees that the product lives up to expectations from the specialized perspective and they are the basis for the following level.

Linguistic Level: For a large portion of the technical products, the international adaptations stop here, where distinctive language versions are delivered. The words and texts of the interface and manuals are deciphered and a few viewpoints like punctuation, vocabulary and language structure are exchanged, however regularly without the consideration of social contrasts.

Cultural Level: The third level incorporates the cultural dimension of the utilization of localizing products. It fundamentally covers two areas: the context of the use and the meaning of symbols, graphics, colors, and metaphors utilized as a part of the user interface. The social connection of the product utilizes and its position within the daily life conveys the data concerning the obliged functionality. To place it in straightforward words, cultural localization is concerned about utilization of the icons, metaphors, message conventions, etc.

Cognitive Level: The list of the entailed functions regarding ontology localization yet to figure out the way to present the localized information to the user. The cognitive level hence goes beyond the immaculate importance of interface segments secured by the cultural level. It encases menu structures, necessities, cooperation styles and systems and also essential cognitive techniques utilized within human-computer interaction. This level is without a doubt the most underestimated; however it has an extraordinary effect on the convenience of a technical product.

Considering that the cognitive level includes a cognizant intellectual action which just could be automated with difficulty, I do not consider that this level ought to be part of ontology localization activity. Also, an extraordinary part of the activities performed in this level of localization are typically considered at the time of designing ontology. In the new approach, these levels have been adjusted looking into the layers of the ontology that are influenced in the localization process. The new levels of localization extent from the linguistic adaption of the ontology to a specific language to a cultural adaption of the ontology to a particular geo-political and cultural environment. In a few cases where the lexical layer of the ontology is changed to document the ontology for an alternate language, unintended semantic shifts may occur in case the term picked as a 1:1 interpretation has distinctive connotations in the target community than in the source community. This might be the situation of the term designating "bibaho (বিবাহ)" (marriage) in Bangla or in German (Ehe). In both cases, marriage is characterized as the union of individuals of diverse genders, although the Spanish term "matrimonio", which is the direct translation of marriage and Ehe that I might discover in any dictionary, has a more extensive scope since it includes individuals of distinctive or the same sex. Even the adaptation of the lexical layer may oblige changes in the conceptualization.

Now, consider another phenomenon. It is possible to imagine a geographical ontology planned by speakers of French. French speakers are familier to make distinction between rivers flowing into the sea ('fleuve') and rivers flowing into other rivers ('rivière') into their ontology (Cimiano et al, 2010). This refinement is obviously not a whimsical one, yet it basically demonstrates how the French culture encounters the world. The way that this difference is straightforwardly lexicalized in the French language (rather than other

languages like English, Spanish or German), makes a French ontology engineer inclined to incorporate this refinement into the ontology. The point when limiting this ontology into an alternate language (say English, Spanish or German), an 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 refinement between rivers flowing into the sea and rivers flowing into other rivers in the conceptualization. This implies that there will be no direct lexicalization regarding one designation that might be utilized as label for each concept, but a paraphrase in the target languages.

2. Ignore the differences and keep just the idea 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-building the first ontology by uprooting distinctions that come about because of granularity layers that are not totally shared by the cultures involved. However, there still exists the option of keeping some cultural specificity at the lexical layer by means of powerful linguistic models that have been developed lately to associate linguistic information to ontologies.

The choice will be finalized by assessing whether the real-world distinction between rivers flowing into the sea and rivers flowing into other rivers is a pertinent one, considering the applications that the target ontology is assumed to support. The way that choices at distinctive layers are obviously subject to one another makes ontology localization a challenging and non-trivial endeavor.

3. Magnitudes of Localization

Localization has a useful vitality as it cultivates the reuse of generally conceptualized information in distinctive semantic and social settings. As already outlined in Section 2, the adaptation techniques of localization may have distinctive implications, that is, diverse layers of the ontology will be influenced by the localization to different degrees. I have distinguished the following extents that focus the sort of localization to be performed:

a. Some domains are plainly "internationalized" or "institutionalized" as a side effect of globalization impacts determined by the necessity to trade information on a worldwide level. This is frequently the case in extremely technical domain, however as I have mentioned earlier, most of the domains are significantly influenced by the culture. In the context of Bangla localization, domains like agriculture, geography, administrations etc are notably culture-bound. The resulting models of the same domain in different communities are going to show an important divergence.

b. The objective of the target ontology could be to have the same function in the target community as the original ontology in the source community. Take again the example of ontologies being used inside public administration, agriculture or as well as geography. It 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. Useful localization consequently intimates the production of another ontology on the groundwork of the old one, adapted to the prerequisites of the target community. In documental localization then again, the purpose is just to help the utilization of the original ontology by members of another (linguistic) community. This does not include the making of a totally new ontology, yet just the documentation of the existing ontology in a different language.

c. I may however report the importance of the classes and relations characterized in the ontology displaying an internationalized space in distinctive languages with the goal that it is open by speakers of different languages. The localization is hence influence just the lexical layer of the ontology for this situation. On account of a culturally impacted domain, the fundamental recognizing rule is whether the ontology should be utilized within an alternate geo-political area and nature, in which case the conceptualization needs to be adapted, or the objective is to permit individuals with a different social and linguistic foundation to gain access to and utilize the ontology. The level of interoperability is not a fresh measurement and just influences the instance of the useful localization of a culturally impacted area. Contingent upon the level of interoperability covered (specifically the granularity at which the ontologies requirement to be interoperable), the conceptualization can change pretty much.

4. Reengineering tools

4.1 Concepts

In general, concepts are considered universal phenomenon. Notwithstanding, their representation in human languages fluctuates. Inside the same language a concept could be eluded with various terms (known the concepts *road, track* and *trail* are represented with the same term in Bangla. Whereas the concept *road* stands for 'an open way (generally public) for travel or transportation', *track* stands for 'any road or path (affording as synonymy) and numerous concepts could be alluded with the same term (known as polysemy) *passage* especially a rough one' and *trail* stands for 'a path or track roughly blazed through wild or hilly country'(Interglot, 2014). Bengali speakers commonly add extra adjectives to realize the differences among these concepts. In this case translating them into the target language increases polysemy. Lexical gaps are those concepts that don't have a concise representation in a given language. Then again, they might be communicated as a free combination of words

(Bentivogli & Pianta, 2000). For example, the concept *causeway* – "a road that is raised above water or marshland or sand" (Interglot; 2014) – is a lexical gap in Bangla. As the lexical gap is a feature of the languages, it does happen with all of them. There can be a gap also from the target to source language. For instance, the Bangla word *bheshal* is a gap in English. The word *bheshal* (ভেসাল) can be represented in English as *a triangular liver driven bamboo structure which is used as a fishing trap*. Note that this word fail to offer a brief representation in English. Accordingly, it is treated as gap in the localization process. The river dependent lifestyle of Bangladeshi people is the source of these concepts that are not identical in different English speaking cultures.

4.2 Senses

Localizing ontologies often deal with several numbers of words which bear multiple senses. These senses have subtle difference in meaning. For instance, the word *fissure* has two senses (Ganbold et al., 2013:11)

> [S^1]: *crack, cleft, crevice, **fissure**, scissure – (a long narrow opening)*
>
> [S^2]: ***fissure** – (a crack associated with volcanism)*

The two concepts connected with the given word are hyponyms of *continental depression* and they might be spoken to with the same word(s) in the target language. Polysemous words in the source language may compare to lexical gaps for a subset of senses.

4.3 Synsets

The precise correlations pertinent to a synset depend upon the part of speech combined with the concept the synset stands for. Besides, two diverse kinds of associations are identified, explicitly semantic and lexical. A semantic association is one that subsists between two synsets and that is supposed to apply to all word forms within the synsets. On the contrary, a lexical association subsists between two

.

specific word forms within two separate synsets. A synset also incorporates a brief definition and typically endows with one or more examples of how the word forms in the synset are used. Words in a synset can be straightforwardly translated into the target language. However, in some cases there might be a lack of translation. For example, the synset *street*– "a thoroughfare (usually including sidewalks) that is lined with buildings"(Interglot, 2014) – has 10 words in its description, of which 2 of them (*thoroughfare, sidewalks*) lack translation into Bangla.

5. Conclusion

In this paper I have proposed a more general meaning of ontology localization as the procedure of adapting an offered ontology to the needs of a certain community, which might be described by a normal language, a basic culture or a certain nature's domain. This definition is more general as it emphasizes that adaptation to a particular language is by all account not the only objective and reason for the localization movement. I have further portrayed the undertaking of ontology localization such as degree of internationalization, purpose (functional vs. documental) and degree of interoperability. From these dimensions I have derived different types of localization activity which affect the two layers (conceptualization and lexicon) in different ways, having different inputs and outputs. I have examined a few genuine Bangla language situations relating to diverse designs along the aforementioned measurements. Finally, by examining different examples from Bangla, I have contended that changes to the conceptualization and to the dictionary are plainly not autonomous from one another yet cooperate in various unforeseeable ways that need to be adapted by the individual, executor or calculation performing or supporting the localization. In fact, ontology localization is in essence an essential and reasonable action with high monetary effect as it permits to reuse ontologies designed for a particular linguistic and social community to fit the needs of that community. In this sense, localization is an exceptional sort of re-engineering activity inheriting all the known difficulties involved in

the task of engineering an ontology. Regarding Bangla language, particular strategies are necessary to adapt for developing ontology localization which is relied upon to be performed at a sensible cost and with great yield. In this sense, the approaching requirement of some international organizations for ontologies that help multilingualism has revealed the lack of methodological support for this action. Over all, the aim of this paper was to shed light on the thought of ontology localization and which may also extend to the case of Bangla language as well. Future work is necessary for creating new or adapting existing ontology designing techniques to the specifics of the ontology localization undertaking and also to create devices to help clients in this assignment.

References

Bentivogli, L., & Pianta, E. (2000). Looking for lexical gaps. In *Proceedings of the Ninth EURALEX International Congress,* Stuttgrat, 663-667.

Cimiano, P et al (2010). A Note on Ontology Localization, *Applied Ontology* Vol. 5, No. 2, Amsterdam: IOS Press, 663-669

Ganbold, A., Farazi, F., Giunchiglia, F. (2013). An Experiment in Managing Language Diversity Across Cultures, DISI - Via Sommarive 5 - 38123 Povo , http://disi.unitn.it, Accessed on: 12.08.2014

Giunchiglia, F., Dutta, B., Maltese, V. (2009), Faceted Lightweight Ontologies, *Conceptual Modeling: Foundations and Applications*, Berlin: Springer. Pages 36-51.

Gruber, T.R.(1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6).

Interglot (2014). Translation Dictionary, Retrived 21 July 2014, from http://www.interglot.com/dictionary/en/en/translate

Mauricio Espinoza, Elena Montiel-Ponsoda, and Asunción Gómez-Pérez. (2009). Ontology localization. In Proceedings of the 5th International Conference on Knowledge Capture (KCAP), 33–40.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990), Introduction to wordnet: An on-line lexical database, *International Journal of Lexicography*, 3(4):235–244.

MSDN (2012). Chapter 1, Understanding Internationalization, http://msdn.microsoft.com/enus/library/cc194758.aspx, Accessed on: 18.08.2014

Sturm, C. (2002). Tlcc - towards a framework for systematic and successful product internationalization. In *4th international Workshop on Internationalization of Products and Systems*, pages 61–70, Austin/Texas, USA. Springer-Verlag.