Designing Domain Ontology: A Study in Lexical Semantics

Samir Karmakar

Indian Institute of Technology Kanpur; Kanpur – 208016; India.

samirk@iitk.ac.in

Preparing a multi-purpose lexicon requires a systematic analysis of inter-conceptual relations. These relations are of two types, namely (i) syntactic and (ii) semantic, which can further be decomposed to capture the greater explanatory adequacy. But the exploration of the lexical structure becomes intricate because of the hidden dynamics of the context; since traditionally, language has been viewed as a totality of lexicon and computation system, and major emphasis has been given to the designing of the computational system, considering the designing of the lexicon internal domain ontology as a mere metaphysical game, when in reality it is a serious epistemic concern, because of having the capacity of licensing inferences. Therefore a lexical level representation should have enough scope to incorporate the contextual information.

Designing domain ontology is important since it tells us about the conceptual constellation within the coherent whole of which the related terms are meaningful. Isolating a term from the corresponding constellation will results into the evaporation of meaning. Furthermore it provides the basis, upon which the entire linguistic structure rests. If so, then how is it possible to construct a lexicon, by divorcing the ontological issues? And at the same time, ontology by itself is not enough, again because of the reason that the higher order typifications of those (grounded) concepts and their corresponding interrelations among the types ultimately results into the consequent super-ordinating levels, containing the syntactic information pertinent to a symbol manipulating system.
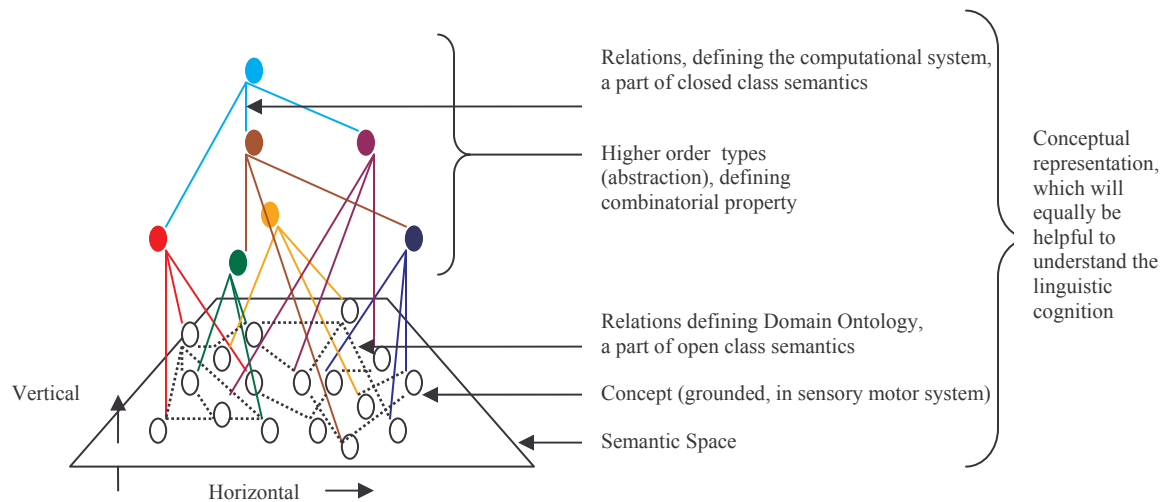
In this paper I would show that the representation of a lexical structure, should include both kind of information which are pertinent to the closed class and as well as the open class semantics, on the basis of examples, cited from English and Bengali.

0. Introduction:

Preparation of a multi-purpose lexicon needs certain attention of theoretical interest, particularly in this age of information technology. Before entering to those theoretical issues one thing I should like to clear that lexicon in no sense is synonymous to the concept of dictionary. The approach to build a lexicon is quite different from that of the dictionary, both in terms of data representation and obviously in terms of goal. While developing a lexicon of a particular language or the languages the task for a linguist is not only to attain the descriptive adequacy of the language function, but also to construct the formalism of meaning conveying mechanism to capture the explanatory adequacy. To proceed further I would like to make a distinction between to levels, namely conceptual and lexical, with an assertion that universalism and as also as

continuation is an property of the conceptual level, in contrast with the fact that lexical level is vulnerable to the linguistic worldview of the language community. The unique theoretical underpinning will be illustrated here in this article, with a paradigm case study.

1. Types of Lexicon:

Lexicons can be classified in two major ways. Each type has its own philosophical prejudices and as also as its own way to model a language. These two major classifications are (i) *sense enumerative model of lexicon* (henceforth SEL) and (ii) *compositional model of lexicon* (henceforth CL) (Pustejovsky 1995). This classification comes into the picture because of the *polymorphic* nature of the languages. The polymorphism arises when word has more than one *sense*.

1. jibaner   natun   *adhyAy*
   (of life)   (new)   (event)

2. baiyer   caturtha   *adhyAy*
   (of book)   (fourth)   (chapter)

Here in these examples, cited from Bengali, the word *adhyAy* has two distinct senses. In (1) it represents the sense of 'event', whereas in (2) it is being used in the sense of 'chapter'. How does one deal such a situation?

1.1 Sense Enumerative Lexicon:

In SEL, whenever the lexicographer faced such a situation, his/her decision is always being guided by a simple dictum. According to SEL, the 'event' expressing *adhyAy* and 'chapter' expressing *adhyAy* would be considered as two separate entries in lexicon, with their corresponding senses. Therefore, in SEL the major claim, which has been made, is that word – sense mapping is one to one and well-defined. Even if a word has multiple senses, the word would be stored as the separate entries along with its corresponding sense. So in Bengali there is two *adhyay*, one means 'event', other 'chapter'. Consider same kind of specification has also been made in the following examples (Jackendoff 1990):

3. *Spatial location and motion:*
   rAmu   pAkhiTA   khnAcAy   *rekhechilo*.
   (Ramu)   (the bird)   (in the cage)   (kept).

4. *Ascription of properties:*
   rAmu   tAr   bouke   sukhI   *rekhechilo*.
   (Ramu)   (his)   (wife)   (happy)   (kept).

5. *Scheduling of activities:*
   rAmu   sombAre   tAr   miTingTA   *rekhechilo*.
   (Ramu)   (on Monday)   (his)   (the meeting)   (kept).

Here in all those above mentioned examples *rekhechilo* is polymorphous, because of having three distinct senses. To deal with all these polymorphic cases, SEL has followed the following dictum (Pustejovsky 1995):

A lexicon *L* is a *Sense Enumeration Lexicon* if and only if for every word *w* in *L*, having multiple senses $s_1,...s_n$ associated with the word, then the lexical entries expressing these senses are stored as $\{w_{s1},...w_{sn}\}$.

But why polymorphism, even after being the linguistic reality, has not been answered in this approach. The most probable answer is not incorporating the issues of *context sensitivity*. I would address this issue in our next section.

1.2 Compositional Model of Lexicon:

In the recent development of semantics, the concept of context earns more and more concern, in both the traditions, namely logical (Peregrin 1999; Saba 2005) and linguistic (Jackendoff 1990; Pustejovsky 1995; Langacker 1991, 1999). Unlike the traditional practices, now the meaning is being conceived as the resultant of the underling complex interactions of the contextual forces. In this tradition, the basic assumptions are as follows:

- Each and every lexeme has a *default meaning*, because of occupying a particular position within the corresponding *ontology*.
- The default meaning of a lexeme has a rich *internal structure*. Lexical representation of this internal structure should capture the ontological status of the lexeme.
- In a polymorphic situation, which meaning would enjoy the default status, is purely determined by the *frequency of occurrences*, and hence *probabilistic*.
- Linguistic polymorphism should be explained in terms of *type shifting* of the default meaning, under the contextual influences.

In this system, any kind of sense-variation is explained in terms of the type shift due to the contextual constraints. Therefore, the claim is that *adhyAy* has a default meaning representation, which gets coerced into the sense of 'event' and 'chapter', under the influence of the *semantic participants* in the *syntactic construction*, stated above in example (1) and (2) respectively. Same is also true in case of (3), (4) and (5).

One point I should emphasize here that *being default* doesn't necessarily entail that it doesn't have any kind of context. Rather in our next section we will see that even the default form also has an *inbuilt* context, and meaning is all about context, even though the context is a vague concept, because of the multiplicity in interpretation.

2. Lexeme Internal Structure: The Functioning of Ontology

Among the sense variations, some are *privileged* because of being default type. But this default type also has some inbuilt context, which is conceptual in nature, hence need *grounding*. Ontology serves this purpose of conceptual grounding.

In ontology, lexemes are being sorted out into different classes, on the basis of their common characteristics; these classes are again grouped into some other subclasses, on the basis of the same principles. As a consequence, greater the distance from the *basic level*, towards *super-ordination*, lesser would be the contextual constraints, for the sake of unity, resulting into the absolutely *underspecified* top-most concept of *entity*. And in the opposite direction, moving toward the *subordinate* level adds the contextual constraints, resulting into the specified sense. Let me put a folk ontology of my Bengali language, to elaborate the issues. **Obviously, this one**

is not a complete ontology, nor even the most efficient one; but I dare with a little hope that it will serve our further purposes to understand the notions of context, semantic participants, type shift, and finally the concept of default meaning. One point should be cleared here that moving away from the *entity*, along the direction of the arrow, will lead us to subordinate levels and finally to the basic levels. Contrariwise, moving towards the *entity*, leads us to the super-ordinate levels.
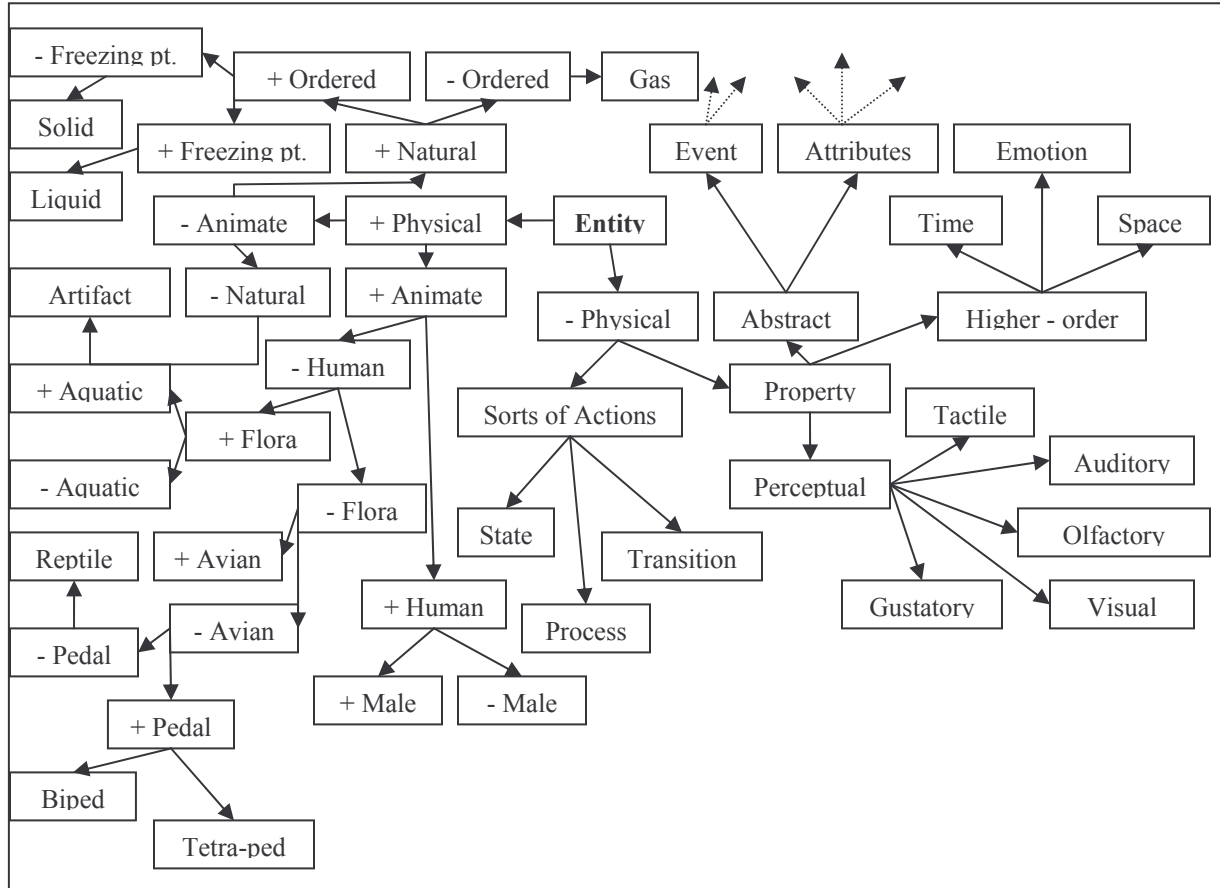

Figure 1: Ontology of the native Bengali speaker

The first major classification of entity is *physical* and *non-physical* concepts. Because of being perceptually grounded, physical concepts are more basic among the conceptually grounded concepts. When a Bengali native speaker faced a word, say *dudh* 'milk', the ontology acts as a filter, which extract information, relevant to the language user. According to the present ontology, these information have two major aspect, that is *physical* and *non-physical*. The physical side of the story tells us about the very nature of the constituent out of which it is made up of. And the non-physical information again are of two types, namely *perceptual* and *abstract*. Perceptual information tell us about its color, taste, shape, size etc. Finally abstract property tells us about its edibility, sweetness etc. After getting filtered, only the relevant concepts of this ontology remain prominent, others *blacked out* (fig. 2). The prominent feature node along with the connections becomes the inbuilt default context for *dudh*, the rest of the region of the semantic space constitutes the *ground* for this inbuilt default concept. The concept of ground is important here, since at the time of type shifting coercion affects the corresponding ground of type, under the intervention of the other contextual forces imposed by the semantic participants in a syntactic construction. We will discuss these issues in proper place.
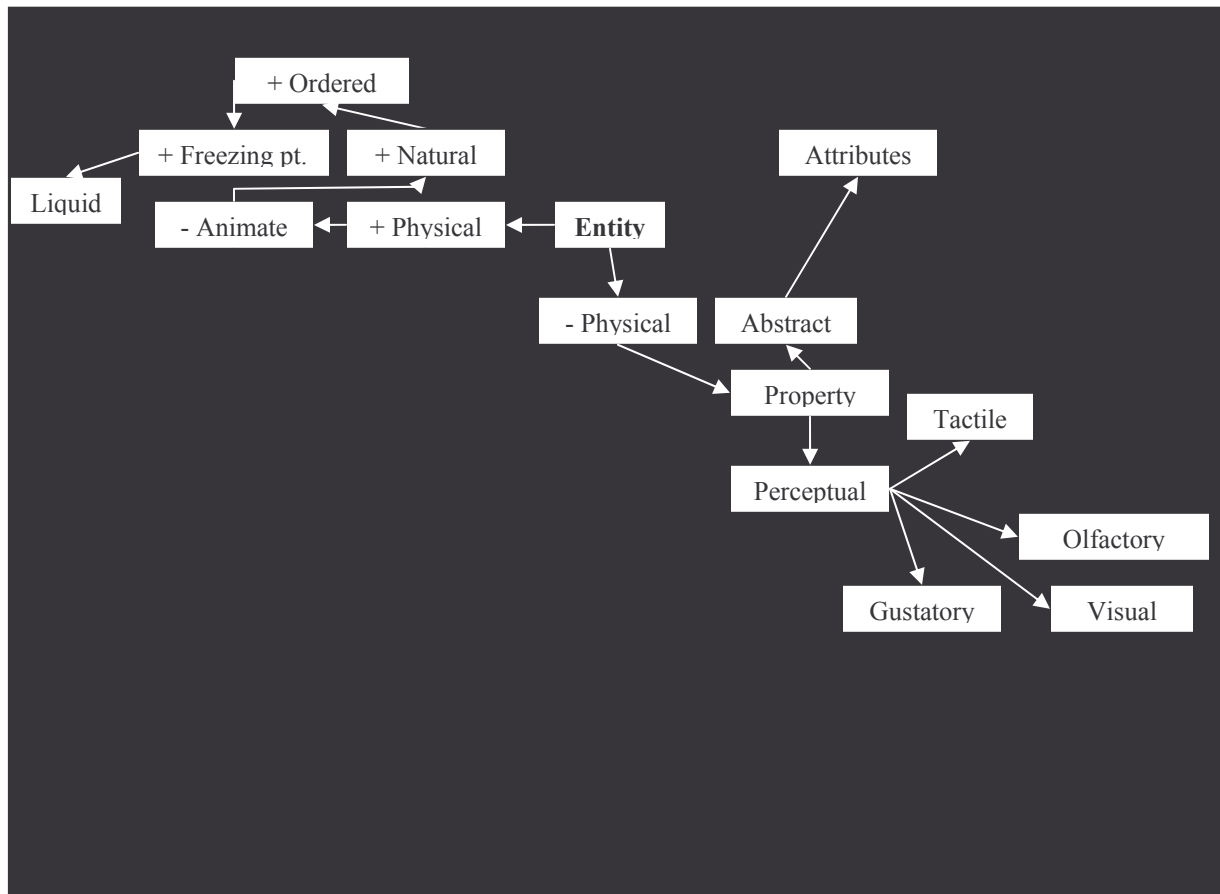
Figure 2: Explaining inbuilt context and grounding of *dudh* 'milk'

2.1 Conceptual Grounding: Relations defining domain ontology, a part of open-class semantics

In our previous section we have seen what we actually mean by inbuilt context and grounding of the perceptually grounded basic concepts. In this section we will try to extend the same scheme in case of defining *adhyAy*. It will also clear certain other aspect of ontology. The default meaning of adhyAy is stated in (6) and (7).

> (6) *adhyAy* = the ordered part of a whole, containing communicative content

The above definition of *adhyAy* implies that the corresponding meaning is the emergent property of the constellation of the other concepts' inbuilt contexts and their resultant effect. Therefore, without invoking the ground it is quiet impossible to have the meaning. Or, more perspicuously, dissociation of any lexical expression from the corresponding semantic space ultimately results into the evaporation of its semantic content. Actually the concepts, forming the ground, provide the background knowledge, which can also be labeled as *scope* (Langacker 1999).

Conceptual grounding, not only help us in case of differentiating the finer-grained sense analysis, but also it helps in case of differentiating the homophonies (ref. 7-8). Homophonous forms are contrastively ambiguous, whereas the other cases what we have discussed still now are complementarily ambiguous (Pustejovsky 1995).

7. (a) tini takhon sAphalyer cURA theke abAdhe *paRchilen*
    (he) (then)   (of success) (peak) (from) (freely)   (was falling).

  (b) tini   takhon   boi   *paRchilen*
    (he)   (then)   (book) (was reading).
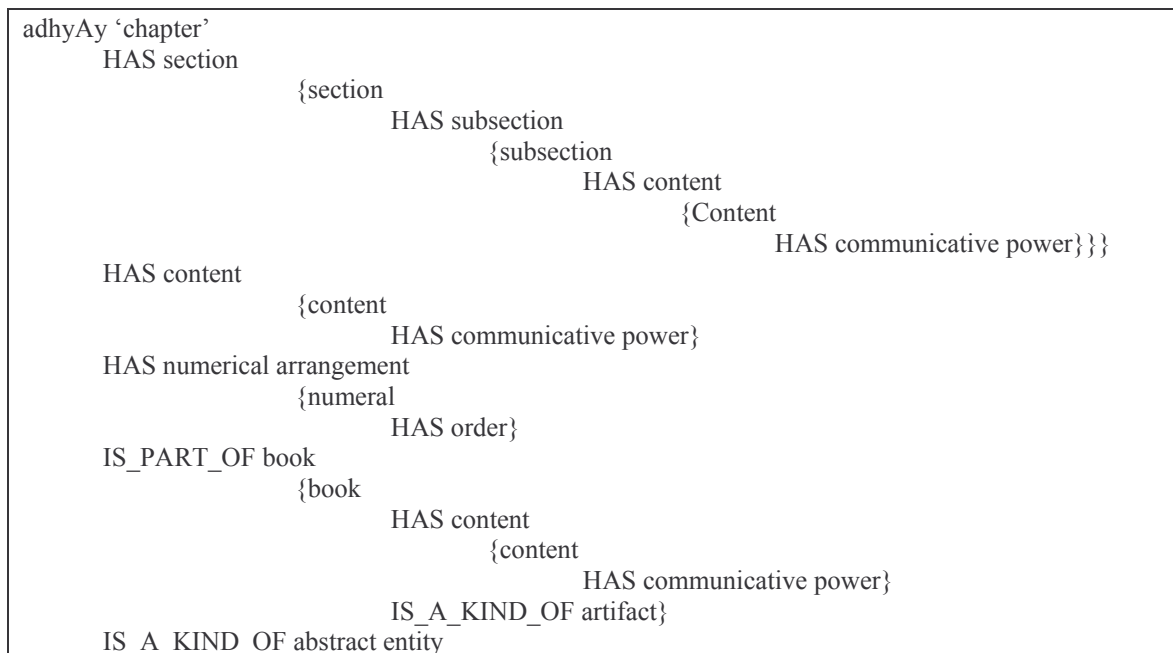
8. (a) gAch   bharA   *kul*
      (tree)   (full of) (jujube)

  (b) *kUl*   barAbar   janapad
      (bank)   (along)   (human habitation)

In all the above mentioned examples, the sense disambiguation becomes possible only because of the reason that their corresponding inbuilt context, which determines the semantic participants and at the same time is also determined by them. Not only these few issues but also the creative use of the lexical expressions can also be explained under the concept of meaning generation, *if and only if a well structured ontological space of Indian languages can be constructed.*

2.2 Ontological computation: Closed class relations

One can classify context in three ways, namely (i) inbuilt context, internal to the lexeme, (ii) scope and (iii) dormant context. First two contextual information are important in case of lexical representation, and inter-lexical relations. In this section the emphasis will be given to explore the meaning relations, namely (i) hyponymy, (ii) hypernymy, and (iii) meronymy. Instead of giving any explicit definition, here these relations will be discussed in the light of our paradigm case.

```
adhyAy 'chapter'
      HAS section
                      {section
                              HAS subsection
                                      {subsection
                                              HAS content
                                                      {Content
                                                              HAS communicative power}}}
      HAS content
                      {content
                      HAS communicative power}
      HAS numerical arrangement
                      {numeral
                              HAS order}
      IS_PART_OF book
                      {book
                              HAS content
                                      {content
                                              HAS communicative power}
                              IS_A_KIND_OF artifact}
      IS_A_KIND_OF abstract entity
```

In above mentioned schema, the meaning relations have been shown, in capital letter. These relations provide us three distinct kind of information. (i) HAS relation is hyponymic, (ii)

IS_A_PART_OF is mereological and (iii) IS_A_KIND_OF is hypernymic. *adhyAy* is represented in the following figure along with these relations, which shows the different dimensions of its meaning, however, once again is not complete. But here incompleteness is not the drawback, since we all are aware that the understanding of the meaning of a particular lexical expression needs the lighting up of the partial ontological relations which are immediately important. I would like to name this partial lighting up as the *locality principle*. Locality principle is related with the *concept proper* whereas the inferences are the logical expansion of this concept proper region.
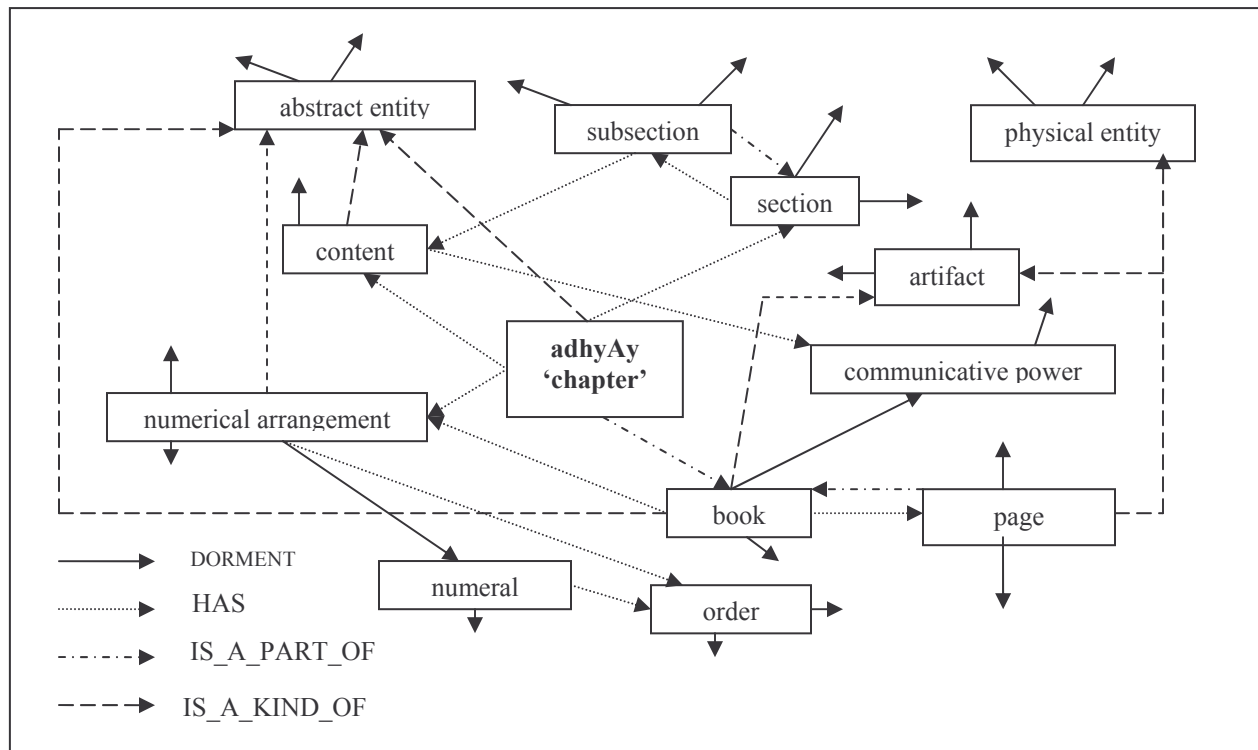
3. Conclusion:

In this article I have argued that language is essentially polymorphic, however, this fact is not at all recognized in the traditional approaches to lexicography. Once we accept the linguistic polymorphism as an essential feature of language function, it becomes quite crucial to talk about the designing of the domain ontology. Designing domain ontology is not a metaphysical game rather it is an ambitious effort to model the context. In this write up, meaning has been viewed as a case of contextual constraints, in terms of grounding, type, and inbuilt context. Furthermore a gesture has also been taken, with the help of a paradigm case, how the domain ontology plays a crucial role in case of semanticality judgment.

References:
Bates, M. 1995. "Models of Natural Language Understanding". In *Proc. Natl. Acad. Sci.*, Vol. 92, pp. 9977-82.
Chomsky, N. 1995. *Minimalist Program*. The MIT Press.
Jackendoff, R.
    - 1990. *Semantic Structure*. The MIT Press.
    - 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford.

Langacker, R.W.

    - 1991. *Concept, Image and Symbol: The Cognitive Basis of Grammar*. Mouton De Gruyter.

    - 1999. *Grammar and Conceptualisation,* Mouton.

Masolo, C. et. al.2003. *The Wonderweb Library of Foundational Ontology, Preliminary Report*. Italy.

Peregrin, J. 1999. ''Pragmatisation of Semantics'', in Turner, K. (ed.) *The Semantics/Pragmatics Interface from Different Points of View*, Elsevier, 419-442.

Pustejovsky, J. 1995. *The Generative Lexicon*. The MIT Press.

Saba, Walid S. 2005. *Language Logic and Ontology – Uncovering the Structure of Common Sense Knowledge*, 2005. Available at: <http://www.aut.edu/agora/ontology.pdf>.