==========
Language in India www.languageinindia.com **ISSN 1930-2940 Vol. 17:5 May 2017**
==========

Visual Onto-thesaurus for Tamil

Dr. Rajendran Sankaravelayuthan & Dr. Anandkumar, M.

==========

Abstract

Tamil Visual Onto-thesaurus (TVOT or simply VOT) is an outcome of an extensive research activity that went on in the field of lexical semantics of Tamil. It went through several stages before being culminated into Tamil visual onto-thesaurus. It depicts our travel from Tamil thesaurus to Tamil word net. It is a lexical resource which amalgamates all sorts of information available in a dictionary, thesaurus and wordNet. The Dravidian wordNets (in which Tamil wordNet is one of the four components) built under the IndoWordNet project depended on an ontology developed by Western conceptualization of the world found in English). This has not taken into consideration the Indian conceptualization of the world depicted in the *nikhandu* tradition. Say for examples, *nikhandus* have classifications such a six types of tastes, nine types of planets (*gragams*), 7 types of *mandalams* (a type of division), 15 *tidis* (15 phases of moon), etc. which are crucial for Indian tradition. In the western oriented WordNet ontology there is no scope for the visualization of concepts depicted in *nikhandus*. Moreover building a wordNet based on Hindi wordNet which in turn is built on English wordNet will take many years to complete and it would miss the conceptualization depicted in Indian tradition. Apart from this the extension approach of building Tamil wordNet using Hindi wordNet cannot fulfill Dravidian conceptualization. A merger approach of building separate wordNets and collapsing them into one would have been a preferable approach. The present visual onto-thesaurus is based on the Indian and Dravidian conceptualization and the process of building one is comparatively very simple. We have the plan to mend it into a generic one so that all the Dravidian languages can be easily accommodated into it.

Key words:

Aristotle, Aristotelian principle, lexical semantics, thesaurus, paper thesaurus, Visual Onto-thesaurus, wordNet, Tamil wordNet, *nikhandu*, *nikhandu* tradition, Dravidian conceptualization, ontology, natural language processing, hyponymy, hyponym, homonymy, homonym, polysemy, hypernymy, hypernym, holonymy, holonym, troponymy, troponym, taxonymy, taxonym, entailment, superordinate term, lexical hierarchy, hierarchy, compatibility, incompatibility, opposition, antonym, antonymy, hierarchical classification, semantic field, semantic network, unique beginner, generative lexicon, linguistic issue, computational issue, congruence relation, lexical relation, lexical inheritance, meronymy, meronym,

taxonomic hierarchy, hierarchy, meronymic hierarchy, entity, entities, events, abstracts, relationals, component, semantic component, componential analysis.

1. Introduction

A paper thesaurus for Tamil was prepared in 1990 based on the principles of componential analysis of meaning propounded by Nida (1975), Indian tradition of *nikhandu* and Aristotelian principle of genera and species and was published in 2001 (Rajendran, 2001), nearly after a decade. Following the paper thesaurus, an Electronic thesaurus for Tamil was attempted and a book on Tamil electronic thesaurus was published in 2006 (Rajendran and Baskaran, 2006). The preparation of wordNet for Tamil was undertaken (2001-2003) with the financial assistance from Tamil Virtual University (renamed now as Tamil virtual academy) and a crude version of it based on the ontology developed by Rajendran (Rajendran, 2001) was submitted to the institute in 2003. After that, from 2009 onwards with the fund received from MHRD and Department of electronics and information Technology of Govt. of India the building of Dravidian wordNets were executed based on Hindi wordNet; nearly 30000 synsets (concepts) have been completed. Still we have a long way to go to achieve the desired target. At present a team from CEN, Amrita University is involved in building onto-thesaurus for Tamil as a part of the project entitled "Computing Tools for Tamil Language teaching and learning". The project is funded by Tamil Virtual Academy, Chennai.

The principles of ontology and various kinds of ontologies and ontological applications are elaborately discussed in this paper as the present VOT has taken into its fold many of the ideas discussed in them.

2. Ontology

The subject of ontology is the study of the categories of things that exist or may exist in some domain. The word ontology is from the Greek *ontos* for being and *logos* for word. One classical dictionary definition of Ontology is: "The branch of metaphysics that deals with the nature of being". It originated with Aristotle's effort to classify things in the world. Ontology is an explicit specification of a conceptualization. As a term borrowed from philosophy, ontology means a systematic account of 'existence'. AI systems deliberates that what "exists" can be represented in a program code. They try to represent knowledge of a domain in a declarative formalism; the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. We can describe the ontology of a program by defining a set of representational terms. In such ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and

formal axioms that constrain the interpretation and well-formed use of these terms. Formally, ontology is the statement of a logical theory (Smith, 2003).

Ontologies prove to be extremely useful in the representation of lexical knowledge. Their renewed interest in lexical semantics and natural language processing (NLP) can be attributed to this use. The meaning of a lexical item is partly determined by the position in the ontology occupied by the concept or concept it expresses. Representing one of the meanings of a word minimally implies (i) distinguishing it by other senses the same word might have, (ii) capturing certain inferences which can be performed from it, and (iii) representing its similarity with the meaning of other words (Busa, et al. 2001: 31). Ontological representation is useful for this purpose. For instance, given the word *mouse*, a proper although minimal representation of its meaning requires distinguishing the sense of 'small rodent' from the one of 'small pointing device for computers'. Moreover, the same representation should be able to capture the fact that being a rodent entails being a mammal, as well as the fact that the sense of *mouse* as 'small rodent' shares with the meaning of other words such as *dog*, or *cat*, the fact of being subtypes of mammal. Ontologies are therefore powerful formal tools to represent lexical knowledge, exactly. The word meanings can actually be regarded as entities to be classified in terms of the ontology types. In this perspective, a given sense can be described by assigning it to a particular type. The ontology structure will then account for entailments between senses in terms of relations between their types. The sharing of the same ontology type can be attributed to the resemblances between word senses (Busa et al 2001: 31).

2.1. Principles of ontology

Hyponymy and its consequence taxonomy are the fundamental building blocks of ontology. Hyponymy and its natural partner, incompatibility, are described by Lyons (1977) as “the most fundamental paradigmatic relations of sense it terms of which the vocabulary is structured”. Lyons (1977) states that taxonomic lexical hierarchies are structured by the relations of hyponymy and incompatibility. The relation of hyponymy imposes a hierarchical structure upon the vocabulary and upon particular fields within the vocabulary; and the hierarchical ordering of lexemes can be represented formally as a tree diagram (Lyons, 1977: 295). It is hard to conceive of any language operating satisfactorily in any culture without its vocabulary being structured in terms of the complementary principles of hyponymy and contrast (Lyons, 1977: 300). Meronymic or partonomic relations are ontological relations that are considered as fundamental as the ubiquitous, taxonomic subsumtions relationship (Pribbenow, 2002: 35). There are numerous lexemes in the vocabularies of languages whose meaning cannot be specified independently of some part-whole relation of sense (Lyons, 1977: 314).

Taxonomy is usually only a hierarchy of concepts (i.e. the only relation between the concepts is parent/child, or sub-class/super-class, or broader/narrower), but in an ontology, arbitrary complex relations between concepts can be expressed too (X married to Y; or A works for B; or C is located in D,

etc.). Although taxonomy contributes to the semantics of a term in a vocabulary, ontologies include richer relationships between terms. It is these rich relationships that enable the expression of domain specific knowledge, without the need to include domain-specific terms.

Ontology has a richer internal structure as it includes relations and constraints between the concepts. Ontology claims to represent a certain consensus about the knowledge in the domain. This consensus is among the intended users of the knowledge, e.g. doctors using a hospital ontology regarding a certain disease, artists relating to historical art and so on. Word vocabulary and ontology are often used interchangeably. But a more strict definition is that a vocabulary is a collection of terms being used in a particular domain that can be structured (e.g. hierarchically) as a taxonomy. This taxonomy when combined with some relationships, constraints and rules, form the ontology. A combination of ontology together with a set of instances of classes constitutes a knowledge base as given below (Breslin et al, 2009: 58):

Vocabulary + structure = Taxonomy

Taxonomy + Relationships, constraints and rules = Ontology

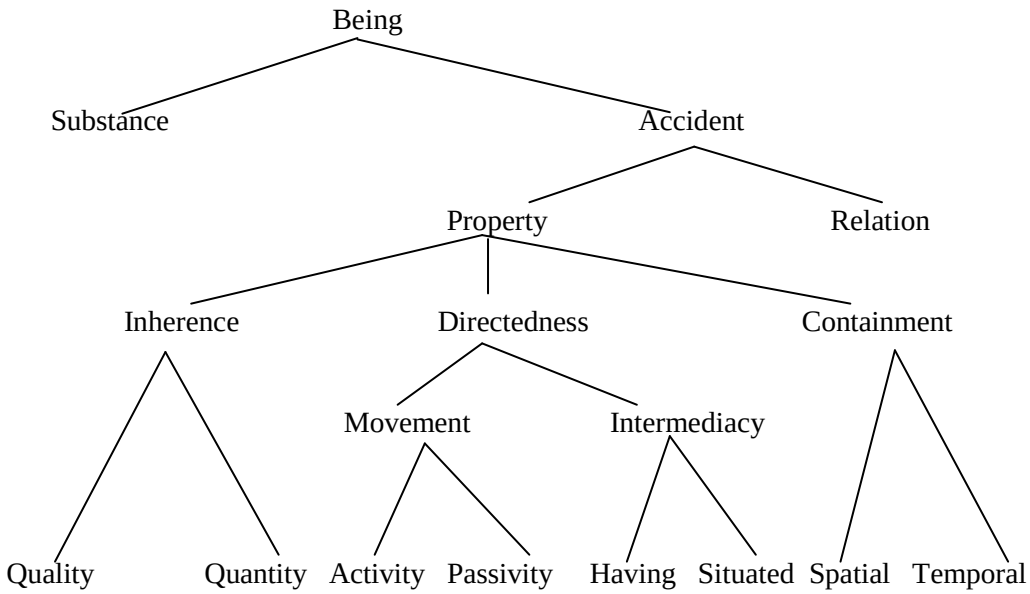
Ontology + instances = Knowledge base

2.2. Ontology of Aristotelian Origin

Aristotle instigated the history of scientific taxonomy. It is predicated on a first philosophy of essentialism. Aristotle's work on natural history and logic laid out taxonomic principles. Taxonomy or the division of things into genera and species is a way of classifying predicates in the logic; it is a refinement of the ten basic categories of predicates. The basic idea was developed in detail in the long tradition of Aristotelian logic. The branching tree diagrams that specified the various genera and species of the category are founded on the classification of the various categories, say, of substance. The taxonomic notions of genus and species were developed in Aristotle's natural history, as opposed to his logic, to handle relations between things, primarily animals (Slaughter, 1982: 15).

Aristotelian philosophy is founded on qualitative, as opposed to quantitative, differences. Aristotle's notions of classification have two sources: the first is in the logical works where he lays down the general theory of classification; the other is in the biological writings where he discusses the problems arising in the classification of animals. There is evidence which indicates that Aristotle's early biological studies were instrumental in his developing taxonomic as logical concepts (Slaughter, 1982:27). The process by which genera are distinguished into species is called logical division. Aristotle has provided the first division, the summa genera, in his ten categories, and within the category of substance, his cosmology and biology were but further elaborations of division. Aristotle offered an ontology which included 10 categories, shown as the leaves in the tree shown in figure 1 (from Sowa, 2009, after Brentano).

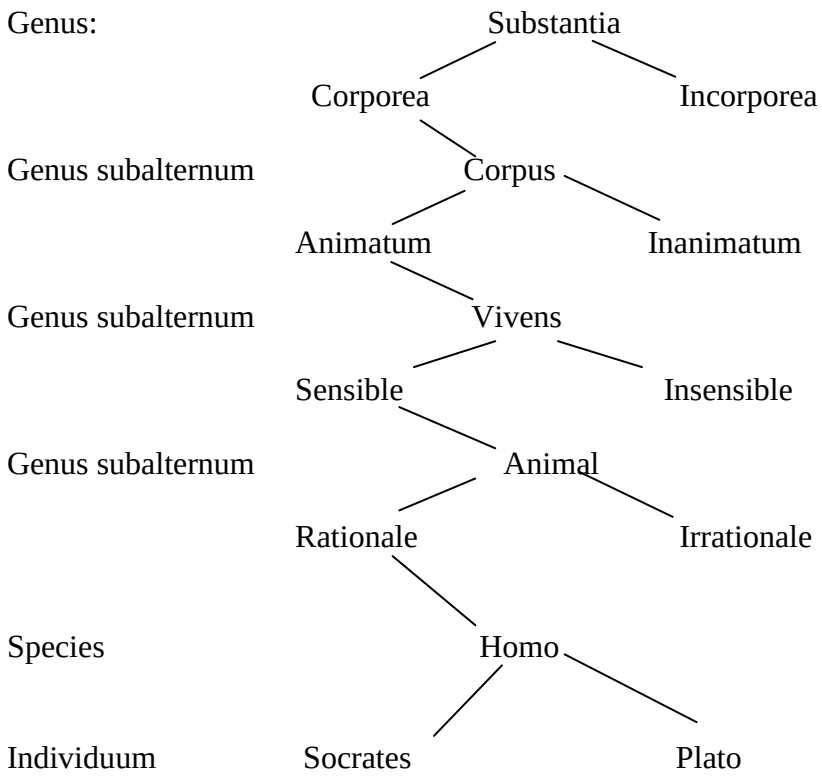
Figure 1



Aristotle did not extend this division to develop a hierarchical arrangement but his commentator Porphyry did and the Porphyrian tree of hierarchically linked genera and species became canonical in the tradition of Aristotelian logic. The tree provides a logical classification of the category of substance as Figure 2 shows (Slaughter, 1982: 29):

Logical Classification of the Category of Substance

Figure 2:



We find in Aristotelian logic books further divisions in each category and the philosophical and scientific lore of the age is included in these divisions and tables by the sixteenth and seventeenth centuries. The examples of Bunddeville (figure 3) and Du Moulin (figure 4) are typical of the encyclopedic nature of these logics (Slaughter, 1982:39).

Figure: 3

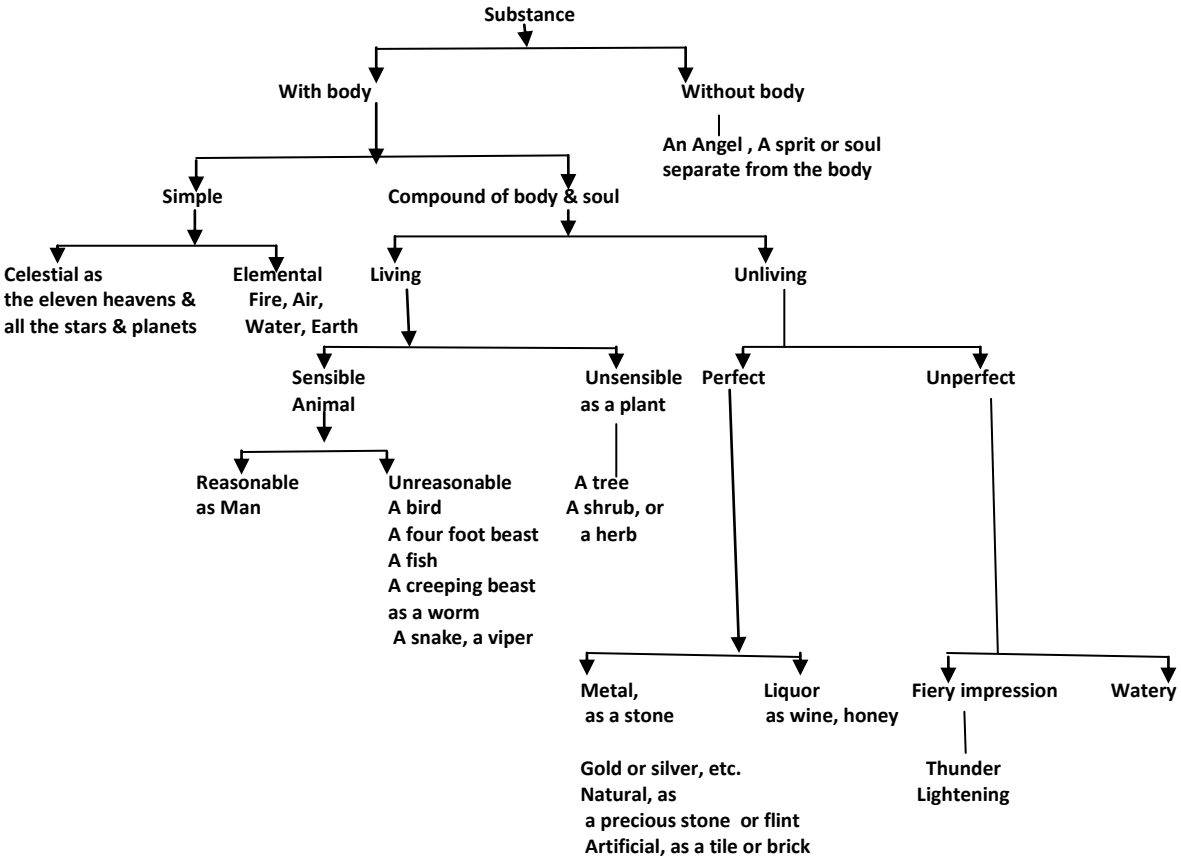
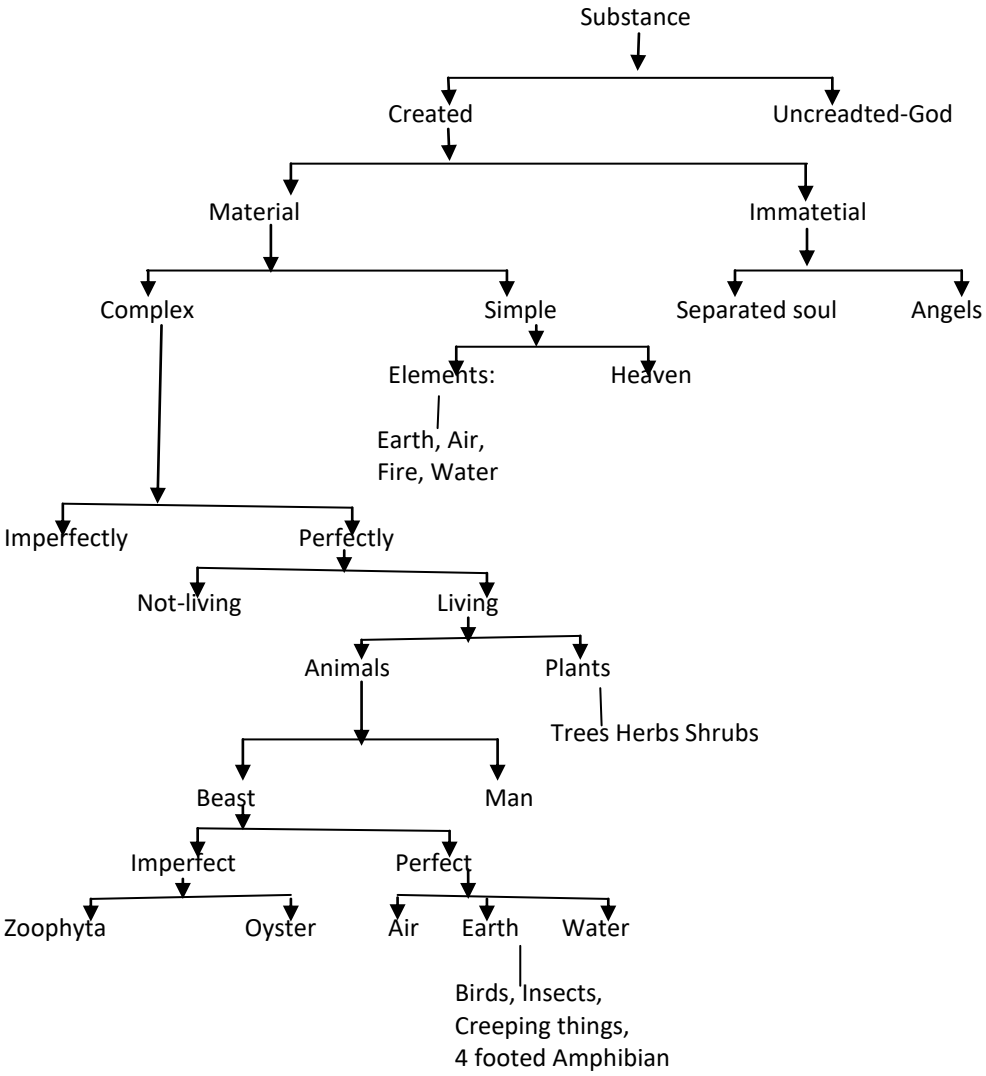

Substance
With body
Without body
An Angel , A sprit or soul
separate from the body
Simple
Compound of body & soul
Celestial as
the eleven heavens &
all the stars & planets
Elemental
Fire, Air,
Water, Earth
Living
Unliving
Sensible
Animal
Unsensible
as a plant
Perfect
Unperfect
Reasonable
as Man
Unreasonable
A bird
A four foot beast
A fish
A creeping beast
as a worm
A snake, a viper
A tree
A shrub, or
a herb
Metal,
as a stone
Liquor
as wine, honey
Fiery impression
Watery
Gold or silver, etc.
Natural, as
a precious stone or flint
Artificial, as a tile or brick
Thunder
Lightening

Figure 4



In scientific sense**,** the Great Chain of Being cannot be called as a taxonomy; rather it can be considered as a predecessor of scientific taxonomy and as a highly elaborated folk taxonomy. Originally, the Aristotelian world view retained and popularized in the concept of the Great Chain of Being. It incorporated all phenomena of nature – everything embraced hierarchically from the heavenly planets to the lowliest worm and the mud in which it burrowed. Mechanistic philosophy made inroads on the higher reaches of the heavens but for the most part it left untouched the world of living creatures (Slaughter, 1982:38).

Aristotle had posited that animate and inanimate natures are two fundamentally different things. Decontextualization and universalization of the words or concepts were attempted. This led to the

development of a scientific (botanical) taxonomy. The following levels of taxa are found (Slaughter, 1982:55):

1. Unique Beginner: e.g. plant, animal
2. Life form: e.g. tree, bush, flower, weed, fern
3. Intermediate: this is an unstable category that manifests itself during a period of adjustment in the taxonomic system and then disappears when a settled, adjusted system is re-established
4. Genus: pine, oak, masterwort
5. Species: ponderosa pine, black oak
6. Variety: northern ponderosa pine, swamp white oak

A Comparison of Berlin's folk taxonomy with a fully developed hierarchy of specialized taxa will reveal two ways of classification of things (Slaughter, 1982: 55).

Figure 5

FOLK TAXONOMY	SCIENTIFIC TAXONOMY
Unique Beginner	Kingdom
Life Form	Phylum
	Class } Specialized taxa
	Order } Specialized taxa
	Family } Specialized taxa
(Intermediate)	
Genus	Genus
Species	Species
	Sub-species
Variety	Variety

2.3. Ontology in Thesaurus

A thesaurus in its widest contemporary sense is a classification of words by concepts, topics or subjects; it need not contain synonyms, and the fact that some items in a given class are synonymous is coincidental. A synonymous dictionary, by contrast, deals with word groups, and does not constitute a conceptual classification the system of kind exemplified by Roget's Thesaurus of English Words and Phrases (1852). This is a crude distinction: thesaurus frequently contains synonyms, and synonymous dictionaries are thesauric, if the groups in them are large and cross-referenced and reflect a wide interpretation of synonymy (Jones, 1986). The following is the plan of classification adopted in Roget's thesaurus (Mawson, 1956).

Figure 6:: Plan of Classification in Roget's' Thesaurus

Class	Section	Sub section
ABSTRACT RELATIONS	I. Existence	
	II. Relation	
	III. Quantity	
	IV. Order	
	V. Number	
	VI. Time	
	VII. Change	
	VIII. Causation	
SPACE	I. In General	
	II. Dimension	
	III. Form	
	IV. Motion	
MATTER	I. In General	
	II. Inorganic	1. Solids
		2. Fluids
	III. Organic	1. Vitality
		2. Sensation
INTELLECT	I. Formation of Ideas	
	II. Communication of ideas	
VOLITION	I. Individual	
	II. Intersocial	1. In General
		2. Possessive relations
AFFECTIONS	I. In General	
	II. Personal	
	III. Sympathetic	
	IV. Moral	
	V. Religious	

2.4. Ontology in *nikhantu* Tradition

The earliest work for information on lexical items in Indian as well as Tamil tradition is *nikhantu* which is comparable to the thesaurus tradition of western community. *Nikhantus* try to give semantic information on vocabulary of a language in a thesauric or classificatory format. The trend was started in

Sanskrit and spread to other Indian languages. Tamil easily adopted Sanskrit oriented *nikhantu* compilation and came out with a number of works.

2.4.1. A*marakosha*

The Amarakosha is a thesaurus of Sanskrit written by the ancient Indian scholar Amarasimha. It is the oldest extant *kosha*. The *Amarakosha* consists of verses that can be easily memorized. It is divided into three *khandas* or chapters. The first, *svargadi-khanda* ("heaven and others") has words pertaining to gods and heavens. The second, *bhuvargadi-khanda* ("earth and others") deals with words about earth, towns, animals and humans. The third, *samanyadi-khanda* ("common") has words related to grammar and other miscellaneous words.

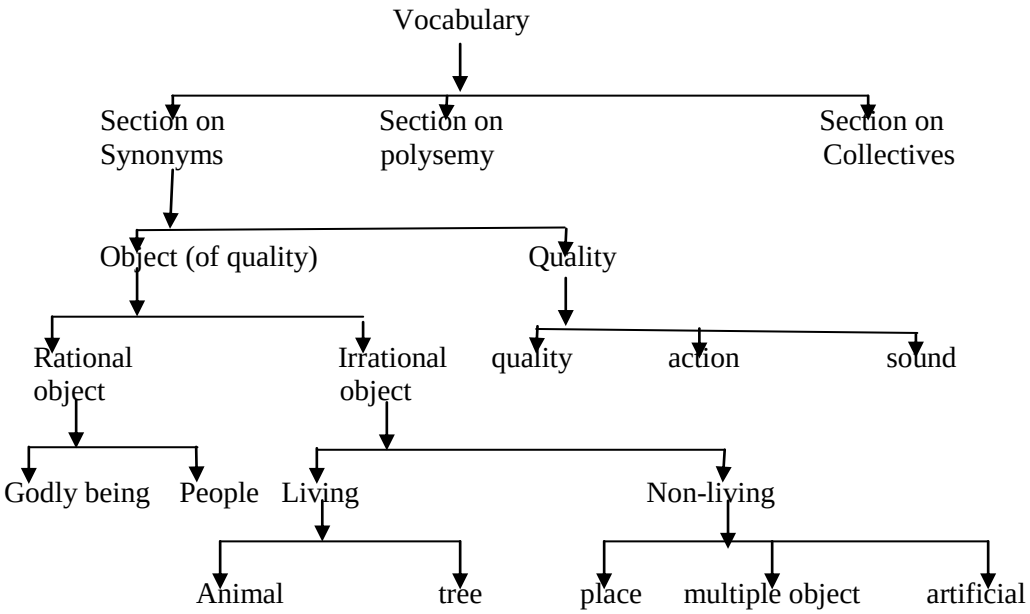
2.4.2. *tivaakaram nikaNTu*

Attempts to classify vocabulary of a language are found in *nikhaNTus*, which lay foundation for the compilation of thesauri or thesaurus dictionaries. The *tivaakaram nikaNTu* (*tivaakarar*, 1958) of early Tamil *nikaNTu* tradition classify words into twelve sections/chapters.

1. Chapter on god names
2. Chapter on human names
3. Chapter on animal names
4. Chapter on tree names
5. Chapter on place names
6. Chapter on multiple object names
7. Chapter on artificial form names
8. Chapter on quality names
9. Chapter on action names
10. Chapter on sound names
11. Chapter on polysemous names
12. Chapter on collective names

This can be classified in the following fashion.

Figure 7



2.4.3. ***naamatiipa NikaNTu* (1810)**

The following classification is found in *ndaamatheepa nikaNTu* (*cuppiramaNiyak* 1930) of Tamil.

Figure 8

Rational creatures: 1. Divinities
2. Human Beings

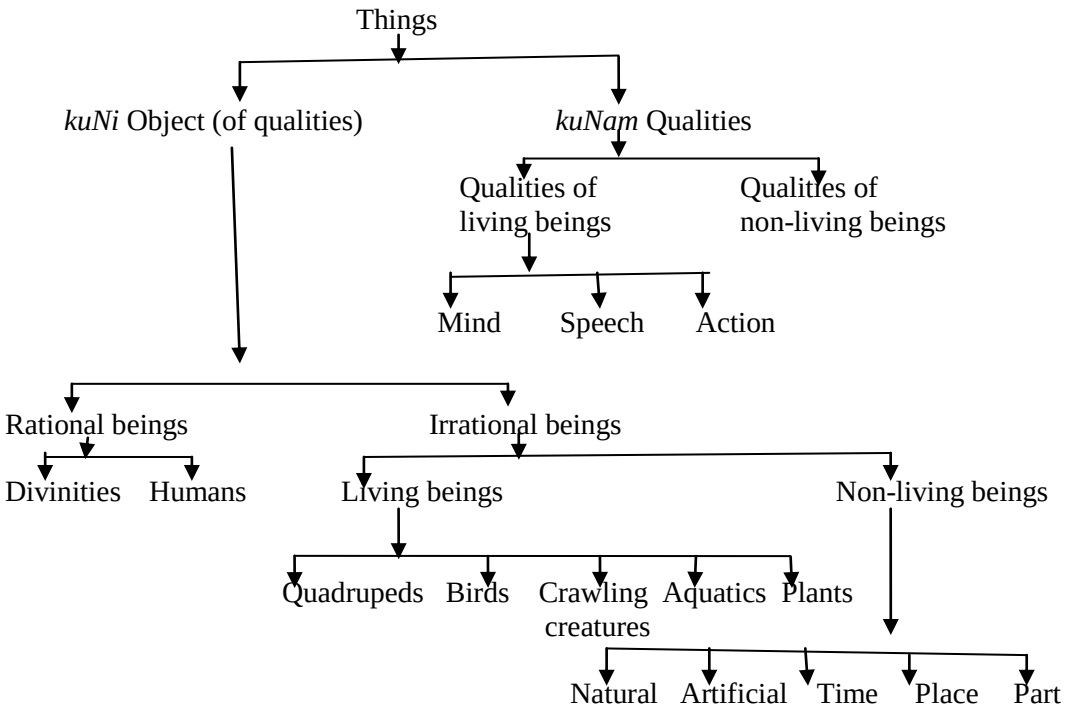
Irrational creatures: 1. Quadrupeds
2. Birds
3. Crawling beings
4. Aquatics
5. Plants

Irrational non-living beings: 1. Natural things
2. Artificial things
3. Place
4. Time
5. Part

Qualities: 1. Living creatures: mental qualities
2. Living creatures: communicative qualities
3. Living creatures: qualities of action
4. Qualities of non-living beings

The classification can be restructured as follows:

Figure 9



2.5. Ontology in Nida's thesauric dictionary

Nida (1975) who was concerned with the preparation of a thesauric dictionary for Greek gives the following as the tentative hierarchical classification of the referential meanings or lexical concepts (Nida1975:178-186). His is a componential approach to meaning. He elaborately discusses about the foundation of his theory classification in his work entitled "Componential Analysis of Meaning: An Introduction to Semantic Structure" (Nida 1975) He has classified the lexical concepts under four categories: entities, events, abstracts and relationals. The following is the outline of his design (Nida, 1975: 178-186).

I. Entities

A. Inanimate

1. Natural

a. Geographical

b. Natural substances

c. Flora and plant products

2. Manufactured or constructed entities

a. Artifacts (non-constructions)

b. Processed substances: foods, medicines, and perfumes

c. Constructions

B. Animate entities

1. Animals, birds, insects

2. Human beings

3. Supernatural power or beings

II. Events

A. Physical, B. Physiological, C. Sensory, D. Emotive, E. Intellection, G. Communication, G. Association, H. Control, I. Movement, J. Impact, K. Transfer, L. Complex activities, involving a series of movements or actions

III. Abstracts

A. Time, B. Distance, C. Volume, D. Velocity, E. Temperature, F. Color, G. Number, H. Status, I. Religious character, J. Attractiveness, K. Age, L. Truth-falsehood, M. Good-bad, N. Capacity, O. State of health, etc.

IV. Relationals

A. Spatial, B. Temporal, C. Deictic , D. Logical, etc.

This classification is based on referential meanings and it is not possible to obtain one to one correspondence between the semantic domain of classes and the grammatical classes. A paper thesaurus and electronic thesaurus are prepared by Rajendran based on the classification given by Nida (1975).

2.6. Ontology in Semantic Fields

Ontology can be related to semantic fields. Aspects of semantic fields presume that the vocabulary of a language is structured in accordance with the structures of grammar and phonology of a language. The words of a language can be classified into sets which are related to conceptual fields and divide up the semantic space or the semantic domain in certain ways. The works of German linguists of a half century ago and that of American anthropologists have led to the development of field theories (Lehrer, 1974: 15). Trier was most influential among the German linguists. Trier distinguishes between lexical and conceptual fields. The lexical field divides the conceptual field into parts, like a mosaic. A word acquires its meaning by its opposition to its adjacent words in the pattern. Field theories are suitable for analysis of some sets of words and unsuitable for others. Trier believed that linguistic fields are not isolated, but rather that they "join together to form in turn fields of higher order, until finally the entire

vocabulary is included" (Lehrer, 1974: 17). Whether or not a progressive synthesis of small fields into larger ones is semantically enlightening is an open question. There is evidence for the view that semantic structures can be looked at in a variety of ways.

Lyons defines the meaning of a term as a function of its relationship to the other term in the lexical field, and the relationships (synonymy, antonymy, class inclusion, incompatibility, etc.) are primitive in his theory (Lehrer, 1974: 22). Synonymy can be defined as a bilateral implication: A and B are synonyms if A $\supset$ B and B $\subset$ A. Class inclusion is unilateral implication. A $\supset$ B, where B is higher in the taxonomy than A, but it is not the case that B $\subset$ A. Class inclusion is the taxonomic relationship 'kind of' relationship, and this is one of the most basic and important notions in the taxonomy. A $\supset$ B, where B is higher in the taxonomy than A, but it is not the case that B $\supset$ A. The highest term in the taxonomy has been called by various names: head word, cover word, superordinate word or archlexeme. Occasionally there is no appropriate head word for a taxonomy, although people make use of a number of devises to fill this gap. Overlap is not permitted in a true taxonomy. Strictly speaking the following chart is not taxonomic (Lehrer. 1974: 24).

Colour				
red	orange	yellow	green	blue
tangerine	Gold	chartreuse	aqua	

But one finds structures that are hierarchically arranged but with some overlap, except in a relatively few domains. One finds overlap even in biological classification. Word contrast in taxonomy is incompatible. If A and B are incompatible, then A $\supset$ not B and B $\subset$ not A. In the chart given below, dog, cat, horse and sheep are incompatible Lehrer. 1974 (: 24).

Animal			
Dog	Cat	Horse	sheep

Although the concepts of synonymy, incompatibility, and class inclusion are fairly clear in most cases, there are areas of overlap and many words are in borderline.

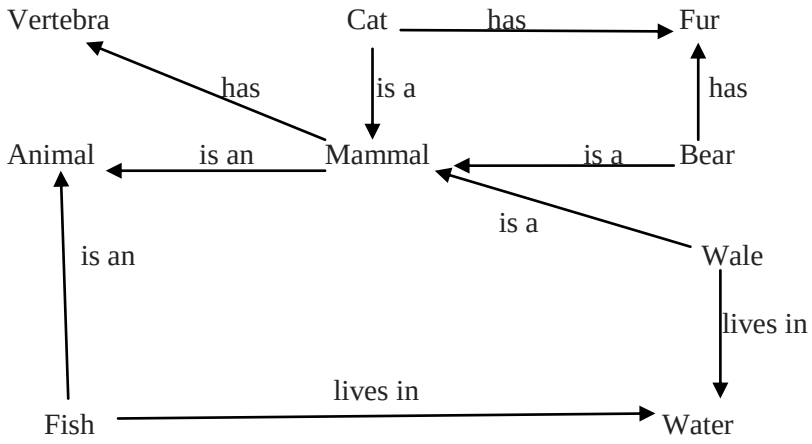
The field theory provides a good model for deciding what to look for and what to describe when dealing with sets of words that are obviously closely related. Sets of words show different types of patterns. For example cooking words, kinship terms and colour terms show different types of patterns exhibiting different types of relations between the words.

2.7. Ontology in Semantic Networks

One of the important applications of ontologies is semantic networks. A semantic network or net is a graphic notation of representing knowledge in patterns of interconnected nodes and arcs (Sowa, 1984: 76). Computer implementations of semantic networks were first developed for artificial intelligence and machine translation, but earlier versions have long been used in philosophy, and linguistics. What is common to all semantic networks is a declarative graphic representation that can be used either to represent knowledge or to support systems for reasoning about knowledge. Some versions are informal, but other versions are formally defined systems of logic.

A semantic network, or frame network, is a network that represents semantic relations between concepts. This is a kind of knowledge representation. It is a directed or undirected graph consisting of vertices, which represent concepts, and edges, which represent semantic relations between concepts. The following example will exemplify the semantic network (Wikipedia on Semantic network).

Figure 10



2.8. Ontology in WordNet

There is a claim from others (not by the creators) that wordNet itself is an ontology. WordNet, (Miller et al 1990) is an example of a semantic network. WordNet is sometimes called an ontology, a persistent claim that its creators do not make. It is an online lexical database initiated for English first. [It is enhanced into Euro-wordNet (Vossen, 1998) consisting of interconnected wordNets of European languages. Tamil WordNet, a component of Dravidian Wordnets (Rajendran 2009, Rajendran et al 2010) which in turn are connected with Indo-wordNet (Bhattacharyya, 2010) which in turn is connected with English or Euro-wordNet is under preparation.] It is a sort of amalgamation of thesaurus and dictionary.

It groups words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. The WordNet by its nature turns to be an ideal lexical accessing system as it links concepts with another concept by multifarious meaning relations. WordNet not only links one concept with another concept through semantic or meaning relations, but also captures the contextual meaning variations of a particular word i.e. the polysemy of a word. The four major syntactic categories (Noun, Verb, Adjective, and Adverb) are treated separately. Nouns are organized in lexical memory as topical hierarchies. Verbs are organized by a variety of entailment relations. Adjectives and adverbs are organized as N-dimensional hyperspaces (Miller et al, 1990).

Some of the most common semantic relations defined are meronymy (X is part of Y, i.e. Y has X as a part of itself), holonymy (Y is part of X, i.e. X has Y as a part of itself), hyponymy (or troponymy) (X is subordinate of Y; X is kind of Y), hypernymy (X is superordinate of Y), synonymy (X denotes the same as Y) and antonymy (X denotes the opposite of Y) (Miller et al, 1990). WordNet properties have been studied from a network theory perspective and compared to other semantic networks created from Roget's Thesaurus and word association tasks. From this perspective the three of them are a small world structure. (Wikipedia on WordNet under Semantic network).

2.8.1. Unique Beginners in WordNet

Based on the hierarchical principle it is possible assume that all nouns are arranged in a single hierarchy. But in WordNet the nouns are separated into several hierarchies, each with a different unique beginner. Relatively distinct semantic fields, each with its own vocabulary, can be correlated to these multiple hierarchies. Unique beginner corresponds roughly to a primitive semantic component in a compositional theory of lexical semantics (Miller, 1998:29).

List of 25 unique beginners for noun source files of EuroWordNet

{act, activity}	{natural object}
{animal, fauna}	{natural phenomenon}
{artifact}	{person, human being}
{attribute}	{plant, flora}
{body}	{possession}
{cognition, knowledge}	{process}
{communication}	{quantity, amount}
{event, happening}	{relation}
{feeling, emotion}	{shape}
{food}	{state}

{group, grouping}	{substance}
{location}	{time}
{motivation, motive}	

These hierarchies vary widely in size and are not mutually exclusive – some cross-referencing in required – but on the whole they cover distinct conceptual and lexical domains. WordNet's nouns are contained in the twenty-five component files. The following is the diagrammatic representation of hyponymic relations among seven unique beginners denoting different kinds of tangible things (Miller 1998).

	{plant, flora}
{living thing, organism}	{animal, fauna}
	{person, human being}
{thing, entity}	{natural object}
	{artifact}
{non-living thing, object}	{substance}
	{food}

The verbs in WordNet are grouped under 15 semantic domains listed below (Miller et al, 1990; Fellbaum, 1998):

1. Verbs of bodily functions and care (Ex. sweat, shiver, faint, etc.)
2. Verbs of change (Ex. change, etc.)
3. Verbs of communication (Ex. stammer, appeal, bet, teach, creak, etc.)
4. Competition Verbs (Ex. fight, etc.)
5. Consumption Verbs (Ex. drink, etc.)
6. Contact Verbs (Ex. hit, scrub, wipe, etc.)
7. Cognition Verbs (Ex. infer, guess, assume, etc.)
8. Creation Verbs (Ex. engrave, print, etc.)
9. Motion Verbs (Ex. gallop, race, fly, swim, etc.)
10. Emotion or Psych Verbs (Ex. amuse, charm, etc.)
11. Stative Verbs (Ex. surround, cross, etc.)
12. Perception Verbs (Ex. watch, spy, etc.)
13. Verbs of Possession (Ex. have, rob, bestow, auction, etc.)
14. Verbs of Social Interaction (Ex. impeach, franchise, excommunicate, etc.)
15. Weather Verbs (Ex. rain, thunder, snow, hail, etc.)

The division of the verb lexicon into semantic domains not only gives one a grip on organizing a large amount of data, but is also necessitated by the absence of a single root verb or "unique beginner" that could head the entire verb lexicon. Within a single semantic field it is frequently the case that not all verbs can be grouped under a single unique beginner.

2.8.2. Organization of Lexical Items in WordNet

In wordNet the four major syntactic categories (Noun, Verb, Adjective, and Adverb) are treated separately. Nouns are organized in lexical memory as topical hierarchies, verbs are organized by a variety of entailment relations, and adjectives and adverbs are organized as N-dimensional hyperspaces. (Miller et al, 1990). The nominal synsets are related to one another by the meaning relations hyponymy-hypernymy, meronymy-holonymy and antonymy. The verbal sysets are related to one another by the meaning relations troponymy and entailment. The adjectives are related mainly by the relation antonymy is a unique fashion.

2.9. Ontology in Generative Lexicon

Pustejovsky (1995), the messiah 'generative lexicon', characterizes a generative lexicon as a computational system involving at least the following levels of representation (Pustejovksy, 1995:61):

1. ARGUMENT STRUCTURE: Specification of number and type of logical arguments
2. EVENT STRUCUTE: Definition of the event type of an expression and its subeventual structure
3. QUALIA STRUCTURE: A structural differentiation of the predicative force for a lexical item
4. LEXICAL INHERITANCE STRUCTURE: Identification of how a lexical structure is related to other structures in this type of lattice

Qualia structure is generally understood as representational tool for expressing the componential aspect of word meaning (Pustejovsky, 1995). The basic vocabulary relies on qualia structure for structuring the semantic/conceptual types. Unless there is a way of structuring word meaning along such multiple dimensions, the model would fail to provide an appropriate semantics for them. Qualia structure specifies four essential aspects of word's meaning (or qualia) (Pustejovsky, 1995:76):

- FORMAL ROLE- provides the information that distinguishes an individual within a larger set. It expresses the ISA relation which applies to all categories of the language.
- CONSTITUTIVE ROLE- expresses a variety of relations concerning the internal constitutions of an entity or event.
- TELIC ROLE- expresses the typical function of an entity, the purpose for carrying out an event, i.e., what the entity is for.
- AGENTIVE- expresses the origin of an entity, or the coming into being of a property or of an event.

Qualia structure is generally understood as a representational tool for expressing the componential aspect of word meaning.

The qualia structure is the core of the generative properties of the lexicon, because it provides a general strategy for creating increasingly specific concepts with conjunctive properties. A simple schematic description of a lexical item, α, using this representation is shown below (Pustejovsky, 1995):

Figure 11

α

ARGSTR = ARG1=x

...

CONST = what x is made of

QUALIA = FORMAL = what x is

TELIC = function of x

AGENTIVE = how x come into being

The lexical structure for *puttakam* 'book' as an object can then be represented as follows (Pustejovsky, 1995: 116):

Figure 12

puttakam 'book'

ARG1= x: *takaval* 'info(rmation)'

ARGSTR = ARG2= y: *pautikapporuL* 'physobj'

takaval. *pautikappaoruL* 'info.physobj_lcp'

FORMAL = *koNTiru* 'holds' (y,x)

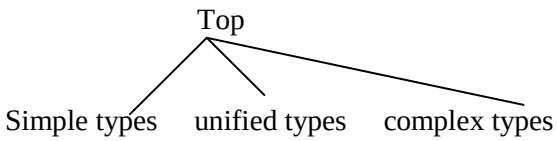
QUALIA = TELIC = *paTi* 'read' (e,w,x.y)

AGENT = *ezutu* 'write' (e', v, x, y)

The above representation of *puttakam* 'book' reveals the following facts: *puttakam* 'book' is a physical object made of papers bounded together into an object; it contains information there by distinguished from notebook; it is written by somebody who is the author of the book and published by somebody who is the publisher of the book.

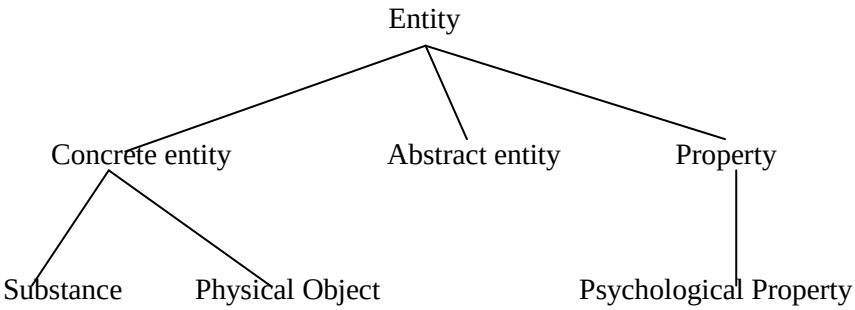
Lexical items are organized in the lexicon in terms of tripartite structure distinguishing between simple types, unified types, and complex types at the top level (Busa et al, 2001: 338).

Figure 13



Intuitively, the type of a lexical item is simple if the item is uniquely defined in terms of a taxonomic relation to another entity in a hierarchy. Natural kind entities are such an example. Similarly, "information," "abstract," are also members of the set of simple types when they do not involve richer information concerning other aspects of meaning (Busa et al, 2001: 338).

Figure 14

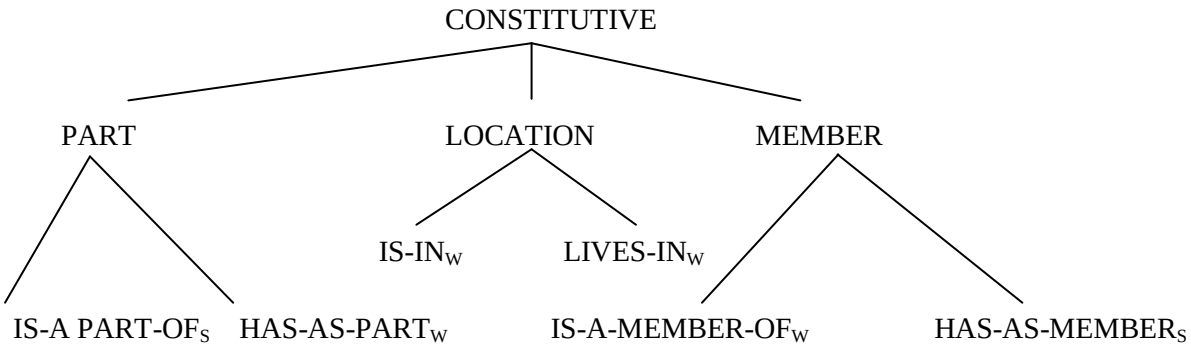


Unified types implement the principle of orthogonal inheritance, which allows a lexical item to fall into multiple classes along different dimensions of meaning. A unified type is created by recursively combining a simple or another unified type with additional elements from qualia structure (Busa et al, 2001).

When using a semantic vocabulary for structuring a large number of word meanings, it turns out that there are lexical items that share, on the surface, the same structural properties (e.g., they may involve TELIC role), but they differ in their linguistic behaviour. For each element in the qualia set we also distinguish between strong functional types and weak types. Weak and strong qualia determine whether or not orthogonal components of meaning give rise to simple or unified types. (Busa, et al., 2001)

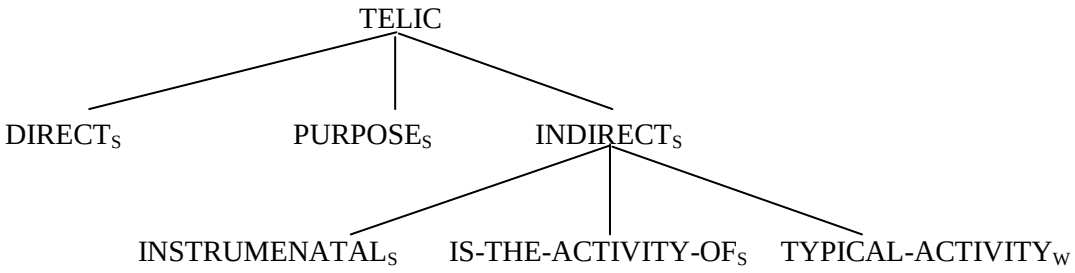
The qualia set for the CONSTITUTIVE role involves a number of subtypes, which expresses different constitutive relations that contribute to the semantic description of various concepts (Busa, et al., 2001: 341).

Figure 15



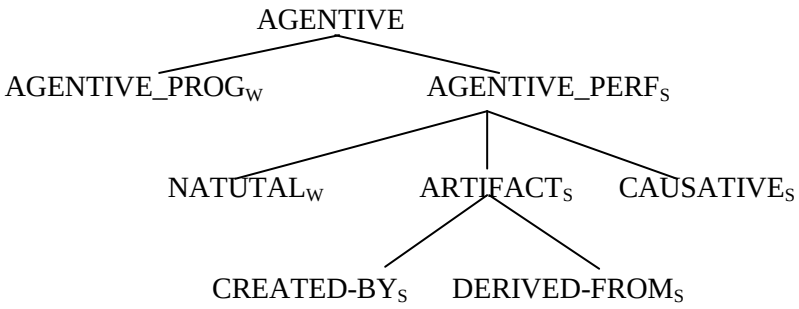
The TELIC involves a set of subtypes that distinguish lexical items according to a number of parameters. The first is discussed in Pustejovsky (1995) and concerns the distinction between DIRECT and INDIRECT TELIC. The PURPOSE TELIC is one associated with verbs, expressing the goal of the agent for performing a given action (Busa et al, 2001: 342).

Figure 16



The AGENTIVE qualia set distinguishes between persistent and temporary properties of the event encoded therein. AGENTIVE_PROGRESSIVE (AGENTIVE_PROG) is distinguished from AGENTIVE_PERFECTIVE (AGENTIVE_PERF). AGENTIVE_PERF is further distinguished for natural kinds, artifactual entities, and for causation involved in complex structures (Busa et al, 2001: 342).

Figure17



3. Tamil Visual Onto-thesaurus

Thesaurus is a in its wider sense is a classification of words by concepts, topics, or subjects. The present Tamil Onto-thesaurus is the extended version of Electronic thesaurus of Tamil focusing more on the ontological features. Two kinds of issues arise in the preparation of Tamil onto-thesaurus:

- Linguistic issues
- Computational issues

3.1. Linguistic Issues

It involves mainly the following four tasks:

1. Developing ontology for Tamil based on structural semantic principles.

2. Establishing semantic domains and sub domains based on distinguishing semantic or componential features of lexical items.

3. Classifying Tamil vocabulary to fit into the ontology developed.

4. Linking words by various semantic or lexical relations such as synonymy, hyponymy-hyperonymy, meronymy-holonymy, compatibility, and incompatibliity.

3.2. Computational Issues

It involves mainly the following three tasks:

1. Conversion of linguistic data base into computer accessible format.
2. Preparation of a tool to provide the facilities for augmenting, entering and editing the raw data, and classifying the lexical items in a semi-automatic way.
3. Creation of user friendly interfaces for accessing the onto-thesaurus in simple manner.

3.3. Ontology of Tamil Vocabulary

The ontology available in Rajendran (1982, 2001), which is founded on the theory of componential analysis of meaning propounded by Nida (1975a), Indian tradition of *nikhandu* and

Aristotelian principle of genera and species is enhanced to suit the present purpose. Following Nida (1975a) the vocabulary of Tamil is grouped initially into four domains: entities which consist of referential meanings of concrete concepts, events which consist mainly of verbs and verbal nouns and abstracts which consist mainly of adjectives and adverbs apart from abstract nouns and relationals which consists of functional words including postpositions, connectives and coordinators. These four domains are further hierarchically classified into sub-domains under which the lexical items are listed. One will be able to capture the meaning of the concerned lexical item from the domain to which it belongs in a hierarchical fashion. A full-fledged detail of building ontology for the vocabulary of a language is available in Nida (1975a).

3.4. Structuring of Vocabulary by Lexical Relations

Lexical semantics offers foundation for structuring vocabulary in terms of lexical relations (Lyons 230-335, Cruses 1986. In the NLP oriented papers, the general practice are to avoid giving linguistic details based on which the system is built. But here we would like to give the lexical semantics of building onto-thesaurus to make it more transparent.

3.4.1. Congruence Relations

There are four basic relations between classes that furnish a model for establishing the fundamental group of sense relations and for defining a set of systematic variants applicable to virtually all other paradigmatic sense relations; they are identity, inclusion, overlapping, and disjunction (Cruse, 1986:86-87).

Identity: class A and class B have same members.

Inclusion: class B is wholly included in the class A

Overlap: class A and class B have members in common but each has members not found in the other

Disjunction: class A and class B have no members in common

These four congruence relations culminate into the four lexical relations discussed below.

3.4.2. Lexical Relations

There are at least four lexical or meaning relations by which lexical items can be linked or related to one another in the ontological structure of Tamil vocabulary. They are synonymy, hyponymy, compatibility and incompatibility (Lyons, 1977; Cruse, 1986:84-111). A word acquires its referential meaning in being a member of a semantic domain by the common features it shares with other members in that domain, and having contrasting features which separate it from other members of the domain. It is the semantic relations among words, such as synonymy, hyponymy, compatibility and incompatibility, which help one to classify and organize words in terms of semantic domains in a structural fashion.

Synonymy: e.g. *puttakam* 'book': *nuul* 'book';

Hyponymy-hypernymy: e.g. *pacu 'cow' : vilangku* 'animal';

Meronymy-holonymy: e.g. *uTal* 'body': *kaal* 'leg';

Compatibility: e.g. *cellappiraaNiv* 'pet': *naay* 'dog'

Incompatibility: *malai* 'hill': *maTu* 'water hole'; incompatibility leads to the relation called opposition which culminates into many types which are discussed below.

Nida (1975a: 68-110) makes use of the following sets of meanings to establish "related meanings of different lexical units": contiguous sets of meanings (e.g. *hat*, *cap*, *beret* and *helmet*), included sets of meanings (e.g. hierarchically related items such as *mammal*: *cat* and *tiger*), overlapping sets of meanings (e.g. *peace*: *tranquility*, *father*: *daddy*, *paper*: *article*) .and complementary sets of meanings (which includes opposites, reversives and conversives/reciprocals). (See Nida 1975a for the proper understanding of his classification of lexical concepts.)

3.4.3. Lexical Inheritance

Hypernymy-hyponymy and meronymy-holonymy assure a lexical item to inherit semantic features as exemplified below:

kuiyl 'quail' – *iniya kuraluTaiya kariya paRavai* 'a black bird with sweet voice'

paRavai – irukaalkaLum alakukaLum uTaiya, uTalin iruppakkangkaLilum paRappataRku eeRRavakaiyil ciRakukaL koNTa vilangkinam 'an animal with two legs and a beak and wings at its sides of its body for flying'

vilangkinam 'animal' – *taanaaka iyangkum uTal uRuppukaLum celluloos ilaata celkaLum uLLa uyiringkaL'* living beings with parts functioning automatically and cell walls without cellulose'

uyirinam - uyir vaazkiRa onRu 'one which lives'

3.4.4. Lexical Oppositions

There are many types of oppositions (Lyons, 1977: 270-290; Cruse, 1986: 197-263, Cruse, 165-176). They can be grouped into two types of oppositions or contrasts based on the number of items involved in the contrast: binary contrast or opposition and non-binary contrast or non-binary opposition. If the contrast is made between two lexical items, it is called binary contrast; if the contrast is made between more than two lexical items, it is called non-binary contrast. The binary-contrasts are listed below with examples from Tamil.

Gradable opposites: The gradable antonyms or opposites occur as end points on a scale, and the denial of one member of the pair does not imply the assertion of the other. E.g. *nalla* 'good': *keTTa* 'bad'; *azakaana* 'beautiful' : *kuruuramaana* 'ugly', *uyaramaana* 'long' : *kuLLamaana* 'short'.

Ungradable opposites or non-gradable opposites: Non-gradable antonymy divides the universe of discourse absolutely without admitting degrees of more or less. E.g. *aaN* 'male': *peN* 'female'; *kalyaaNamaana* ‘married’: *kalyaaNamaakaata* ‘un-married’, *uyiruLLa* ‘live’: *uyiraRRa* ‘dead’.

Complementaries: Complementaries do not fall on a scale, though in certain borderline cases something like a scale might exist. E.g. *pakal* 'day: *iravu* 'night', *uNmai* 'truth': *poy* 'false', *aaN* 'male': *peN* 'female' *veRRi* 'success': *toolvi* 'failure'.

Privative Opposites: In privative opposition, one item of the pair denotes some positive property and the other denotes the absence of that property. They are referred as privative opposites. E.g. *uyiruLLavai* 'living beings': *uyirillaatavai* 'non-living beings'.

Equipollent opposites: In equipollent opposition, both the contrasting lexemes denote a positive property. They are referred as equipollent opposites. E.g. *aaN* 'male': *peN* 'female'.

Converses: Converseness is the relationship that holds between such pairs of words (i.e. converses) in which one can be considered as the reverse of the other. Converseness is distinguished from antonymy and complementarity. e.g, *kaNavan* 'husband': *manaivi* 'wife', *kol* 'kill': *kollappaTu* 'be killed', *vaangku* 'get': *koTu* 'give'; 'X is Y's husband' means 'Y is X's wife'. There are various types of converses.

1. Converse pairs of social roles, e.g. *vaitiyar* 'doctor' : *nooyaaLi* 'patient', *ejamaanar*/*ejamaani* 'boss': *veelaiyaaL* 'servant';
2. Converse pairs of Kinship terms, e.g. *tandtai* 'father'/*taayaar* 'mother' : *makan* 'son'/*makaL* 'daughter';
3. Converse pairs of temporal relations, e.g. *munnaal* before': *pinnaal* 'after', *munnar* earlier': *pinnar* 'later';
4. Converse pairs of spatial relations, e.g. *munnaal* 'in front': *pinnaal* 'at the back', *meelee* 'above': *kiizee* 'below'

Directional opposites: e.g. *meelee* 'above' : *kiizee* 'below', *vandtuceer* 'arrive': *puRappaTu* 'start', *vaa* 'come': *poo* 'go'

Orthogonal opposites (perpendicularly opposites**)**: e.g *vaTakku* 'north' : *meeRku* 'west', *vaTakku*: *kizakku* 'east'; i.e. *meeRku* ‘west’ and *kizakku* ‘east’ are perpendicularly opposite to *vaTakku* ‘north’

Antipodal opposites (diametrically opposites): *vaTakku* 'north' : *teRku* 'south'; *kizakku* 'east' : *meeRku* 'west'; i.e. *vaTakku* ‘north’ is diametrically opposite to *teRku* ‘south’ and *kizakku* ‘east’ is diametrically opposite to *meeRku* ‘west’

Non-binary Opposition: As stated above, if the opposition involves more than two lexical items such contrasts are called non-binary opposition (Lyon, 1977:271). There a number of types of non-binary contrasts. They are dealt under non-branching hierarchies.

3.4.5. Hierarchies

Hierarchies (Cruse 1986:112-118) are of two types: branching hierarchies and non-branching hierarchies.. The branching hierarchy shows tree structure, where as non-branching hierarchy does not show tree structure.

Figure 18

Branching hierarchies

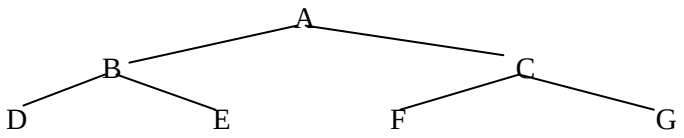


Figure 19

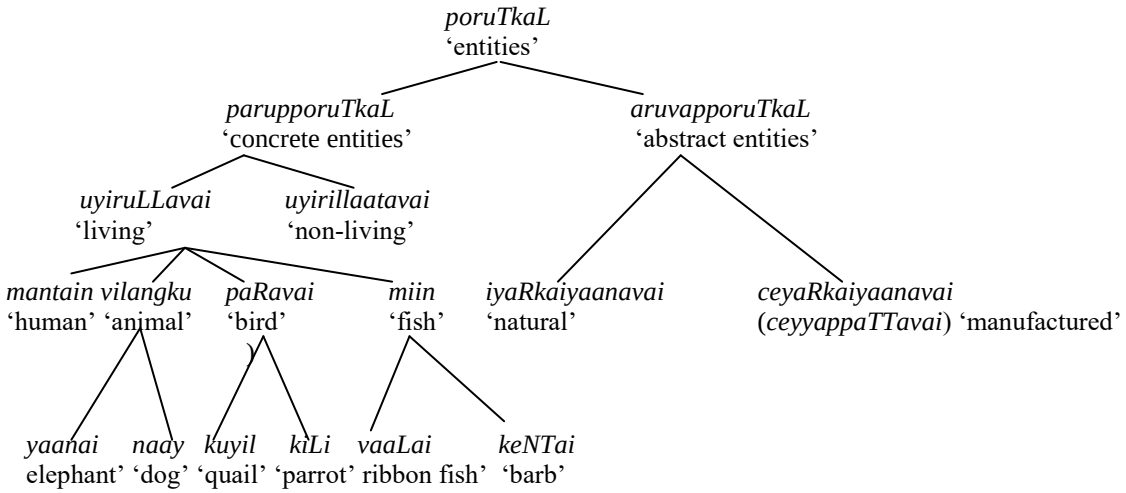
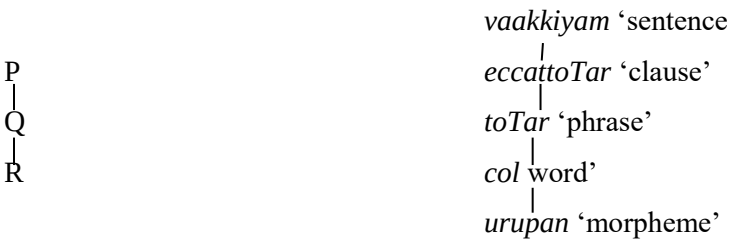


Figure 20

Non-branching hierarchies



3.4.5.1. Taxonomic Hierarchies

The first major type of branching lexical hierarchy is the outcome of the hyponymy-hypernymy relation between lexical items. The sequence of hyponymy-hypernymy relation leads to branched-hierarchical structure of a set of vocabulary items which show this pair of relations among themselves. Taxonomic hierarchies (Cruse, 1986: 136-155) are more liberal than hyponymy-hypernymy hierarchies. The following is an example.

Figure 21

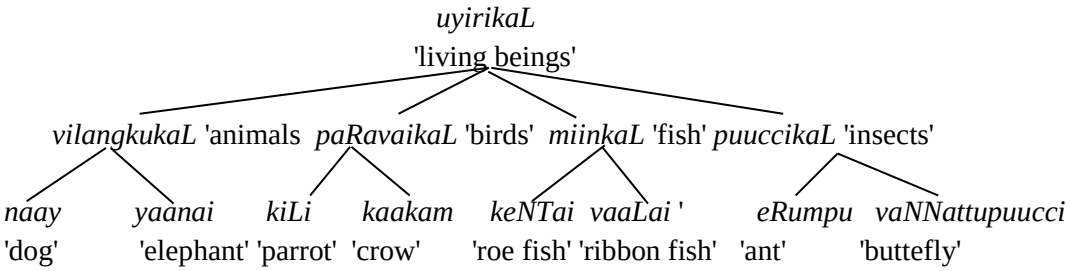
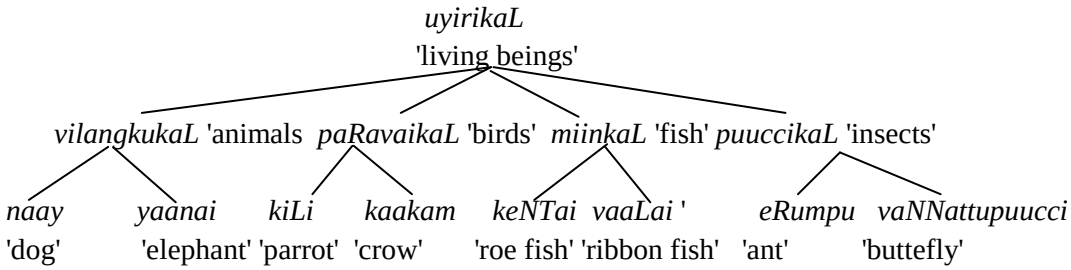


Figure 22

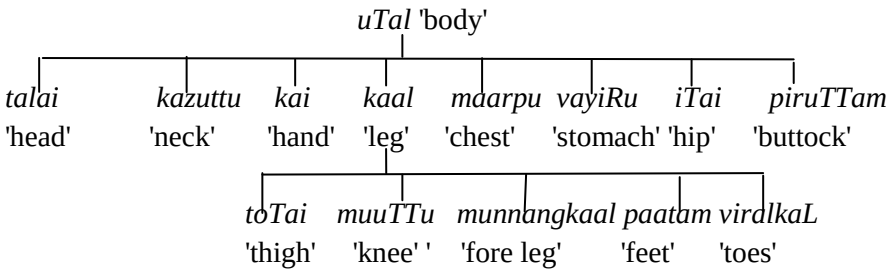


3.4.5.2. Meronymic Hierarchies

The second major type of branching lexical hierarchy is the part-whole type which is called meronomies. (Cruse, 1986: 157-180). Meronymic hierarchies are the result of meronymy-holonymy

relation shown by the lexical items. The meronymy-holonymy relation also gives hierarchical structure to a set of vocabulary. The following is an example.

Figure 23



3.4.5.3. Non-branching Hierarchies

Non-binary opposition leads to a number of types of non-branching hierarchies. They are listed below:

Bipoles: The bipoles is the simplest kind of linear structure found in a pair of opposites. They are simply oppositions which we have discussed earlier (Cruse, 2000: 189). e.g. *niiLamaana* 'long': *kuTTaiyaana* 'short', *viraivaaka* 'fast': *metuvaaka* 'slowly'.

Bipolar chains: The bipolar chains imply a scale on which a pair of opposites operates denoting different degrees of the property. They have implicit superlative terms of opposite polarity at each end of the scale (Cruse, 2000: 189). The following is the example: *mikanuNNiya* 'very minute', *nuNNiya* 'minute', *ciRiya* 'small', *periya* 'big' *mikapperiya* 'very big'

Monopolar Chains: In monopolar, there is no sense that the terms at the ends of the chain are oriented in opposite directions. Degree, stages, measures, ranks and sequences come under monopolar chains (Cruse , 2000: 190).

Degrees: Degrees incorporate as part of their meaning different degrees of some continuously scaled property such as size or intensity (Cruse, 2000: 190). e.g. *paaRai* 'mound': *kunRu* 'hillock': *malai* 'hill' : *maamalai* 'mountain'

Stages: Stages are points in a life cycle of something and normally involve the notion of progression (Cruse 2000: 190). e.g. *kuzandtaip pruvam* 'child stage': *vaalipap paruvam* 'young stage' : *mutir paruvam* 'adult stage': *mutu paruvam* 'old stage'.

Measures: Measures are based on part-whole relationship, with each whole divided into a number of identical parts (Cruse, 2000: 190). *cekaNTu* 'second', *miniT* 'minute', *maNi* 'hour', *naaL* 'day', *vaaram* 'week' *maatam* 'month', *aaNTu* 'year'.

Ranks: The lexical items under ranks entails a sequential order which is not gradual (Cruse, 2000. 191). e.g. *virivuraiyaaLar* 'lecture', *mutunilai virivuraiyaaLar* 'senior lecture', *iNaippeeraaciriyar* 'reader', *peeraaciriyar* 'professor'. Lyons (1977: 290) includes numeral under rank as an unique type. E.g. *onRu* 'one', *iraNTu* 'two' .. *pattu* 'ten', *irupatu* 'twenty, … *nuuRu*, *iRunuuRu*, …

Sequences: Sequences are lexical items which are ordered but do not bear increasing property as in the case of previous chains (Cruse, 2000: 191). e.g. *iRandta kaalam* 'past tense', *nikaz kaalam* 'present tense', *etir kaalam* 'future tense'; *kaar kaalam* 'rainy season' *kuuLir kaalam* 'cold season', *munpani kaalam* 'early cold season', *pinpanik kaalam* 'late cold season', *iLaveeniR kaalam* 'milder hot season', *mutuveeniR kaalam* 'hot season'.

Cyclical Sets or Cycles: The sequential lexical item (Cruse, 1986:187-190) can entail a cyclical order of time in the natural arena. e.g. *kaarkaalam* 'rainy season' *kuutirkaalam* 'cold season', *munpanikkaalam* 'early cold season', *pinpanikkaalam* 'late cold season', *iLaveenirkaalam* 'milder hot season', *mutuveenirkaalam* 'hot season'; *canavari* 'January', *pepravari 'February'*, *maarc* 'March', *eeppiral* 'April', *mee* 'May', *juun* 'June'…}; *njaayiRu* 'Sunday', *tingkaL* 'Monday', *cevvaay* 'Tuesday' , *putan* 'Wednesday', *viyaazan* 'Thursday', *veLLi* 'Friday', *cani* 'Saturday'.

3.4.6. Propositional series or grids

Cruse (1986: 118-133) discusses about propositional series elaborately under lexical configurations. Propositional series is a type of lexical configuration; the other lexical configuration is hierarchy. He uses term grid for the same in another occasion (Cruse, 2000:191-193). The grids generated by recurrent sense relations, or which comes to much the same thing, by recurrent semantic components. The unit of a gird is the cell, which consists of four lexical items, any one of which must be uniquely predictable form the remaining three (Cruse, 2000:191). The followings are the examples of cells: *pacu* 'cow': *kanRu* 'calf' :: *kutirai* 'horse': *kuTTi* 'foe' ; *paampu* 'snake' : *kunjcu* 'young one':: *palli* 'lizard' : *kunjcu* 'young one' ; *yaanai* : *piLiRu 'trumpet'* :: nari 'fox': *uuLaiyiTu* 'howl'. Propositional series or grids help us to relate lexical items by different types of meaning relations. Even derivative relation can be expressed using grids; e.g. *kuLi* 'bathe' : *kuLittal* 'bathing:: *kuTi* 'drink':: *kuTittal* 'drinking'.

3.4.7. Clusters

Cruse (2000: 193) prefers to make use of the term clusters in the place of synonyms. For him clusters are essentially groups of synonyms. He opines that the name is intended to indicate that the sharpness and complexity of structuring is much less than in other types of fields. For him they are somewhat informal groups. He identifies two main types of cluster, centred cluster and non-centred cluster. The centred cluster has more-or-less clear core of one or two, and a penumbra of more peripheral items. In non-centred clusters, the items spread over a spectrum of sense, but there is no superordinate item. It should be noted that Nida (Nida, 1975a) makes use of the term "contiguous sets of meanings". (The opinions of Nida and Cruse are taking into consideration while building VOT for Tamil.)

3.5. Creation of database for VOT

As mentioned earlier, the lexical items are arranged into four major categories entities, events, abstracts and relationals in line with Nida (1975a). Each category requires different representation in the VOT because of their inherent componential features. The organization of lexical items in the data base is discussed below:

3.5.1. Entities in VOT

Nida's (1975a) classification of entities is given above. Rajendran (1983, 2001) has elaborately studied entities and made an operational classification of entities using the principles of componential analysis. Entities are represented as nouns in the surface level or formal level. It is proposed to make use of Nida's classification for organizing entities in VOT. Relations pertaining to entities can be captured by lexical relations such as synonymy, hyponymy, compatibility, incompatibility and meronymy which have been elaborately discussed in the previous sections.

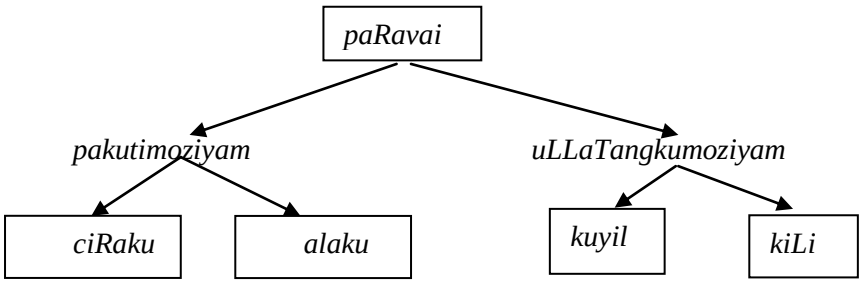
Entities in VOT contain lexical items denoting concrete objects. Like any thesaurus, synonymy is captured in VOT as a set of lexical items denoting the same meaning and labeled as *iNaiccoRkaL* 'the lexical items having the same meaning'. The notion of synonymy (referred here as *iNaimoziyam*) does not entail interchangeability in all contexts. By that criterion, natural languages have few synonyms. The more modest claim is that synonyms can be interchanged in some contexts. Although synonymy is a semantic relation between word forms, the semantic relation that is most important in organizing entities is the relation of subordination (or class inclusion or subsumption), which is called hyponymy. It is this semantic relation that organizes entities into a lexical hierarchy. Each hyponym leads on to a more generic hypernym. Hyponymy-hypernym relation cannot be represented as a simple relation between word forms. Hyponymy is a relation between lexicalized concepts, a relation that is represented in VOT by the label '*uLLaTangkumoziym* 'hyponymy' between the appropriate lexical concepts. A lexical hierarchy can be reconstructed by following the trail of hyponymically-hypernumically related lexical items.

For each hyponymic relation we can add a corresponding hypernymic relation that points in the opposite direction. What emerges from this manner of representing hyponymy and hypernymy is a lexical hierarchy. Hierarchies of these sorts are widely used by computer scientist as a means of representing knowledge. The entities in VOT from a lexical inheritance system; a systematic effort has been made to connect hyponyms with their hypernyms (and vice versa). VOT presupposes a linguistic knowledge of anaphoric relations; an anaphor can be a hypernym of its antecedent.

More generally, a hyponym can replace a more specific term whenever the context ensures that the substitution will not produce confusion. It is of some interest that these levels are shallow. In principle, of course, there is no limit to the number of levels in inheritance system can have. Lexical inheritance systems, however, seldom go more than ten levels deep, and the deepest examples usually contain technical levels that are not part of the everyday vocabulary. Some hierarchies are deeper than others:

Another important relation which helps in the tree representation of entities is part-whole relation. Part-whole relation between entries is generally considered to be a semantic relation, called meronymy. It is comparable to synonymy, antonymy, and hyponymy. The relation has an inverse: if x is a meronym of y, then y is said to be a holonym of x. For concrete objects like bodies and artifacts, meronymy can help to define a basic level. Meronyms are distinguishing features that hyponyms can inherit. Consequently, meronymy and hyponymy are intertwined in complex ways. For example, if *alaku* 'beak' and *ciRaku* 'wing' are meronyms of *paRavai* 'bird', and if *kuyil* 'koel' is a hyponym of bird, then by inheritance, beak and wing must also be meronyms of *kuyil* 'koel'. In VOT the relation homonymy is referred as *uLLaTangkumoziyam* and the relation meronymy is referred as *pakutimoziyam.* The above mentioned scenario is depicted in VOT by the tree-viewer (explained at the end) as follows:

Figure 24



It has been said that distinguishing features are introduced into noun hierarchies primarily at the level of basic concepts; some claims have been made that meronym is particularly important for defining

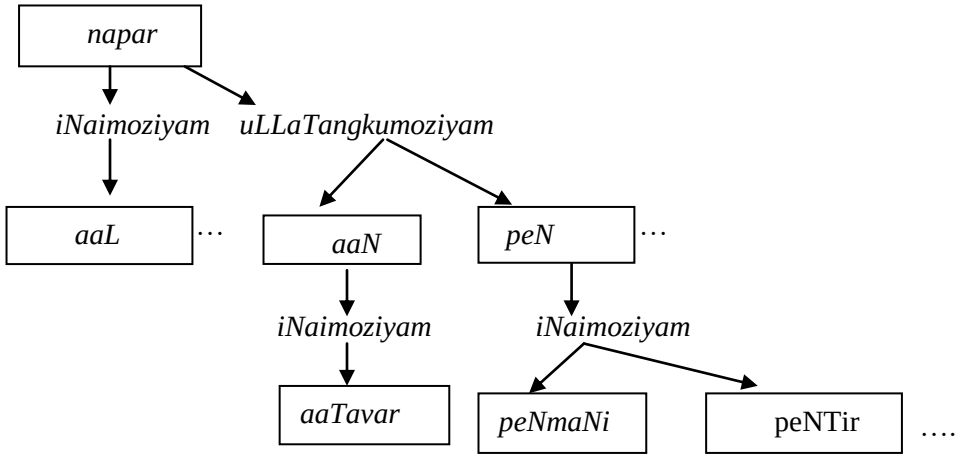
basic concepts. Meronymy is often compared to hyponymy: both are asymmetric and transitive, and both can relate terms hierarchically. In many instances transitivity seems to be limited for meronymy. For example, *piTi* 'handle' is a meronym of *katavu* 'door' and *katavu* 'door' is a meronym of *viiTu* 'house', yet it sounds odd to say *viiTTiRku piTi irukiRatu* 'The house has a handle' or *piTi viiTin oru pakuti* 'The handle is a part of the house'.

The strongest psycholinguistic indication that two words are antonyms is that each is given on a word association test as the most common response to the other. Semantic opposition is not a fundamental organizing relation between nouns, but it does exist and so merits its own representation in onto-thesaurus.

napar 'person', *aaL* 'person' {*aaN* 'male person', <*aaTavar* 'male person>} *peN* 'female person', <*peNmaNi* 'female person', *peNTir* 'female person'>}

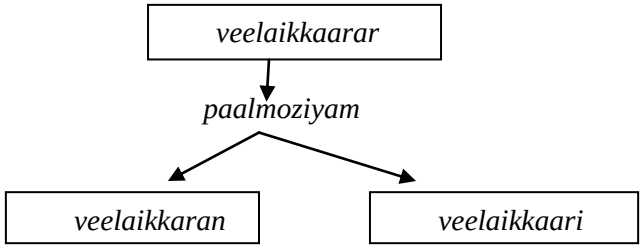
Note that hyponyms arranged (or included) under hypernym(s) by using the curly brackets ({}) and synonyms are arranged (or included) under the respective word-form by using the ankle brackets (< >). This will give the following visualization in the tree-viewer:

Figure 25



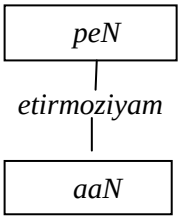
In Tamil certain human entities have three forms; one is the epicene form, another is male form and the third is the female form; the epicene form implies a kind of respect when compared to the other two forms. For example for the word *servant* there are three forms: *veelaikkaarar* 'servant' (epicene form), *veelaikkaaran* 'male servant' and *veelaikkaari* 'female servant'. In the tree viewer the gender marked forms are given under epicene form with the label *paalmoziyam* 'gender marked relation'.

Figure 26



The relation between the three forms are established by giving the gender marked forms under the epicene form by using the special separators Antonymy is a lexical relation between words, rather than a semantic relation between concepts. The antonyms are related to one another by using the special separators: % and *. For example *peN* 'female person' is given under *aaN* 'female person by using the above mentioned separators as follows: *aaN %peN**. In the tree-viewer this will be represented by using the label *etirmoziyam* 'antonymy' as given below.

Figure 27



When all three kinds of semantic relations – hyponymy, meronymy, and antonymy – are included, the result is a highly interconnected network of entities.

The term 'function' has served many purposes, both in psychology and linguistics. The functional feature of a nominal concept is intended to be a description of something that instances of the concept normally do, or that is normally done with or to them. For example, it seems natural to say that the function of a pencil is to write or the function of knife is to cut, to say that the function of a canary is to fly or to sing seems a bit forced. Nominal concepts can play various semantic roles as arguments of the verbs that they co-occur with in a sentence (Miller et al 1990).

katti 'knife – *veTTu* 'cut'

kuzi 'hole' – *tooNTu* 'dig'

paTam 'picture' - *varai* 'draw'

There are also linguistic reasons to assume that a thing's function is a feature of its meaning. It should be mentioned here that Pustejovsky (1995) in his generative lexicon talks about telic role in his qualia structure. At present VOT does not relate entities with their 'function' or telic role. This will be incorporated in VOT latter.

Relations	Subtypes	Example
Synonymy		*puttakam* 'book' to *nduul* 'book'
Hypernymy-Hyponymy		*vilangku* 'animal' to *paaluuTTi* 'mammal'
Hyponymy-Hypernymy		*pacu* 'cow' to *paaluuTTi* 'mammal'
Holonymy-Meronymy	Wholes to parts	*meecai* 'table' to *kaal* 'leg'
,,	Groups to members	*tuRai* 'department' to *peeraaciriyar* 'professor'
Meronymy-Holonymy	Parts to wholes	*cakkaram* 'wheel' to *vaNTi* 'cart'
,,	Members to groups	*paTaittlaivar* 'captain' to *paTai* 'army'
Binary Opposites	Antonymic (gradable)	*ndallavan* 'good person' to *keTTavan* 'bad person'
,,	Complementary	*aaTavar* 'man': *makaLir* 'woman'
	Converse	*kaNavan* 'husband' *to manaivi* 'wife'
,,	Privative (opposing features)	*ahRiNai* 'irrational' to *uyartiNai* 'rational'
,,	Equipollent (positive features)	*aaN* 'male' to *peN* 'female'
,,	Reciprocal Social roles	*vaittiyar* 'doctor' to *ndooyaaLi* 'patient'
,,	Kinship Relations	*ammaa* 'mother' to *makaL* 'daughter'
,,	Antipodal Opposition	*cikaram* 'peak' to *aTi* 'foot (of mountain)'
	Orthogonal opposition	*ciRumi* 'girl' : *ciRuvan* 'boy' *and peNTir* 'woman'

“	Degrees	*paaRai* 'mound': *kunRu* 'hillock': *malai* 'hill' : *maamalai* 'mountain'
,,	Ranks	*virivuraiyaaLar* 'lecture', *mutunilai virivuraiyaaLar* 'senior lecture', *iNaippeeraaciriya* 'reader', *peeraaciriyar* 'professor'
Compatibility		*ndaay* ‘dog’ to *cellappiraaNi* ‘pet’

3.5.2. Organization of Events in VOT

The semantic domain EVENTS comprises of verbs and the abstract nouns derived from them. Nida’s (1976b) tentative classification of events into twelve semantic domains based on componential analysis has been given already. Events are mostly realized in the surface level as verbal forms. Rajendran (1978) classified verbs into 31 groups out of which nine are major important semantic domains. The important semantic domains identified by him based on componential analysis of verbs are: 1. Verbs of movement (i.e. change of position), 2. Verbs of transferring (change of possession), 3.Verbs of change of state (change of shape, condition, etc), 4. Verbs of impact, 5. Verbs of senses, 6. Verbs of emotion, 7. Verbs of intellection, 8. Verbs of communication and calling, 9. Verbs of association. Each major domain is divided into sub domain by taking into account distinguishing semantic component. This classification need second look to make it more user-friendly. Even though verbs do not show hierarchical ordering, a quasi-hierarchical ordering is possible by taking into account certain pertinent distinguishing semantic features. For wider coverage of verbs, it is proposed to follow the twelve-way classification of verbs by Nida (1975a) and this tentative classification is liable to change to accommodate more verbs.

3.5.2.1. Polysemous Nature of Verbs

The verbs are fewer in number than nouns in Tamil and at the same time verbs are more polysemous in nature than nouns. The semantic flexibility of verbs makes the lexical analysis of verbs difficult. A look at the Tamil corpus or Tamil dictionary will reveal the polysemous behaviour of verbs. The polysemy will be captured in line with Nida (1975). He elaborately discusses about the representation of polysemy of the verb *run* (Nada: 138-150) in his thesaurus. The following table outlines the different senses in *ooTu* is used (Rajendran,1978).

Different senses	**Examples**
I. Movement	
1. run as animals	*avan pattu mail tuuram ooTinaan*

	'He ran for about ten miles'
2. run as a liquid in a channel, river, tube, vessel, etc. (Note: *poo* 'go' can replace *ooTu* in this context.)	*aaRRil veLLam ooTukiRatu* 'The water is running in the river'
3. work as a machine (the movement of which can be seen from the movement of wheels).	*kaTikaaram ooTukiRatu* 'The clock is running'
4. run as vehicles (generic locomotion); ply.	*rayil taNTavaaLattil ooTukiRatu* 'Train moves on tracks' *kappal taNNiirril ooTukiRatu* 'Ship moves in water' *cennaiyilirundtu kanniyaakumaarikku bas ooTukiRatu* 'Buses are playing between Chennai and Kanyakumari'
5. escape	*avan miinaip piTikkap poonaan, aanaal atu ooTiviTTatu* 'He tried to catch the fish, but it ran away' *avan vaNNattuppuucciyaip piTikkap poonaan, aanaal atu ooTiviTTatu* 'He tried to catch the butter-fly, but it ran away'
6. elope (Note: The compound *ooTippoo* 'having run go' also gives the meaning 'elope away')	*avaL avan kuuTa ooTiviTTaaL* 'She eloped away with him'
II. abstract movement	
1. run or go on as a performance, a business, an organization, life, etc. (Note: *ndaTa* 'walk; happen' can be used in the place of *ooTu* in all these contexts. Running of dance or drama performance cannot be denoted by the verb *ooTu*.)	*anta tiyeeTTaril oru cinimaa ooTukiRatu* 'A cinema is running in that theatre' *viyaapaaram ndanRaaka ooTukiRatu* 'The business is going on well' *kampani ndanRaaka ooTukiRatu* 'The company is running well' *vaazkkai eppaTiyoo ooTukiRatu* 'The life is going on somehow'
2. pass quickly as time.	*ndaan inkee vandtu muunRu varuTankaL ooTiviTTana* 'Three years have passed after my coming over

	here’
3. by capable of comprehending, doing work, etc. (Note: *poo* ‘go’ is synonymous to *ooTu* in this context. Note: The verb in this context receives a dative-subject. The verb *vaa* ‘come’ can also be used in the place of *ooTu*. *ooTu* in this context when compounded with the negative auxiliaries *illai* and *maaTu* gives the meaning ‘be incapable or paralyzed’.	*enakku kaNakku ooTum* ‘I can comprehend mathematics’ *avanukku kottaveelai ooTum* ‘He can do masonry’ *enakku kaNakku ooTavillai* ‘I could not comprehend mathematics’ *enakku veelai ooTamaaTTeen enkiRatu* ‘I am unable to work’ *avanaip paarttatum enakkuk kaiyum kaalum ooTavillai* ‘(As soon as I saw him I could not operate my hands and legs) I was paralyzed by seeing him’ *enakku onRumee ooTavillai* ‘I could not do anything (I am inactive).

Nida (1975a) elaborately discusses about treating “different meanings of the same lexical units”. At present polysemy is not taken care of explicitly in VOT. The verbal polysemy is a challenging problem. The conceptual classification of lexical items, itself will take care of certain amount of polysemy. The contrastive polysemy is very well taken care of in VOT. Only certain types of complementary polysemy need to be taken care of. This will be done at the later stage of VOT. Predictable polysemy can be tackled by incorporating more features in VOT. Metaphorical and metonymic extensions of meaning of lexical items can be resolved by incorporating Pustejovsky’s (1995) principles of generative lexicon.

3.5.2.2. Componential Features of Verbs

Verbs can be paraphrased in terms of finer semantic features. The decompositional nature of verbs can be exploited for the interpretation of verbs denoting complex events in terms of verbs denoting simple events. For example the verb *kol* ‘kill’ can be decomposed into ‘cause not to become alive’. The verb *eRi* ‘throw’ can be decomposed into ‘cause an object to move away from one’s possession by force’. The decompositional nature of verbs reveals the entailment relation existing between verbs. For example, the entailment of simple verb under causative verb (ex. *ooTu* ‘run’ vs. *ooTTu* ‘cause to run’) is understood by decompositional nature of verbs. The decompositional features of verbs can be captured by the componential analysis of verbs into finer semantic components (Leech, 1974). All types of lexical relations such as synonymy, entailment, hyponymy and troponymy and sentential properties such as

presupposition, inconsistency, tautology, contradiction, and semantic anomaly can be mapped clearly if verbs are decomposed into componential features. The decompositional means of relating verbs or events will be considered in VOT.

3.5.2.3. Synonymy among Verbs

Synonymy is a rare phenomenon in verbal domain. Verbal domain exhibits only a few truly synonymous verbs. Take for examples the words *paTi* 'read' and *vaaci* 'read'. *avan puttakam paTikkiRaan* 'He is reading a book' can entail *avan puttakam vaacikkiRaan* 'He is reading a book'. The relation existing between *paTi* and *vaaci* is synonymy and *paTi* and *vaaci* are synonyms, at least in this context. Truly synonymous verbs are difficult to find, mostly quasi synonymous verbs are found in Tamil. The existence of a simple and a parallel compound forms (noun + verbalizer) prompts synonymy (quasi synonymy) in verbal system of Tamil.

kol 'kill' and *kolai cey* 'murder

vicaari 'enquire' and *vicaaraNai cey* 'investigate'

The synonymous expressions of many verbs show that they are manner elaborations of more basic verbs. For example, *viniyooki* 'distribute' can be considered as an elaboration of the basic verb *koTu* 'give'. The more effective way of depicting the lexical and semantic relations among verbs is to establish these relations in terms of different senses of each verb. VOT makes use of synonymy to relate one verb (verbal concept) with another verb (verbal concept) whenever it is possible.

3.5.2.4. Lexical Entailment and Meronymy

Lexical entailment refers to the relation that holds between two verbs when the statement *"*X entails Y*"*. For example, *kuRaTTai viTu* 'snore' lexically entails *tuungku* 'sleep' because the sentence *avan kuRaTTai viTukiRaan* 'he is snoring' entails *avan tuungkukiRaan* 'he is sleeping'; the second sentence is true if the first one is true. Lexical entailment is a unilateral relation: if a verb V1 entails another verb V2, then it cannot be that case that V2 entails V1. For example, *uRangku* need not entail *kanavukaaN*.

The entailment relation between verbs discussed above is similar to meronymy found between nouns, but meronymy is more suitable to nouns than to verbs. Fellbaum and Miller (1990) argue that, first, verbs cannot be taken as parts in the same way as nouns, because the parts of verbs are not analogous to the parts of nouns. Most nouns and noun parts have distinct, delimited referents. The referents of verbs, on the other hand, do not have the kind of distinct parts that characterize objects, groups, or substances. Componential analyses have shown that verbs cannot be broken into referents denoted solely by verbs. It is true that some activities can be broken down into sequentially ordered sub-

activities, say for example *camai* 'cook' is a complex activity involving a number of sub-activities. Consider the relation between the verbs *vaangku* 'buy' and *koTu* 'pay'. Although neither activity is a discrete part of the other, the two are connected in that when you buy something, somebody gives it to you. Neither activity can be considered as a sub-activity of the other. Consider the relations among the activities denoted by the verbs *kuRaTTaiviTu* 'snore', *kanavukaaN* 'dream', and *uRanku* 'sleep'. Snoring or dreaming can be part of sleeping, in the sense that the two activities are, at least, partially, temporally co-extensive; the time that you spend snoring or dreaming is a proper part of the time you spend sleeping. And it is true that when you stop sleeping you also necessarily stop snoring or dreaming. The relation between pairs like *vangku* 'buy' and *koTu* 'pay' and *kuRaTTaiviTu* 'snore' and *uRagnku* 'sleep' are due to the temporal relations between the members of each pair. The activities can be simultaneous (as in the case of *vaangku* 'buy' and *koTu* 'pay' or one can include the other (as in the case of *kuRaTTaiviTu* 'snore' and *uRangku* 'sleep'). VOT makes of entailment to relate one verb with another verb whenever it is possible.

3.5.2.5. Hyponymy among Verbs

Some verbs seem more generic than others. For example, *koTu* 'give' describes a wider range of activities than *viniyooki* 'distribute'. The hyponymous relation of the kind found in nouns cannot be realized in verbs. The sentence frame, *An x is a y*, which is used to establish hyponymous relation between nouns is not suitable for verbs, because it requires that *x* and *y* be nouns. The scrutiny of hyponyms and their superordinates reveals that lexicalization involves different kinds of semantic expansions across different semantic domains (Miller 1990, Felbaum, 1998). The analysis of verbs of motion in Tamil (Rajendran, 1978) reveals the fact that the semantic component such as +DIRECTION (eg. *eeRu* 'climb up' vs *iRanku* 'climb down'), +MANNER (eg. *ndazuvu* 'slip down' vs *vizu* 'fall') + CAUSE (eg. *ooTu* 'run' vs. *ooTTu* 'cause to run', +SPEED (e.g. *uur* 'crawl' vs *ooTu* 'run) added to the common semantic component +MOVE establish co-hyponymous relation found among verbs of motion. Miller (1991) makes use of the term troponymy to establish this type of relation existing between verbs. "When two verbs can be substituted into the sentence frame To V1 is to V2 in a certain manner, then V1 is a troponym of V2" (Miller, 1991:228). For example, *ndoNTu* 'to walk unevenly' is a troponym of *ndaTa* 'walk' as the former entails the latter.

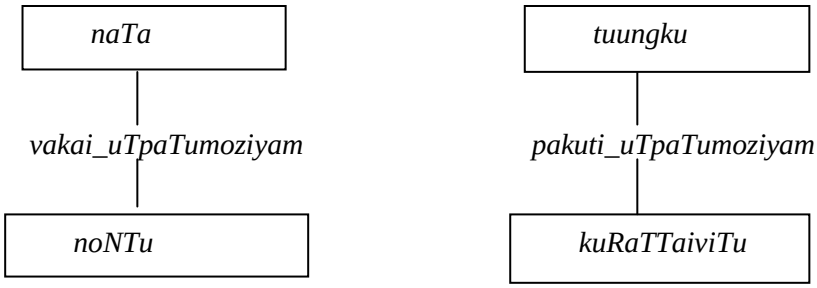
3.5.2.6. Troponymy and Entailment

Troponymy is a particular kind of entailment in that every troponym of a more general verb X also entails Y (Miller 1990, Felbaum, 1998). Consider for example the pair *noNTu* 'limp' and *naTa* 'walk'. The verbs in this pair are related by troponymy: *noNTu* is also *naTa* in a certain manner. So *noNTu* is a

troponym of *naTa*. The verbs are also in entailment relation: the statement *avan noNTukiRaan* 'he is limping' entails *avan naTakkiRaan* 'he is walking'.

In contrast with pairs like *noNTu* 'limp' *and naTa* 'walk', a verb like *kuRaTTaiviTu* 'snore' entails and is included in *tuungku* 'sleep', but is not a troponym of *tuungku*. Similarly *vaangku* 'buy' entails *koTu* 'give', but is not a troponym of *koTu* 'give'. The verbs in the pairs like *kuRaTTaiviTu* 'snore' and *tuungku* 'sleep' are related only by entailment and proper temporal inclusion. It can be generalized that the verbs related by entailment and proper temporal inclusion cannot be related by troponymy. If the activities denoted by two verbs are temporally co-extensive, they can be linked by troponymy. Troponymy represents a special kind of entailment. In VOT the lexical items linked by troponymy are plotted in tree viewer as given below. The relation between *naTa* amd *noNTu* are referred as *vakai_uTpaTumoziyam* 'type-troponymy' and the relation between *tungku* and *kuRaTTaiviTu* are referred as *pakuti_uTpaTumoziyam* 'part-troponymy'.

Figure 28



Troponyms can be related to their superordinates in various ways, subsets of which tend to come together within a given semantic domain. In the semantic domain of verbs of communication, troponyms denotes the speaker's objective or drive for communicating. Even though troponymy culminates in hierarchical structure for verbs parallel to hyponymic structure for nouns, they vary significantly. Verbs tend to have superficially branched structure. In most case, the number of hierarchical levels does not exceed four. Moreover, within a semantic domain, not all verbs can be grouped into a single hierarchy, under a single term. VOT makes use of paraphrases (or descriptions) to relate certain pairs of verbal concepts or events.

3.5.2.7. Opposition Relations and Entailment

Opposition relations are psychologically significant not only for adjectives, but also for verbs. It is found that after synonymy and troponymy, opposition relations are the most frequently coded semantic relations in building database for verbs (Miller 1990, Felbaum, 1998). The semantics of opposition

relations among verbs is complex. As for as Tamil is concerned there is no morphologically derived opposite verbs. Some of the oppositions found among nouns are absent in verbs. A number of binary oppositions have been shown by the verbs that include converseness, directional, orthogonal, and antipodal oppositions. Active and passive forms of transitive verbs can be taken as showing converse opposition. *avan avaLaik konRaan* is in converse relation with the passive expression *avaL avanaal kollappaTTaaL*. Thus active-passive pairs of transitive verbs in Tamil show converse opposition. The relation between the verbs *vaangku* 'buy' and *vil* 'sell' is rather more complex. The lexical items that are directionally opposite are in directional opposition. The relationship which hold between the pairs such as *vandtuceer* 'arrive' and *puRappaTu* 'reach', *vaa* 'come':and *poo* 'go' is directional opposition. Under this category are the verb pairs such as *uyar* 'rise' and *taaz* 'go down', *eeRu* 'ascend' and *iRangku* 'descend'. There are many other oppositions with reference to change of state, manner, speed, etc. as exemplified below:

kaTTu 'build' : *iTi* 'demolish'

kaTTu 'tie' : *aviz* 'untie'

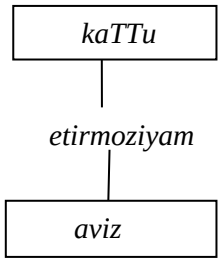
ottukkoL 'agree' : *maRu* 'disagree'

uLLizu 'inhale' : *veLiviTu* 'exhale'

ndaTa 'walk' : *ooTu* 'run'

The opposition between verbs is represented in VOT as follows:

Figure 29



Not only the opposing features, even the presence or absence of a feature can also keep two items in opposition relation. These contrasting or distinguishing features can be arrived at by componential analysis of verbs (Rajendran, 1978). The componential analysis of verbs shows that many verb pairs in an opposition relation also share an entailed verb. For example the pair *jeyi/vel* 'succeed' and *tool* 'fail' entails *muyal* 'try'. The relation between the fist and the second are referred as *muRkooL_uTpaTumoziyam* 'presupposed-troponymy' .

3.5.2.8. Causation and Entailment

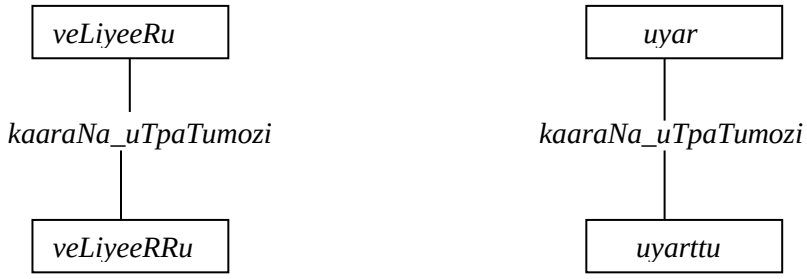
The causative relation exists between two verbal concepts: one is causative (e.g. *koTu* 'give) and the other is resultative (e.g. *peRu* 'get'). Causation can be considered as a specific kind of entailment as denoted by the following examples (Miller 1990, Felbaum, 1998).

veLiyeeRRu 'expel' entails *veLiyeeRu* 'leave'

uyarttu 'raise' and *uyar* 'rise' (temporal inclusion)

We have distinguished four different kinds of lexical entailment that systematically interact with the semantic relations mapped in VOT. The concerned pairs are linked by the relation referred as *kaaraNa_uTpaTumozi* 'cause_entailment'.

Figure 30



3.5.2.9. Syntactic Properties and Semantic Relations

In recent years there is a trend incorporating syntactic properties in the lexicon itself. Viewing verbs in terms of semantic relations can also provide clues to an understanding of the syntactic behaviour of verbs. Incorporating the syntactic properties of verbs in VOT has to be explored for the better understanding of verbal concepts or events.

3.5.2.10. Summing up of Relations of Events in VOT

The following table sums up the lexical relations to be captured in the verb net.

Relations	**Definition/sub types**	**Example**
Synonymy	Replaceable events	*tuungku* 'sleep' → *uRangku* 'sleep'
Meronymy- Hypernymy	From events to superordinate events	*paRa* 'fly' → *pirayaaNi* 'travel'
Troponymy	From events to their subtypes	*naTa* → *noNTu* 'limp'

Entailment	From events to the events they entail	*kuRaTTaiviTu* 'snore' muyal 'try' *tuungku* 'sleep'
"	From event to its cause	*uyar* 'rise' → *uyarttu* 'raise'
"	From event to its presupposed event	*vel* 'succeed' → *muyal* 'try'
"	From even to implied event	*kol* 'murder' → *iRa* 'die'
Antonym	Opposites	*kuuTu* 'increase' → *kuRai* 'decrease'; *kaTTu* 'build' *iTi* 'demolish'
"	Conversensess	*vil* 'sell' → *vaangku* 'buy'
"	Directional opposites	*puRappaTu* 'start' → *vandtuceer* 'reach'
Derivatives	Verb to verbal noun	*paTi* 'study' → *paTippu* 'education'

3.5.3. Organization of Abstracts in VOT

As we noted already, Nida (1978) classifies abstracts into following classes:

1. Time, 2. Distance, 3. Volume, 4. Velocity, 5. Temperature, 6. Color, 7. Number, 8. Status, 9. Religious character, 9. Attractiveness, 10. Age, 11. Truth-falsehood, 11. Good-bad, 12. Capacity, 13. State of health, etc. (Nidia, 1978)

Nida considers abstracts as meanings which can be realized at the outset as adjectives and adverbs. Dixon (1982) has suggested that the lexical items that are generally found to get included in the category of adjectives can be grouped into seven distinct semantic types. They are:

- Dimension (ex. *kuTTaiyaana* 'short', *kuRukalaana* 'narrow')
- Physical Property (ex. *periya* 'big', *cinna* 'small')
- Colour (ex. *veLLai* 'white', *kaRuppu* 'black')
- Human Propensity (ex. *kuruTTu* 'blind', *ceviTTu* 'deaf')
- Age (ex. *putiya* 'new', *pazaiya* 'old')
- Value (*ndalla* 'good', *keTTa* 'bad')
- Speed (ex. *veekamaana* 'quick', *metuvaana* 'slow')

Rajendran (2001) has classified abstracts in which adjective forms a part into 38 sub-domains by taking into account the componential features of meaning and classification of Nida (1978) and Dixon (1982). The inherent adjectives and adverbs are comparatively less in number than the derived adjectives and adverbs in Tamil. In VOT the derived abstract concepts (i.e. adjectives and adverbs) are represented mainly in their nominative forms and the adjectival and adverbial concepts related to them by derivative relation. The lexical sets are built taking into account the above mentioned classification and the adjectival and adverbial concepts are related to their nominal counterparts by certain relation discussed below.

Abstracts in VOT contain mainly adjectives and adverbs apart from abstract nouns. Noun modification is primarily associated with the syntactic category "adjective." Similarly verb modification is associated with the syntactic category "adverbs". Adjectives have their sole function the modification of nouns, whereas modification is not the primary function of noun, verb, and prepositional phrases. The lexical organization of adjectives is unique to them, and differs from that of the other major syntactic categories, noun and verb. Three types of adjectives can be distinguished: Descriptive adjectives (Ex. *periya* big, *kanamaana* 'heavy'), Relational adjectives (Ex. *poruLaataara* 'economic', *cakootara* 'fraternal'), Reference modifying adjectives (Ex. *pazaiya* 'old', *munnaaL* 'former') (Miller, 1998b).

3.5.3.1. Descriptive Adjectives

A descriptive adjective is one that ascribes a value of an attribute to a noun. For example, *atu kanamaana cumai* 'that luggage is heavy' presupposes that there is attribute *eTai* 'WEIGHT' such that *eTai (cuma*i 'luggage') = *kanam* 'heavy'. In the same way *taazndta* 'low' and *uyarndta* 'high' are values of HEIGHT (Miller, 1998b).

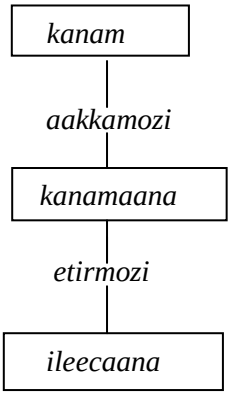
VOT has to link the descriptive adjectives with the appropriate attributes. The descriptive adjectives require a semantic organization which differs drastically form that of nouns. The hyponymic relation that builds nominal hierarchies is not available for adjectives. It is not possible to say that one adjective 'is a kind of' some other adjective. Relating descriptive adjectives with the particular noun they pertain to is known by the term pertainymy.

3.5.3.2. Antonymy in Adjectives

Antonymy is the basic semantic relation that exists among descriptive adjectives. The word association testes reveal the importance of antonymy in adjectives (Miller, 1998:48-52). As the function of descriptive adjectives is to express values of attributes, and that nearly all attributes are bipolar, antonymy becomes important in the organization of descriptive adjectives. Antonymous adjectives express opposing values of an attribute. For example, the antonym of *kanamaana* 'heavy' is *ileecaana* 'light' that expresses a value at the opposite pole of the *kanam* 'WEIGHT' attribute (Miller, 1998).

Antonymy, like synonymy, is a semantic relation between word forms. The problem is that the antonymy relation between word forms is not the same as the conceptual opposition between word meanings.

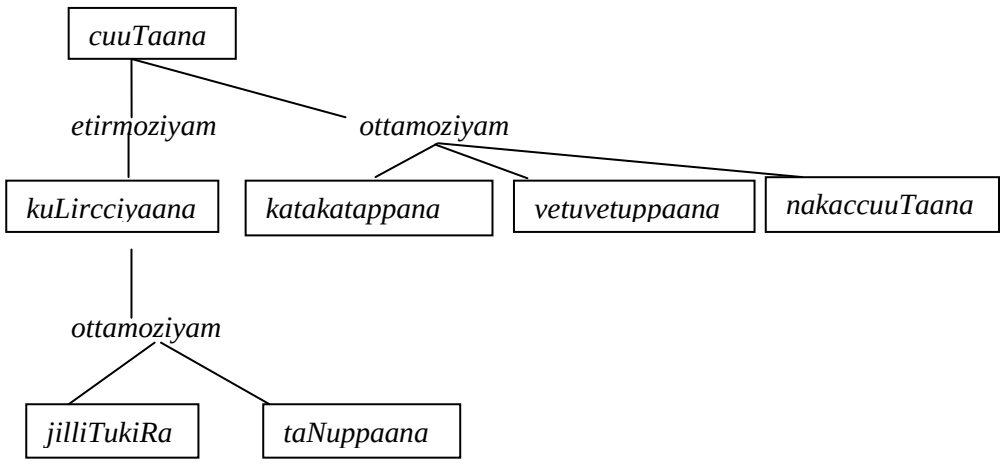
Figure 31



3.5.3.3. Similarity in Adjective

Adjectives show bipolar structure (Miller, 1998:50-52). A set of adjectives show similarity of meaning with an adjective which is antonymous with another set adjectives which show similarity with another adjective. The following examples show the existence of a bunch of adjectives denoting hotness against the bunch of adjectives denoting coldness. We can link them through their typical representatives as shown below. Similarity relation is mentioned in VOT as *ottamoziyam*. *Ottamoziyam* 'similarity' relation is different form *iNaimoziyma* 'synonymy'.

Figure 32



It can be inferred that *periya* 'big' and *ciRiya* 'small' form an antonymous pair and *niiNTa* 'long' and *kuTTaiyana* 'short' form another antonymous pairs. These pairs demonstrate that antonymy is a semantic relation between words rather than concepts.

3.5.3.4. Gradation in Adjectives

Gradation (Miller, 1998: 52-53) is one of the important properties found among adjectives. Most discussions on antonymy distinguish between contradictory and contrary terms. This terminology is originated in logic, where two propositions are said to be contradictory if the truth of one implies the falsity of the other and are said to be contrary if only one proposition can be true but both can be false.

uyiruLLa 'alive' – *cetta* 'dead' (Contradictory terms)

kuNTaana 'fat' – *melindta* 'thin' (Contrary terms)

Contraries are gradable adjectives, contradictories are not. Gradation therefore must also be considered as a semantic relation organizing lexical memory for adjectives.

cuuTu 'warmth'	*vayatu* 'age'
kotikkiRa 'very hot' *cuuTaana* ' hot' *vetuvetuppaana* 'warm *iLanjcuuTaana* 'warm' *kuLirndta* 'cold'	*vayataana* 'old' *ndaTuttara vayataana*'middle aged' *iLamaiyaana* 'young'

For some attributes gradation can be expressed by ordered strings of adjectives, all of which point to the same attribute noun in onto-thesaurus.

3.5.3.5. Markedness in Adjectives

Markedness (Miller, 1998:53-54) is an important property found among adjective. Binary oppositions frequently have a marked term and an unmarked term. That is, the terms are not entirely of equivalent weights, but one (the unmarked one) is neutral or positive in contrast to the other. Marked/unmarked distinction is found in polar oppositions such as the following:

uyarndta 'high'/*taazndta* 'low'

vayataana 'old'/*iLamaiyaana* 'young'

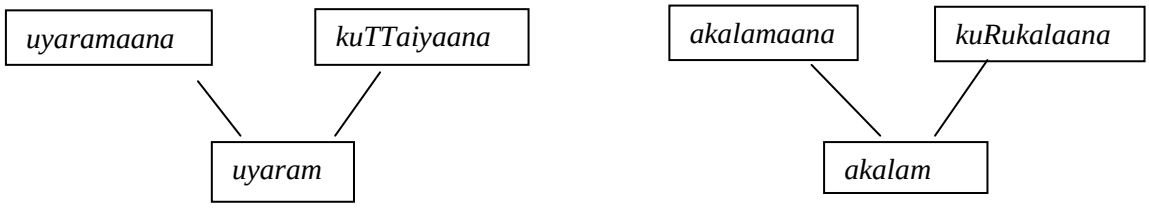
niiLamaana 'long'/*kuTTaiyaana* 'short'

akalamaana 'wide'/*kuRukalaana* 'narrow'

We measure things by *uyaram* 'height' rather than *kuTTai* 'shortness'. While asking questions about *uyaram* 'height', we say *atu evvaLavu uyaramaana tuuN* 'How high that pillar is?'rather than *atu evvaLavu kuTTaiyaana tuuN* 'How short that pillar is?'. A question *X evvaLavu kuTTaiyaanatu* 'How

short is X?' is felt to contain the assumption that X is short, while no equivalent assumption is present in *X evvaLavu uyaramaanatu* 'How high is X?' That is, if the two antonyms contrast with reference to a scale of measurement, the unmarked one is capable of referring to a point on that scale, thereby neutralizing the contrast. Thus the primary member, *uyaramaana* 'high' is the unmarked term; the secondary member, *kuTTaiyaana* 'short' is the marked one. They are related to the attribute noun *uyaram* 'height'. VOT captures the relation between marked and unmarked terms and their cross reference to their variable property.

Figure 33



3.5.3.6. Polysemy in Adjectives

Polysemy is found among adjectives (Miller, 1998: 54-56) as a limited number of adjectives are used to attribute a considerable number of nouns. For example, the use of *nalla* in the following phrases illustrates the polysemous nature of it. The semantic interpretation of adjectives depends on the head noun they modify. Many adjectives take on different meanings when they modify different nouns. The following example will exemplify this statement.

nalla kaalam 'good time'

nalla naaNayam 'good coin'

nalla naNpan 'good friend'

nalla ceruppu 'good sandal'

Adjectives are choosy about the nouns they modify. The general rule is that if the referent denoted a noun does not have attribute whose value is expressed by the adjective, then the adjective-noun combination requires a figurative or idiomatic interpretation. For example, *caalai* 'road' can be long because roads have LENGTH as an attribute, but stories do not have LENGTH, so *niiNTa* 'long' does not admit literal readings, but admits idiomatic interpretation (Miller, 1998). The selectional preferences of the adjectives are captured in VOT by suitably organizing them.

3.5.3.7. Reference-modifying and Referent-modifying Adjectives

Distinction has to be drawn between reference modifying and referent-modifying adjectives (Bolinger, 1967). For example *pazaiya* 'old' in the phrase *en pazaiya ndaNpan* 'my old friend' does not refer the referent who is a person as old, but attributes the friendship as old, whereas *pazaiya* in *pazaiya paattiram* 'old vessel', *pazaiya* attributes directly the vessel itself. Similarly, in the following phrase, both the adjectives attribute the quality of being criminals and the quality of being ministers respectively, rather than the persons.

neRRaiya kurravaaLikaL inRaiya mandtirikaL

'yesterday's criminals are today's ministers'

Some reference modifying adjectives may have direct antonyms as in the case of descriptive adjectives.

neRRaiya 'past' vs. *innaaLaiya* 'present'

mundtaiya 'past' vs. *inRaiya* 'present'.

3.5.3.8. Colour Adjectives

Colour terms have to be given different treatment (Miller, 1998: 54-57). They need to be organized differently than other adjectives in VOT. They can be both nominal as well as adjectival. As adjectives, they can be graded and conjoined with other descriptive adjectives. But they differ from the descriptive adjectives as the pattern of direct and indirect anotonymy does not hold good for colour adjectives. Only one colour attribute is clearly described by direct antonyms: LIGHTNESS, whose polar values are expressed by light/drark. In VOT, however, the opposition '*niRamuLLa*/*niRamaRRa* 'colored/colorless' is used to introduce the names of colours.

3.5.3.9. Relational Adjectives

Relational adjectives (Miller, 1998: 59-60) include of a large and open class of adjectives. Relational adjectives can be defined by using the phrase 'of, relating/pertaining to or associated with some noun', and they play a role similar to that of a modifying noun. For example, *cakootara* 'fraternal', as in *cakootra paacam* 'fraternal love' relates to *cakootaran/cakootari* 'brother/sister', and *poruLaataara* 'economical', as in *poruLaataara eRRa taazvu* 'economical difference', is related to *poruLaataaram* 'economics'. As far as Tamil is concerned noun form is used mostly in the place of relational adjective in English. For example,

icaik karuvi 'musical instrument'

paR cuttam 'dental hygiene'

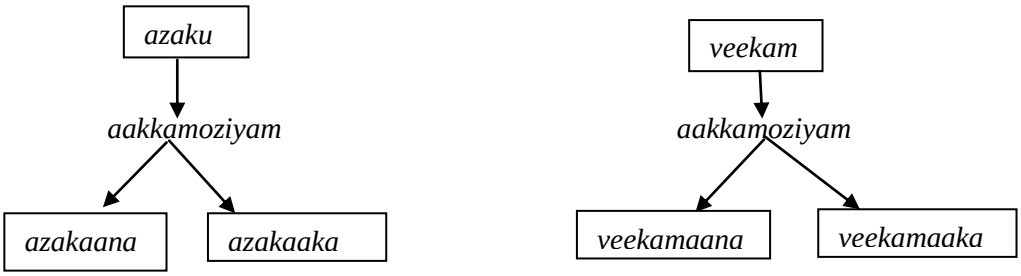
Since relational adjectives do not have antonyms, they cannot be incorporated into the clusters that characterize descriptive adjectives. And because their syntactic and semantic properties are a mixture of those of adjectives and those of nouns used as noun modifiers, rather than attempting to integrate them

into either structure VOT treats relational adjectives separately with cross references to the corresponding nouns.

3.5.3.10. Adverbs

Most of the adverbs are derived from nouns by adding suffix in the similar way adjectives are derived from nouns. The derived adjectives and adverbs need to be linked with the nouns form which they are derived. For example *aazak-aana* 'beautiful' and *azak*-aaka are derived from the noun *azaku* 'beauty'; similarly *veekam-aana* 'fast (adj.)' and *veekam-aaka* 'fast (adv.)' are derived from the noun *veekam* 'speed'. This is captured in VOT by linking the derived forms with the noun by the relation *aakkamoziyam* 'derivative relation'

Figure 34



The derived adjectives and adverbs inherit the semantic property of the noun form which they are derived. The semantic organization of adverbs is simple and straight forward. There is no tree structure (Miller, 1998:61).

3.3.5.11. Summing up of Relations in Abstracts

Relations	Subtypes	Example
Synonymy		*tukkam* 'sorrow' to *tunpam* 'sorrow', *cangkaTam* 'sorrow', *tuyaram* 'sorrow'
Hypernymy-Hyponymy		*uNarcci* 'feeling' to *makizcci* 'happiness'
Hyponymy-Hypernymy		*paccai* 'green' to *niRam* 'colour '
Holonymy-Meronymy	Wholes to parts	*vaaram* 'week' to *naaL* 'day'

Meronymy-Holonymy	Parts to wholes	*injcu* 'inch' to *aTi* 'feed'
Binary Opposites	Antonymic (gradable)	*nalla* 'good' to *keTTa* 'bad'
,,	Temporal Relations	*munnar* 'before' to *pinnar* 'after'
,,	Orthogonal or perpendicular opposition	*vaTakku* 'north' to *kizakku* 'east' and *meeRku* 'west'
,,	Antipodal Opposition	*vaTakku* 'north' to *teRku* 'south'
Multiple opposites	Serial	*onRu* 'one', *iraNTu* 'two', *muunRu* 'three', *ndaanku* 'four'
,,	Cycle	*njaayiRu* 'Sunday' to *tingkaL* 'Monday' .. to *cani* 'Saturday'

Relations	**POS linked**	**Example**
Antonymy (gradable i.e. contrary)	Adjective-adjective	*azakaana* 'beautiful': *kuruurmaana* 'ugly'
Antonymy (non-gradable i.e. contradictory)	Adjective-adjective	*uyiruLLa* 'alive': *cetta* 'dead'
Derivational	Adjective-noun	*azakaana* 'beautiful': *azaku* 'beauty'
Attributive	Noun-adjective	*vaTivam* 'size': *cinna* 'small'
Relational	Adjective-noun	*poruLaataara* 'economical': *poruLaataaram* 'economy'
Similarity	Adjective-adjective	*paaramaana* 'heavy': *kanamaana* 'heavy'
Derivational	Noun-adjective/adverb	*azaku* 'beauty': *azakaana* 'beautifull', *azakaaka* 'beautifully'
Similarity	Adverb-adverb	*veekamaaka* 'fast' *: viraivaaka* 'fast'

3.5.4. Organization of Relational in VOT

Nida listed relational concepts under the headings spatial, temporal, deictic, logical, etc. VOT has also followed his approach.

3.6. User-friendly Interface for Accessing VOT

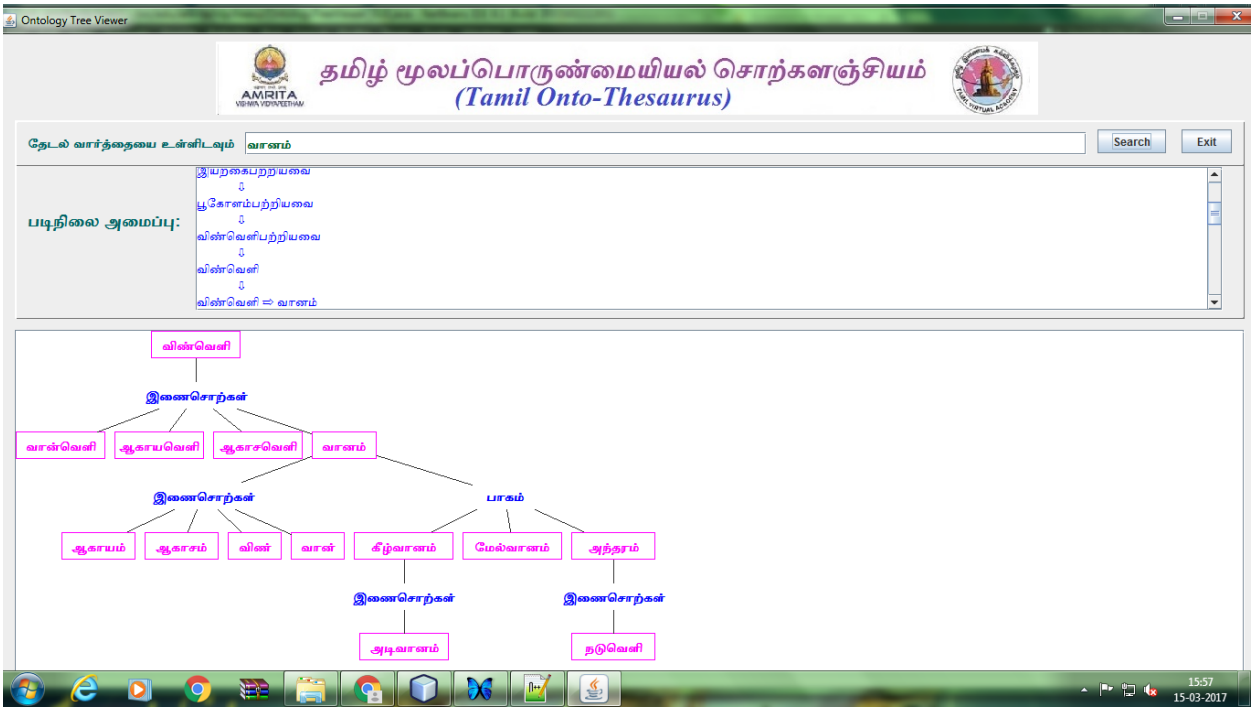
A few user friendly interfacs have been prepared from which one access the infomration needed form the onto-thesaurus of Tamil.

3.6.1. Tree Viewer for VOT

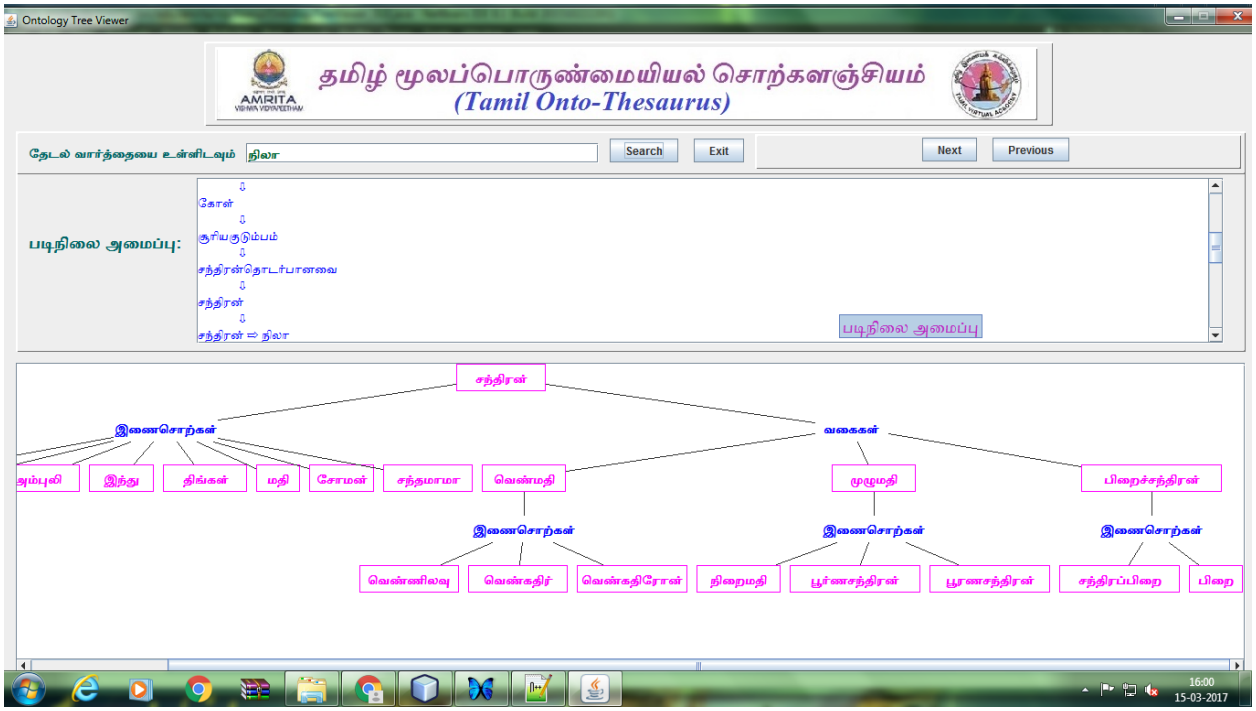
NLP or Linguistic researchers who work in syntax often want to visualize parse trees or create linguistic trees for analyzing the structure of a language. TreeViewer software provides an easy to use interface to visualize or create simple linguistic trees. This software is written entirely in Java. This tree viewer has been converted to depict the ontological structure of Tamil vocabulary. The semantic relations such as synonymy, hyponymy-hypernymy, meronymy-holonymy, oppositions, entailments, etc are captured by the tree viewer. The tree viewer gives the meaning of a given word in a hierarchical fashion as given below. The tree representation is converted into an ontology based visual thesaurus.

3.6.2 Sample Tree Structure of VOT

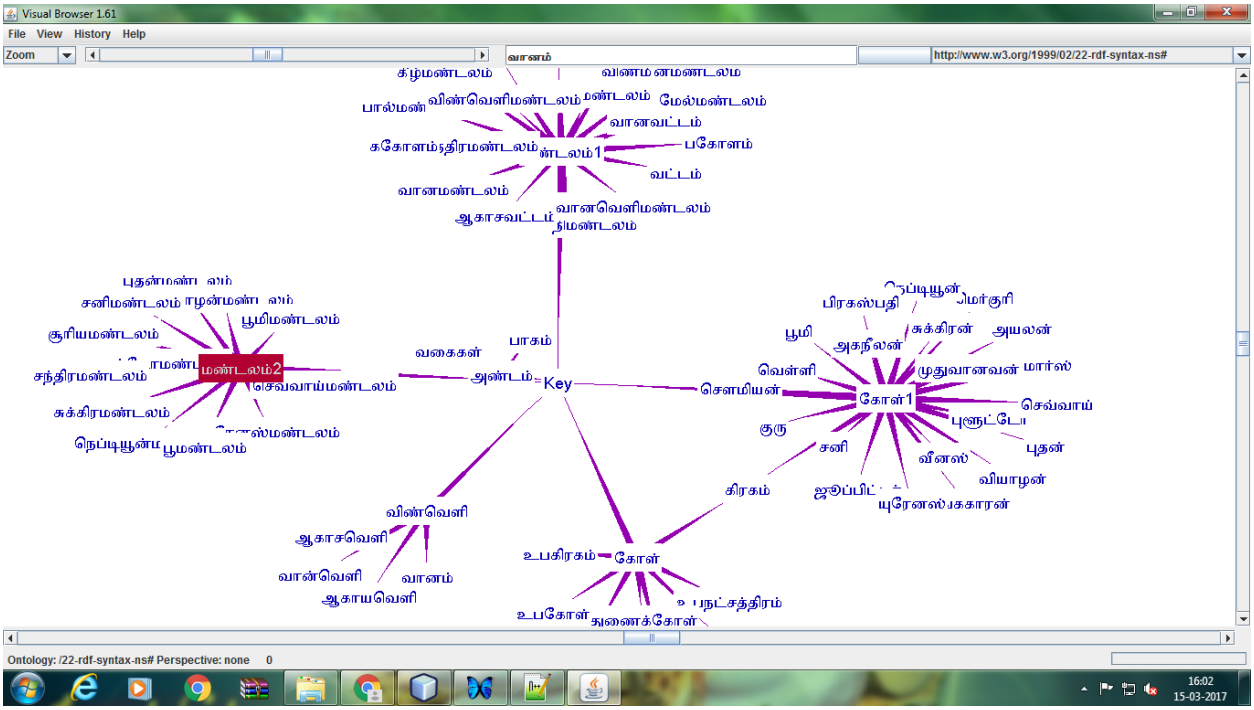
The samples of snap shots given below exemplify the working nature of VOT.



The GUI gives the hierarchical details of a lexical item for which the query is made as shown in the following screen shot.



The tree structure is converted into visual thesaurus as exemplified below:



4. Conclusion

The coverage of the vocabulary at present is only 50000 lexical items. We hope to improve on it in the near future. We like to accommodate all kinds of lexical and meaing relations or linkages a user expects from VOT. All the information availabel to a word and a set of words will be incorporated in VOT. The present onto-thesurus system will be converted into a generic system so as to accommodate all the other Dravidian languaes. Such a sort of onto-thesaurus will have wide range of uses which inclue information retrieval across Dravidian languages, machine translation across Dravidian languages and building knowledge based systems for Dravidian languages.

==========

References

Bhattacharyya, P. 2010. "IndoWodNet." In: N.S. Dash, P. Bhattacharyya and J.D. Pawar. (eds.). 2010. The wordNet in Indian Languages. Springer.

Bolinger, D. 1967. "Adjectives in English: Attribution and Predication." Lingua 18.1-34.

Bolinger, D. 1972. Degree Words. The Hague: Mouton.

Breslin, J.G., Passant, A. and Decker, S. 2009. Social Semantic Web. Springer, Verlag, Berlin, pp. 58.

Busa, F., Calzolari, N., and Lenci, A. 2001. "Generative Lexicon and SIMPLE Model: Developing Semantic Resources for NLP." In: Bouillon P. and Busa, F. (ed.) The language of Word Meaning, Cambridge University Press, Cambridge, pp. 333-362.

Busa, F., Calzolari, N., Lenci, A. and J. Pustejovsky. 2001. "Building a Semantic Lexicon: Structuring and Generating Concepts." In: H. Bunt et al. (ed.) Computing meaning, Volume 2, 29-51, Kluwer Academic publishers, Netherlands.

Cruse, D.A. 1986. Lexical semantics. Cambridge University Press, Cambridge.

Cruse, D. A. 2000. Meaning in Language: An Introduction to semantics and pragmatics. Oxford University Press, Oxford.

cuppiramaNiyak kaviraayar. (muulappatitppu 1930*, maRupatippu* 1985*). naamatiipanikaNTu. (uraiyaaciriyar: vaiyaapurippiLLai.) tamizp palkalaikkazakam, tanjaavuur*.

Dixon, R.M.W. 1982. Where have all the adjective gone? Mouton Publishers, Berlin

Gruber, T. R. 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Systems Laboratory, Computer Science Department Stanford University Stanford, California.

iraaceendtiran, ca. 2001. *taRkaalat tamiz coRkaLanjciyam* (a modern Thesaurus for Tamil). Tamil University, Thanjavur.

iraaceendtiran ca. & *paaskaran* ca. 2006 *tamiz mincoRkaLanjciyam* (an electronic Thesaurus for Tamil). Tamil University, Thanjavur.

Fellbaum, C. (ed.). 1998. WordNet: An Electronic Lexical Database. The MIT press, Cambridge, Massachusetts.

Fellbaum, C. 1998. “A Samantic Network of English verbs.” In: Fellbaum, C. (ed.). WordNet: An Electronic Lexical Database. The MIT press, Cambridge, Massachusetts, pp. 69-104.

Jones, K.S. 1986. Synonymy and Semantic Classification. Edinburgh University Press, Edinburgh.

Kashyap, L., S.R. Joshi, S.R. and P. Bhattacharyya. 2010. "Insights on Hindi WordNet Coming from the IndoWordNet." In: N.S. Dash, P. Bhattacharyya and J.D. Pawar. (eds). 2010. The wordNet in Indian Languages. Springer.

Leech, G.N. Semantics. Harmondsoworth: Penguin.

Lehrer, A.J. 1974. Semantic Fields and Lexical structure. Amsterdam: North Holland.

Lyons, J. 1977. Semantics, volume I and II. Cambridge University Press, Cambridge.

Mawson, C.O.S. 1956. Roget's International Thesaurus of English Words and Phrases. Pocket Books, INC., New York.

Miller, K.J. 1998. “Modifiers in WordNet.” In: Fellbaum, C. (ed.). WordNet: An Electronic Lexical Database. The MIT press, Cambridge, Massachusetts, pp. 47-67.

Miller, G.A. Beckwith, R., Fellbaum, C., Gross, D., and Miller, K.J. 1990. “Introduction to WordNet: An on-line lexical database.” International Journal of Lexicography, 3, 235-244.

Miller, G.A. 1991. Science of Words. New York: Scientific American Library.

Miller, G.A. 1998. “Nouns in wordNet.” In: Fellbaum, C. (ed.). WordNet: An Electronic Lexical Database. The MIT press, Cambridge, Massachusetts, pp. 23-46.

Nida, E.A. 1975a. Componential Analysis of Meaning: An Introduction to Semantic Structure. Mouton, The Hague.

Nida, E.A. 1975.b. Exploring Semantic Structure. The Hague: Mouton.

Pustejovsky, J. 1995. The Generative Lexicon. MIT Press, Cambridge, MA.

Pribbenow, S. 2002. “Meronymic relationships: From classical mereology to complex part-whole relations.” In: Green, R., Bean, C.A. and Myaeng, S.H. (ed,). 2002. The semantics of relationships: An Interdisciplinary perspective. Kluwer Academic publishers, Dordrecht/Boston/London, pp. 35-50.

Rajendran, S. 1978. Syntax and Semantics of Tamil Verbs. Ph.D. Thesis. University of Poona, Poona.

Rajendran, S.1982. Semantics structure of Tamil Vocabulary. Post-doctoral Research Report. Deccan College, Poona.

Rajendran, S. 2001. *taRkaalat tamizc coRkaLanjciyam* [Thesaurus for Modern Tamil]. Tamil University, Thanjavur.

Rajendran. S. 1983. Semantics of Tamil Vocabulary. Report of the UGC sponsored postdoctoral work (manuscript). Deccan College Post-Doctoral Research Institute, Poona.

Rajendran, S. 1995. "Towards a Compilation of a Thesaurus for Modern Tamil". South Asian Language Review. 5.1:62-99.

Rajendran, S. 2002. "Preliminaries to the preparation of Wordnet for Tamil." Language in India 2:1, March 2002, www.languageinindia.com

Rajendran, S. 2009. "Dravidian WordNet." In: Proceedings of Tamil Internet Conference 2009. Cologne, Germany, October, 2009

Rajendran, S. 2010. "Tamil WordNet." In: Proceedings of the Global WordNet Conference (GWC 10) 2010, IIT, Bombay.

Rajendran, S. 2016. "Tamil Thesaurus to WordNet." In: Conference Papers of 15th Tamil Internet Conference 2016. International Forum for Information Technology in Tamil, September 2016, 1-9.

Rajendran, S., S. Arulmozi, B. Kumara Shanmugam, S. Baskaran, and S. Thiagarajan. 2002. "Tamil WordNet." In: Proceedings of the First International Global WordNet Conference. Mysore: CIIL, 271-274.

Rajendran, S. and Baskaran, S. 2002. Electronic Thesaurus for Tamil. In: Proceedings of the International Conference on Natural Language Processing. NCST, Mumbai, 2002.

Rajendran, S. and S. Baskaran. 2006. *Tamizh mincoRkaLanjciyam* [Electronic thesaurus for Tamil]. Tamil Universtiy, Thanjavur.

Rajendran,S., Shivapratap G, Dhanalakshmi V and Soman K.P. "Building a WordNet for Dravidian Languages." In: Proceedings of the Global WordNet Conference (GWC 10), 2010 IIT Bombay.

Saint-Dizier, P and Viegas, E. 1995. "An introduction to lexical semantics from a linguistic and a psycholinguistic perspective," In: Saint-Dizier, P and Viegas, E. (ed.) Computational lexical semantics. Cambridge: Cambridge University Press, pp. 1-29.

Roget, P.M. 1852. Thesaurus of English Words and Phrases. London.

Smith, B. 2003. "Ontology." In: Luciano Floridi (ed.), Blackwell Guide to the Philosophy of Computing and Information. Blackwell, Oxford, pp. 155-166.

Sowa, J.F. 1984. "Conceptual Structures: Information Processing in Mind and Machine". Addison-Wesley Publishing Company, Reading, Massachusetts.

Sowa, J.F. 2009. Building, sharing, and merging ontologies (modified on 18.01.2009). http://www.jfsowa.com/ontology/ontoshar.htm

tivaakaram. 1958. *ceentan tivaakarm. kazakam, cennai*.

Vossen, P. (ed.) 1998. EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Dordrecht: Kluwer Academic Publishers.

==========

Dr. Rajendran Sankaravelayuthan
s_rajendran@cb.amrita.edu

Dr. Anandkumar, M.
m_anandkumar@cb.amrita.edu

Centre for Computational Engineering and Networking (CEN)
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham
Amrita Universtiy
India