INTERNATIONAL JOURNAL OF ENGLISH LANGUAGE, LITERATURE AND TRANSLATION STUDIES (IJELR)
A QUARTERLY, INDEXED, REFEREED AND PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
http://www.ijelr.in

RESEARCH ARTICLE

Vol. 4. Issue.2., 2017 (April-June)

INTERNATIONAL STANDARD SERIAL NUMBER INDIA
2395-2628(Print):2349-9451(online)

BUILDING ONTOLOGY OF TELUGU VOCABULARY USING MACINE REDABLE DICTIONARIES OF TELUGU

SAMPATH BABU TOKALA
Doctoral Scholar, Centre for ALTS, School of Humanities, University of Hyderabad
Hyderabad
tsampathbabu@gmail.com

ABSTRACT
Dictionary possesses relevant information about the lexical items it describes. This is true to the machine readable dictionary too. They are supposed to describe lexical items without any ambiguity, taking into account its relations with other lexical items such as synonymy, antonymy, hyponymy and meronymy. A dictionary makes use of these relations to describe a lexical item explicitly making use of other lexical items. For example, *deer* is described as a type of animal and *parrot* is described as type of bird, *low* described as antonymous to *high*, *leg* is described as a part of a body and so on. This way of description provides us the ontological information about the lexical items. One can extract ontological information from dictionaries if the dictionary is properly written to capture the meaning of lexical items by providing its relations with other lexical item. It must be informative. A dictionary should describe a book as a concrete object made up of paper bound into an object which contains information meant for reading and is written by a writer or edited by an editor and is published by somebody. By this description we can infer that one can read a book as it contains information, one can write or edit a book, one can publish a book and one can tear a book as it is made of paper and so on. The present paper attempts to build an ontology of Telugu vocabulary by making use of machine readable dictionaries.

Keywords: Synonymy, Hypernomy, Hyponomy, Meronomy, Holonymy, Antonymy, Homonymy and Polysemy

1. INTRODUCTION

A machine-readable dictionary (MRD) is a dictionary in an electronic form which can be made use of getting all sorts of information a dictionary can provide by means of queries in the use friendly interface. A application software connected with the dictionary helps us to make such queries and get the necessary or needed information. It may be a single language explanatory dictionary or a multi-language dictionary to support translations between two or more languages or a combination of both. Translation software between multiple languages usually apply bidirectional dictionaries. An MRD may be a dictionary with a proprietary structure that is queried by dedicated software (for example online via internet). It can also be a dictionary that has an open structure and is available for loading in computer databases and thus can be used via various

software applications. Conventional dictionaries contain a lemma with various descriptions. An MRD may have additional capabilities and is therefore sometimes called a smart dictionary. An example of a smart dictionary is the Open Source Gellish English dictionary.

The term dictionary is also used to refer to an electronic vocabulary or lexicon as used for example in spelling checkers. If dictionaries are arranged in a subtype-supertype hierarchy of concepts (or terms) then it is called ataxonomy. If it also contains other relations between the concepts, then it is called an ontology. Search engines may use either a vocabulary, taxonomy or an ontology to optimize the search results. Specialized electronic dictionaries are morphological dictionaries or syntactic dictionaries.

In general usage, a thesaurus is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms), in contrast to a dictionary, which provides definitions for words, and generally lists them in alphabetical order. The main purpose of such reference works for users "to find the word, or words, by which [an] idea may be most fitly and aptly expressed" – to quote Peter Mark Roget, architect of the best known thesaurus in the English language. Although including synonyms, a thesaurus should not be taken as a complete list of all the synonyms for a particular word. The entries are also designed for drawing distinctions between similar words and assisting in choosing exactly the right word. Unlike a dictionary, a thesaurus entry does not give the definition of words.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 ROGETS' THESAURUS OF ENGLISH WORDS AND PHRASES (1852). This is a crude distinction: thesaurus frequently contains synonyms, and synonymous dictionaries are thesauri, if the groups in them are large and cross-referenced and reflect a wide interpretation of synonymy (Jones, 1986).

> In the recent years, MRDs have been made use of as lexical resources for various purposes (Wilsk et al, 1996:161-181). Extensive linguistic information on lexical semanticsis available in MRDs. MRDs give scope for semantic analysis and we can harvest many types of semantic information encoded in it by professional lexicographers. It is a well-known fact that a wealth of implicit information lies within dictionaries. This lexical information can be made explicit for use by computational means. In dictionaries, the lexicographers define words in terms of other words, which themselves have definitions elsewhere in the dictionary. The computational linguists have to “unpack” the meanings that have between encoded in dictionaries by lexicographers by converting them into a computationally tractable form. The motivation for creating lexical database in this way is to take advantage of the extensive lexicographic work that goes into creating the dictionary in the first place.

Lexical semantics and in artificial intelligence elucidate ontology. Their borders are to some extent unclear. An ontology as a formal system aims at representing the different concepts and their related linguistic realizations of a given domain by means of basic elements. At its large vista an ontology can also incorporate in its fold different forms of encyclopaedic knowledge about the domain, and common-sense knowledge as well as rhetorical and metaphorical knowledge and expressions. Ontologies manifest a significant link between knowledge representation and computational lexical semantics. Ontologies are widely used as formal devices to manifest the lexical content of words. Ontology perform a vital role in various language engineering (LE) tasks such as content-based tagging, word sense disambiguation, multilingual transfer, etc. explicate elucidate illuminate

The skeletal structure of ontology is available in Nida's componential analysis of meaning (1975). This paper aims to explain building an ontology of Telugu vocabulary by making use of the skeletal structure available in Nida (1975) and by extracting ontological relations from MRDs of Telugu.

2. Ontology

Ontology is the philosophical study of the nature of being, becoming, existence or reality, as well as the basic categories of being and their relations. Traditionally listed as a part of the major branch of philosophy known as metaphysics, ontology often deals with questions concerning what entities exist or may be said to exist and how such entities may be grouped, related within a hierarchy, and subdivided according to

similarities and differences. Although ontology as a philosophical enterprise is highly hypothetical, it also has practical application in information science and technology, such as ontology engineering**.**

The definition of ontology is a very delicate task. The main point is to be able to define real primitive or basic elements of knowledge associated to the domain. This definition is based on several factors, among which are (Saint-Dizier and Viegas: 20):

- the degree of granularity one wants to attain in the system.
- the reference to already existing partial ontologies which must homogeneously be integrated into the system (e.g. spatial or temporal ontologies)
- the reference to theories of linguistic knowledge representation, such as the Lexical Conceptual Structure (LCS) or the Conceptual Graphs framework)

An ontology based on a formal language composed of (Saint-Dizier and Viegas, 20) :

- sets of entities, often structured and typed
- a set of relations and operations and their related properties
- a finite set of basic predicates describing primitive states and actions of the domains of interpretation within the application and
- a set of functions operating on the entities. These functions are usually defined from a corpus analysis.

This language is usually a sub-set of first-order logic. It is used to represent linguistic as well as application-dependent knowledge at various levels of generality.

Ontologies proved to be extremely useful is the representation of lexical knowledge. This is the main reason for their renewed interest in lexical semantics and natural language processing (NLP). Representing the meaning of a word minimally implies (i) distinguishing it from 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(Busaet la 2001:31). 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 because 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. Finally, resemblances between word senses will correspond to the sharing of the same ontology type.

Hyponymy and its consequence taxonomy are the fundamental building blocks of ontology. Hyponymy and its natural partner, incompatibility, are described by Lyons as “the most fundamental paradigmatic relations of sense it terms of which the vocabulary is structured” (Lyons, 1968: 453). Lyons states that taxonomic lexical hierarchies are structured by the relations of hyponymy and incompatibility.

Meronymic or partonomic relations are ontological relations that are considered as fundamental as the ubiquitous, taxonomic subsumtions relationship (Pribbenow, 2002:35).There is no doubt about the importance of parts for human cognition. Parts and their relations to each other play an important role in visual and auditional perception. Our vocabulary contains a variety of words for different kinds of part or ways of portioning a whole. The part-whole relationship is important for natural language.

A number of situations and classes of applications make use of ontological descriptions. Ontological knowledge has been integrated in several systems to enhance their excellence and generalization. Applications such as intelligent user-front ends in natural language and automatic retrieval of documents make use of ontological knowledge. A number of practical and theoretical approaches like the following makes use of ontological knowledge implicitly:

- Semantic networks and conceptual graphs

- Lexical conceptual Structure
- Generative lexicons
- Scenarios and scripts and
- Discourse management systems.

2.1 Ontology of Aristotelian Origin

The history of scientific taxonomy starts with Aristotle and is predicated on a first philosophy of essentialism. Taxonomic principles were laid out both in Aristotle's work on natural history and in his logic. In the logic, taxonomy or the division of things into genera and species is a way of classifying predicates; it is a refinement of the ten basic categories of predicates. In the long tradition of Aristotelian logic, the basic idea was developed in detail. Classification of the various categories, say, of substance, took the form of branching tree diagrams that specified the various genera and species of the category. In Aristotle's natural history, as opposed to his logic, the taxonomic notions of genus and species were developed to handle relations between things, primarily animals.

There are two sources for Aristotle's notions of classification: 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. 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. He 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 1 shows (Sluagheter, 1982:29)

Logical Classification of the Category of Substance

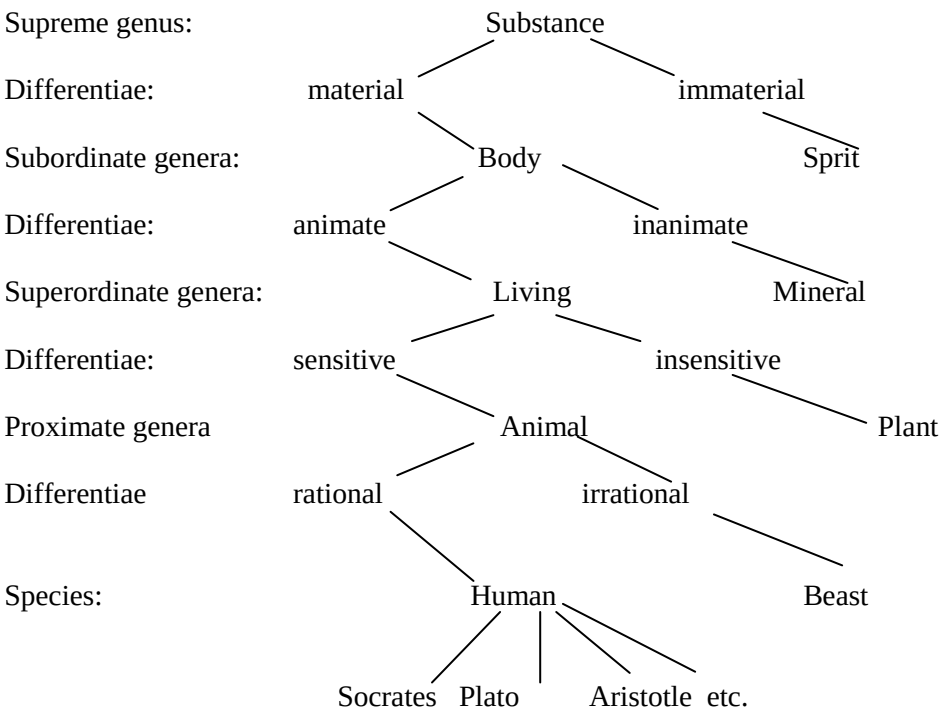

Figure 1

By the sixteenth and seventeenth centuries what we find in Aristotelian logic books is further divisions in each category and in these divisions and tables is included the philosophical and scientific lore of the age. The examples of Bunddeville (figure 2) (Slaughter, 1982:30). and Du Moulin (figure 3) are typical of the encyclopedic nature of these logics(Slaughter, 1982:31).

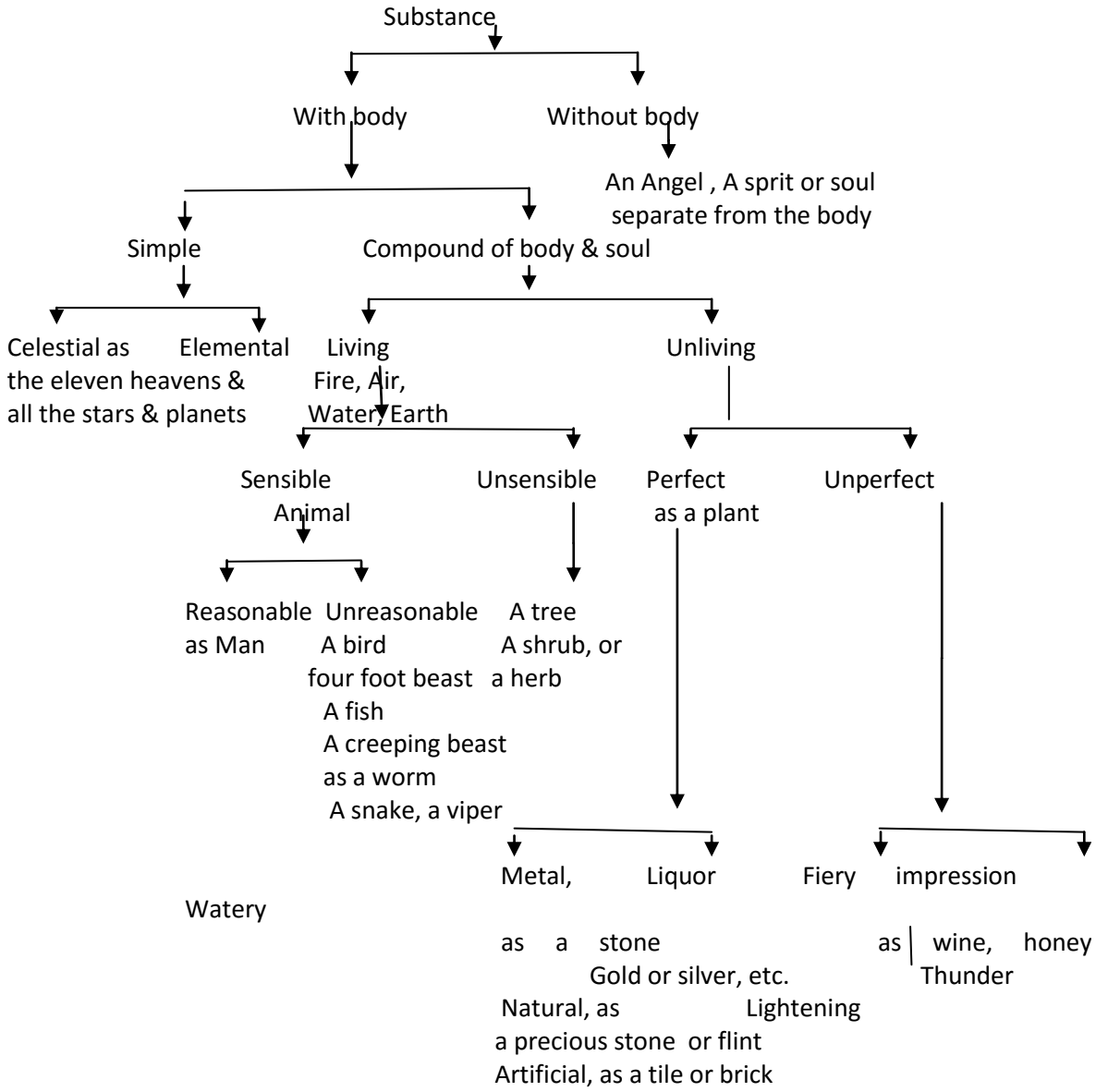

Figure: 2

Originally, the Aristotelian world view, retained and popularized in the concept of the Great Chain of Being, encompassed 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.

Aristotle had posited that animate and inanimate natures are two fundamentally different things. Decontextualization and universalization of the words or concepts were attempted. This leads to the development of a scientific (botanical) taxonomy. The following levels of taxa are found (Slaughter, 1982:55):

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

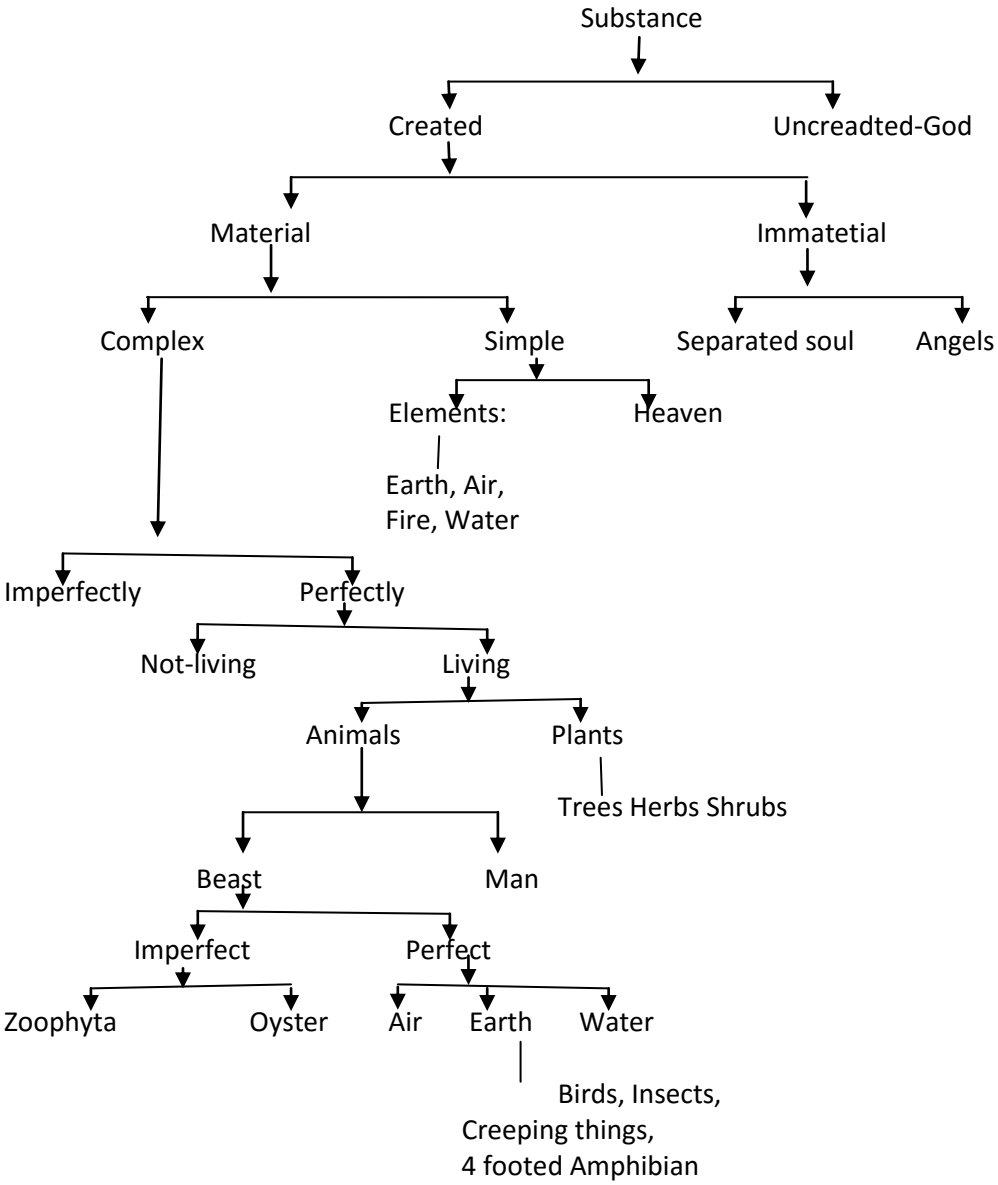

Figure 3

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).

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

2.2 Ontology in thesaurus

Thesaurus which tries to classify lexical items under meaning-wise headings, furnish us with an ontology for classifying the vocabulary of a language. Roget's thesaurus provides us a very interesting ontology which is given below:

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	

2.3 Classification in Dictionary (nigantuvu**)**

Attempts to classify vocabulary of a language are found in dictionaries (nigantuvulu***)***, which lay foundation for the compilation of thesauri or thesaurus dictionaries. The following classification is found in dictionary (nigantuvu***)*** of Telugu

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:

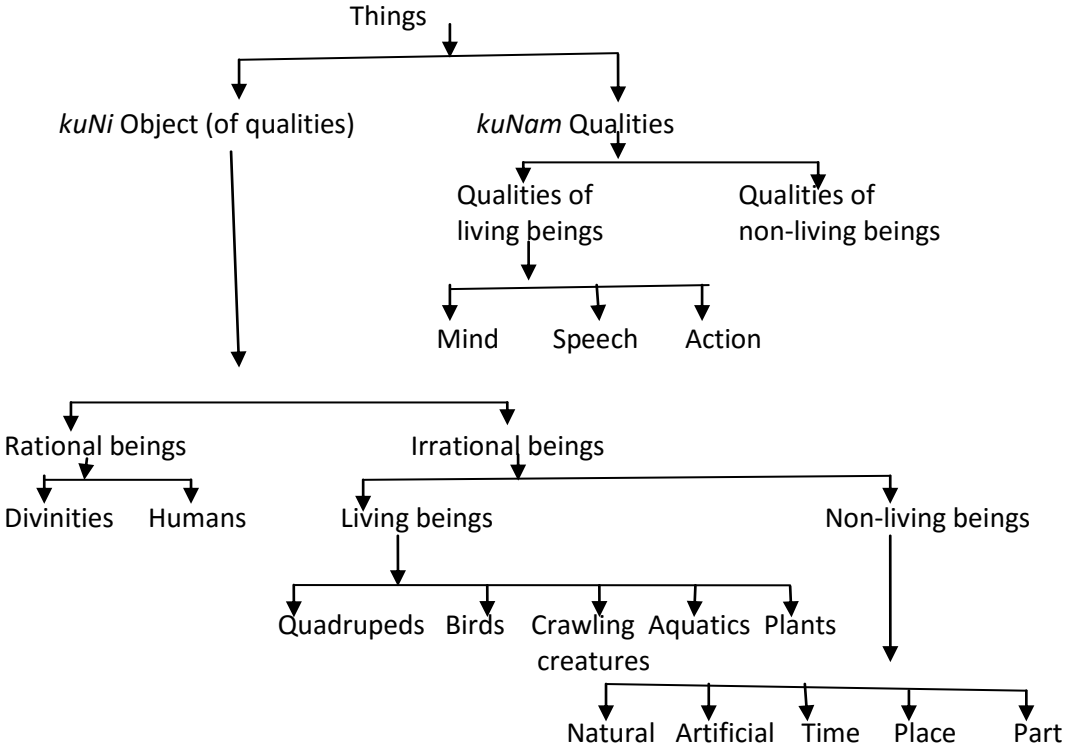

Figure 4

2.4 Ontology in Nida's Thesauri Dictionary

Nida (1975a) who was concerned with the preparation of a thesauric dictionary for Greek gives the following as the tentative hierarchical classification of the lexical items (Nida: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. Relational: 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.

2.5 Ontology and 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 has led to the development of field theories. 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. 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 (1977) 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. Synonymy can be defines 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. Word contrast in a taxonomy are incompatible. If A and B are incompatible, then A $\supset$ not B and B $\subset$not A.

Animal			
Dog	Cat	Horse	sheep

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 are 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.

2.6 Ontology and 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. 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.

It is necessary to build a knowledge base that includes hierarchical information in order to process language effectively. It is not difficult to argue that the knowledge base should 'know' facts like a poodle IS-A dog, a dog IS-A mammal, and a mammal IS-A animal. In the knowledge base, poodles will have all the properties of dogs, and dogs will have all the properties of mammals, etc. and although there are differing opinions about whether this knowledge is inferred at that time of processing or inferred earlier and stored in the knowledge base, this is nonetheless crucial information which must be available for language processing. In general, the network/hierarchy efforts are more narrowly devoted to detailing hypernymic and hyponymic relations (at other times referred to as IS-A relation or AKO [a-kind-of] relations) between words and word senses defined in the dictionary. These are meaningful (semantic) relations explicitly intended to support inheritance. Dictionaries provide a rich source from which we can extract this kind of information automatically on a large scale.

The notion of automatically converting information available in MRDs into a "machine-tractable" one (i.e. Machine Tractable Dictionary, shortly MTD) is the central theme of this paper. The present proposal becomes relevant in the context of building a WordNet for Telugu. Nida (1975) ontological classification will be used for building the basic framework for building ontology of Telugu vocabulary. This will be completed by the hierarchical and lexical relations available in a standard Telugu dictionary.

2.7 Ontology in WordNet

WordNet has been built for Telugu (Arulmozi, 2010) based on Hindi 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. 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.

3. Network of meaning relations

The semantic relations that have to be considered for building ontology for Telugu vocabulary are the following pardigmantic relations between words or concepts: Synonymy, Hypernomy, Hyponomy, Meronomy, Holonymy, Antonymy, Homonymy and Polysemy.

3.1 Synonymy

The lexical items which have the same meaning or which share same componential features are synonyms and the relationship existing between them is synonymy. Synonymy does not necessarily mean that the items concerned should be identical in meaning, i.e. interchangeable in all contexts. Synonymy can be said to occur if lexical items are close enough in their meaning to allow a choice to be made between them in some contexts, without there being any difference in the meaning of the sentence as a whole. Take, for examples, the words *pustakʌm* 'book' and *granḍʌm* 'book'. *aṭanu oka pustakʌm čaḍiva:du 'he read* a book' can entail *aṭanu granḍʌm čaḍiva:du* 'he read book'. The relation existing between *pustakʌm* and *granḍʌm* is synonymy.

3.2. Hyponymy-Hypernymy

Hyponymy is the relationship which exists between specific and general lexical items, such that the former is included in the latter. The set of terms which are hyponyms of same superordinate term are co-hyponyms. Take for example the lexical items *a:vu* 'cow' and *jantuvu* 'animal'. *iḍioka a:vu* 'this is a cow' and *iḍi oka barre* 'this is a buffalo' unilaterally entail 'this is an *animal*'. The relationship existing between *cow* and *buffalo* with animal is hyponymy and cow and *buffalo* are co-hyponyms.

One sense is a hyponym of another if the first sense is more specific, denoting a subclass of the other

'caru' *car* is a hyponym of *'rava:na bandi' vehicle*

'kukka' *dog* is a hyponym of *'jantuvu' animal*

'ma:midi' *mango* is a hyponym of *'pandu' fruit*

Conversely

'rava:na bandi' *vehicle* is a hypernym/superordinate of 'caru' *car*

'jantuvu' *animal* is a hypernym of *'kukka' dog*

'pandu' *fruit* is a hypernym of *'ma:midi' mango*

3.3. Meronymy-holonymy

The part-whole relation between nouns is generally considered to be a semantic relation, called meronymy, comparable to synonymy, antonymy, and hyponymy.The relation has an inverse: if X is a meronymy 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 'mukku' *beak* and 'rekka' *wing* are meronyms of 'pakʃi' *bird*, and if 'ka:nari pakʃi' *canary* is a hyponym of 'pakʃi' bird, then by inheritance, 'mukku' *beak* and 'rekka' *wing* must also be meronyms of 'pakʃi' *canary*.

The lexical items which overlap in terms of meaning and do not show systematic include-included relation and have some semantic traits in common, but differ in respect of traits that do not clash are said to

be compatible.Take for example the words 'kukka' *dog* and 'pempudu jantuvu' *pet*. A dog could be a pet, but neither all pets are dogs nor all dogs are pets. The relationship existing between *dog* and *pet* is compatible.

3.4. Incompatibility

Incompatibility refers to sets of items where the choice of one item excludes the use of all the other items from that set. Take for example the words *cat* and *dog*, 'it is a cat' can entail'it is not a dog'. The relation existing between 'pilli' *cat* and 'kukka' *dog* is incompatible. Both come under the superordinate term *animal*. Thus the incompatible items can be co-hyponyms of a superordinate item, that is items which are in incompatible can be related to one another by hyponymous relation. All kinds of oppositions can be included under incompatibility. If the opposition is between two lexical items, it is called binary opposition and if the opposition is between many lexical items it is called many-member opposition.

3.5. Consolidated relations for ontology development

The following tables will give the meaning relations that have to be considered for building ontology of Telugu vocabulary.

Table of lexical relations for Nouns

Relations	**Subtypes**	**Example**
Synonymy		pustakʌm 'book' granḍʌm 'book'
Hypernymy-Hyponymy		*jantuvu or pra:ni* 'animal' to *ksiradam or pa:liccu jantuvu* 'mammal'
Hyponymy-Hypernymy		*a:vu* 'cow' to *ksiradam or pa:liccu jantuvu* 'mammal'
Holonymy-Meronymy	Wholes to parts	*Me:d͡ʒa balla* 'table' to *ka:lu* 'leg'
,,	Groups to members	*sa:ka* 'department' to *a:ča:rjudu* 'professor'
Meronymy-Holonymy	Parts to wholes	čakram 'wheel' to *la:gudu bandi* 'cart'
,,	Members to groups	*jattu lekadalasa:radi* 'captain' *to sainjam or sena* 'army'
Opposites	Antonymic (gradable)	*manchi vjakti* 'good person' to *chedda vjakti* 'bad person'
,,	***Complementary***	*ra:tri* 'night' to *pagalu or ro:ju* 'day'
,,	Privative (opposing features)	*hetu viruddamaina* 'irrational' to *hetu baddamaina* 'rational'
,,	Equipollent (positive features)	*maga or* puruʃudu 'male' to *a:da or* s̱tre: 'female'
,,	Reciprocal Social roles	vaiḍudu 'doctor' to *rogi* 'patient'
,,	Kinship Relations	t̪alli 'mother' to kuma:rt̪e 'daughter'
,,	Temporal Relations	munḍu 'before' to t̪aruva:t̪a 'after'
,,	Orthogonal or perpendicular	ut̪t̪aram 'north' to t̪u:rupu 'east' and paccima 'west'
,,	***Antipodal Opposition***	ut̪t̪ara 'north' to ḍakʃina'south'
Multiple opposites	***Serial***	*okati* 'one' *rendu* 'two' *mu:du* 'three' *na:lugu* 'four'
,,	***Cycle***	*a:diva:ram* 'Sunday' *so:mava:ram* 'Monday' *saniva:ram* 'Saturday'
Lexical association	***Collocation***	*simham* 'lion' to *garginchu* 'roar'
,,	Morphological relations	čaḍuvu 'study' to čaḍuvu *kunna vyakti* 'educated man'
Compatibility		*kukka* 'dog' to pempudu 'pet'

Table of lexical relations for Verbs

Relations	Definition/sub types	Example
Synonymy	Replaceable events	niḏrapovu 'sleep' → niḏrapovu 'sleep'
Meronymy- Holonymy	Events to super-ordinate events	*eguru* 'fly' → *praja:nam* 'travel'
Troponymy	Events to their subtypes	naduču 'walk' → *kuntuta* 'limp'
Entailment	Events to the events they entail	*gurakapettu* 'snore' → niḏrapovu 'sleep'
,,	Event to its cause	*edugu* or *lechu* 'rise' → *paikettu* 'raise'
,,	Event to its presupposed event	*sa:dinchu* 'succeed' → *prajtnamceyu* 'try'
,,	Event to its implied event	*hatja* 'murder' → maranińču 'die'
Antonym	Opposites	penču 'increase' → ṯagginču 'decrease'
,,	Conversensess	*ammu* 'sell' → *konu* 'buy'
,,	Directional opposites	*prarambincu* 'start' → *čerukonu* 'reach'

Table of meaning relations for adjectives and adverbs

Relations	POS linked	Example
Antonymy (gradable i.e. contrary)	Adjective-adjective	*andamaina* 'beautiful': *chandalamaina* 'ugly'
Antonymy (non-gradable i.e. contradictory)	Adjective-adjective	*bratikivunna* 'alive': čanipoina 'dead'
Derivational	Adjective-noun	*andamaina* 'beacutiful': *andam* 'beauty'
Attributive	Noun-adjective	*kolata* 'size': činna 'small'
Relational	Adjective-noun	*a:rdikamaina* 'economical': *ardika* 'economy'
Similarity	Adjective-adjective	*peddamottam* 'heavy': *cha:la* 'heavy'
Derivational	Noun-adjective/adverb	*andamaina* 'beautifull', *azakaaka* '*andamaina': azaku*andam 'beauty'
Similarity	Adverb-adverb	*veganga* 'fast' *: vegam* 'fast'

The present proposal of extracting this information from MRD automatically is a challenge.

4. Extraction of semantic information form MRDs

Much of the information needed for our purpose can be found in a dictionary. The problem, of course, is that none of this comes without great effort. Every dictionary is printed in a different format, and every format is designed to be both information-bearing and pleasing to the eye. This combination makes it very difficult to remove or reformat the strings and symbols in the dictionary automatically, to decide, by program, what is intended by the lexicographer, what to keep, and what to leave out. Extraction of semantic information from dictionaries entails certain assumptions about the extent of knowledge in a dictionary, about where that knowledge is located, and how that knowledge can be extracted from the language of dictionary definitions. These assumptions are methodological assumptions because they underlie the decisions made in choosing one method for semantic analysis rather than another. These assumptions are about sufficiency, extricability, and bootstrapping (Wilks et al.1990, 1996:141-142).

Sufficiency addresses the issue of whether a dictionary is a strong enough knowledge base for Tamil, especially as regards linguistic knowledge and, above all, the knowledge of the real world needed for subsequent text analysis. Different positions have been taken toward sufficiency within computational lexicography. Some researchers believe that there is not enough knowledge in dictionaries in principle; in other words, that certain specific semantic information is not available anywhere in a dictionary and hence must be derived from another, outside, source. Other researchers believe that dictionaries may indeed contain sufficient knowledge, although that knowledge may be implicit, because that knowledge can be made explicit by using information from entries in other parts of a dictionary.

Extricability is concerned with whether it is possible to specify a set of computational procedures that operate on an MRD and extract, through their operation alone and without any human intervention, general and reliable semantic information on a large scale, and in a general format suitable for, though independent of, a range of subsequent NLP tasks.

Bootstrapping refers to a process of collecting the initial information that is required by a set of computational procedures that are able to extract semantic information from the sense definitions in an MRD. The initial information needed is commonly linguistic, normally syntactic and case information, which is used during the processing of dictionary sense definitions into an underlying representation form which semantic information is then extracted. Boost strapping methods can be internal or external. Internal bootstrapping methods obtain the initial information needed for their procedures from the dictionary itself and use procedures to extract that information. By contrast, external bootstrapping methods obtain the initial information for their procedures by some method other than the use of those procedures themselves. The initial information may be from some source external to the dictionary or may be in the dictionary but impossible to extract without the use of the very same information. There are differences of opinion in computational lexicology regarding extricability and bootstrapping.

The straight forward approach to creating an MTD is to segment the dictionary into a collection of records, one for each word sense, composed of a set of labeled constituent fields, where each filed is roughly equivalent to a single meaningful element in a sense definition. The procedure makes the contents of the dictionary explicit, and is the so-called lexical database approach for forming an MTD. The more difficult approach is to process the labeled constituent fields in some way, in order to retrieve the implicit knowledge that lies buried there. This is the semantic approach to creating an MTD.

As mentioned earlier, dictionaries resort to defining a word or concept by using some other word concept spelling out the meaning relations between them. Say for example, *book* will be defined as type of reading material;*knife* will be defined as a sort of instrument;*green* may be defined as a type of colour, or *leg* may be defined as a part of a body; *long* will be defined by associating it with length; *heavy* will be defined by associating it with weight; and so on. Thus dictionary definitions give amble scope for stabling various kind of sense or meaning relations and thereby help in building ontology of the vocabulary defined in the dictionary.

In addition to research into constructing network of genus terms form MRDs, various proposals in the past have suggested that different specialized link types should be added to the net-work. There are approaches to forming semantic networks from dictionary text using hand-coding have also been tried. The Princeton group (Beckwith et al. 1989) is now in the process of creating a network of nouns, verbs, and adjectives in accordance with psycholinguistic guidelines. The nodes in their work are synonym sets and its links are of many types. The nodes of their network are synonymy sets and its links are of many types. Within category of nouns there are both hypernym, hyponym, meronym (is-part), and holonym (reverse of meronym) links. Within the verb category are troponym (manner-of) and entailment (implicatgion) links. The following tables depicts various sense relations proposed in Telugu wordnet (Arulmozi,2010). In the Telugu wordnet, sense relations are established manually by using dictionaries.

4.1. Extraction of synonymy relation

The dictionaries of past tried to give synonymous items to the lexical item concerned. The earlier types of dictionaries in Telugu try to give synonymous term of the least understood lexical item. Even *puttu* 'to give birth' gives synonymous item to certain lexical items. This helps in establishing synonymous relations between lexical items.

a:me-ku bidda puttindi

that lady to child born-3rd-sing-female

she has got a child

4.2. Extraction of hyponymy-hypernymy relation

Dictionary definitions of nouns are normally written in such a way that one can identify for each headword, a "genus term", and these are related via an IS-A relation. Information following the genus term,

the differentia, serves to differentiate the headword from other headwords with the same genus. Let us look at the meaning of the katti found in *kattu*.

katti or ča:ku*'knife'*

paḍunina ančulugaligina ko:jada:niki upajoginche oka parikaramu

'a cutting instrument of flat meal plate with sharp edges'

In the definition *parikaramu* or *sa:danamu* 'instrument' is the genus term of the headword and *oka* čaḍunina: paḍunina ančulu*gala inupa lohamu* 'flat meal plate with sharp edges' and *ko:se vastuvu leka kattirinče vastuvu* 'that which can cut' yields differentia. The standard technique for creating a semantic network form noun definitions is to identify the IS-A relationships of this type. The standard technique for creating a semantic network form noun definitions is to identify the IS-A relationships of this type.

We can view the process of building a semantic hierarchy of IS-A links from nouns definitions as two folds: (1) to find the genus term (or terms) in the definitions and (2) to disambiguate them. Here we are concerned with the research projects in this area, the MRD's available for finding the hierarchies and networks for Telugu lexical items, and the degree of human intervention required.

The analysis of Telugu dictionary reveals us that about 90% of the noun definitions in the dictionary have the genus term, which is the head of the first noun phase (NP), and for these cases the standard method for identifying the genus term is to parse (or partially parse) the definitions and extract the head of the first NP. All researchers in this area have applied this method. However, all dictionaries seem to have a set of noun definitions that do not conform to this rule. Constructing taxonomies from the genus term thus forces one to take a stand on how to treat a rather large class of noun definitions which are not as 'standard' as the definition given above for *katti* 'knife'. The characteristic property of many of these definitions in that the head of the first NP (the usual place to find a genus term) seems vacuous, and another easily identifiable noun in the definition gives information about the headword. The analysis of definitions found in Telugu dictionary reveals us the following structure:

kattera 'scissors':

lohamuto tayaru cesina rendu poduva:ti padunaina baddalugala va:tini addamuga amarchi guddalanu mariyu dudila:ntiva:tini kattirincutaku upayogince parikaramu.

Word by word English gloss: scissors: two equal long metal strips_ACCacross_LOCone_GEN on one_ADV having joined their sharp inside_part_bycotton_cloth

things_of_that_sort_ACCcut_INFbe_useful_RP one instrument

'an instrument made up of two long metal strips joined cross-wise which is used to cut materials like cotton clothes'

The head occurs at the end of a clause for Telugu. Here *parikaramu* 'instrument' can be considered as the genus of the term *kattera* 'scissors'. These items which comepreceding the genus term can be considered as attributes. *upajogapadedi* 'that which is useful' identifies the telic features of the lexical item, i.e *kattirincutaku* 'for cutting'.

aratichettu 'plantain tree' *okarakamaina aratipallani ice čettu* 'a kind of 'plantain tree'

The key word *okarakamaina arati* 'a kind of' helps in finding the genus term *čettu* 'tree'.

Some genus terms are synonymous in nature. The following example will illustrate the point.

si:tala pa:nijamau 'cool drink':

dehanni čallaga unčenduku upjogince pa:nijamu 'the drink which made to become cold'

čallagaunčenduku ka:ranamajje pa:niyamu 'the drink which is caused to become cool'

In this definition the word *pa:niyaamu* 'drinks' which comes before the key word *okarakamaina* 'kind of' is the genus term.

Sometimes the genus term is far removed from the concerned word conceptually.

kaḷḷemu 'Bridle': *gurrmu jokka perugunu adupulouncutaku notiki taglinchi unche vastuvu*

'a thing which is a long strap joined to a horse's mouth and head in order to control its running movement'

In this kind of description, the genus term identified *vastuvu*'thing' is far removed from the immediate genus.

4.3. Extraction ofmeronymy-holonymy relation

Dictionaries while defining a word, always look for the possibility of defining it by means of meronymy-holonymy relation. If you look for the definition or meaning of the word *ka:lu* 'leg', you will definitely get the information that it is a part of the body. Part-and-whole relation (i.e. meronymy-holonymy relation) in the following case is difficult to identify, though the genus term can be easily identified.

velu 'finger': *velu sariramlo oka ba:gam* 'finger is part of the body'

The genus term *sarirabagam* 'body part' is explicit. But its meronymic relation is with *čejji* 'hand' or*ka:lu* 'leg' is difficult to establish though the genitive *jokka* 'of' can help us to infer that the holonym of *velu* 'finger' is čejji 'hand' or *ka:lu* 'leg'. *ba:gam* 'part' is the key term which can help us to establish meronymy-holonymy relation.

kannu 'eye'čudadaniki *upajoginche sariraba:gam*

'seeing_for be useful_RPbody_part'

'the body part useful for seeing'

In this definition *sariraba:gam* 'body part' is the genus of *kannu*. It more or less follows the previous pattern. The telic feature is given and other attributive or contentive feature is not given. *kannu* is the polysemous word and the identification of telic feature disambiguate it form other selecting other genus.

Many subtypes of mernymic relations such as member of, part of, piece of, constituent of, etc. can be established making use of dictionary definition.

4.4. Extraction of Made of relation

Sometimes the 'is made up of' relation can be established between lexical items by means of dictionary definition. The key word could be *čejabadina* 'that which is made of'

mruduvaina chekkato ṯajaručejabadina *oka rakamaina petta*

'a kind of box made up of deal wood'

The genus term is petta 'box'. *ṯajaručejabadina* is in 'is made up of 'relation with *mrudujaina čekkamukkalu* 'deal wooden planks'.

4.5. Extraction of different types of oppositions

Many different types of oppositions, both binary and multi-nary can be established by making use of dictionary definitions. Some of the dictionaries try to define certain words or concept by making use of antonymous relation. It is likely that *čallani 'cold'* could be defined by relating it with *vedi 'hot'*. You will number of instances of helices, circles, ranks, and proportional series in the dictionary definition itself.

4.6. Extraction of troponymy and entailment relations

You will find out instances of using of troponymy and entailment relations in the dictionary while defining certain lexical items. For example, *guraka* 'snore' will be defined as a part of *niṉra* 'sleep', or 'sore' entails 'sleep'. 'Succeed' many define as the outcome of 'try' or 'make effort'; 'limp' may be defined as a type of 'walk'. *pagiliunna* 'being broken' can be defined as the causative effect of *vira:mam* or *viruču*'break'.

5. Conclusion

MRDs are knowledge source for extracting semantic information about words. They can be very well exploited for building ontology for Telugu vocabulary. The identification of a satisfactory genus term is not straightforward in all cases, and the problem associated with classifying the relationships expressed are difficult and numerous. A through study of this shadowy area is necessary in order to make optimal use of the semantic information available in MRDs.

Reference and Notes:

Arulmozi, S. 2010. Telugu wordNet.Proceedings of the International Global WordNet Conference 2010, IIT, Bombay.

Busa, F., Calzolari, N. and Lenci, A. 1998. "Generative Lexicon and the SIMPLE Model: Developing Semantic Resources for NLP." In: Bouillon P. and Busa, F (ed.) Language of Word Meaning. Cambridge: Cambridge University Press. pp 333-349.

Busa, F., Calzolari, N.,Lenci, A. and Pustejovsky, J. 2001."Building a semantic lexicon: structuring and generative concepts." In: J. Bunt, H et al (eds.). 2001. Computing meaning, volume 2, pp 29-51.

Cruse, D.A. 1986.Lexical Semantics. Cambridge, England: Cambridge University Press.

Cruse, D.A. 2000. Meaning in Language: An Introduction to Semantics and Pragmatics. Oxford New York: Oxford University Press.

Cruse, D.A. 2002. "Hyponymy and its varieties", In: Green R, Bean, S.A., and Myaeng, S.H. (eds.). 2002. The Semantics of Relationships: An Interdisciplinary perspective, pp 3-22.

Hearst, M.A. 1999. "Automated Discovery of WordNet Relations." In Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database. Cambridge, Mass. MIT Press.

Lyons, J. 1968. Introduction to Theoretical Linguistics. London & New York: Cambridge University Press.

Lyons, J. 1977. Semantics Volume 1. Cambridge: Cambridge University Press.

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

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

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

Miller, G.A. 1998b. "Modifiers in WordNet." In: Fellbaum, C. (ed.). WordNet: An Electronic Lexical Database. The MIT press, Cambridge, Massachusetts, pp. 47-67.

Pribbenow, S. 2002. "Meronymic Relationship: From classical mereology to complex part-whole relations", In: Green R, Bean, S.A., and Myaeng, S.H. (eds.). 2002. The Semantics of Relationships: An Interdisciplinary perspective, pp 35-50.

Pustejovky, J. 1996. The Generative Lexicon.Cambridge, Mass. MIT Press.

Pustejovky, J. and Boguraev, B. 1993. "Lexical Knowledge Representation and Natural Language Processing." In F. Pereira and B. Grosz (eds.), Natural Language Processing. Cambridge, Mass.: MIT Press.

Rajendran, S. Creating Generative Lexicon from MRDs: Tamil Experience. Rajeev Sangal et al (edited). Recent Advances in Natural Language Processing: Proceedings of the International Conference of Natural Language Processing (ICON 2003), 2003, 83-91.

Rajendran, S. Visual Onto-thesaurus for Tamil, Language in India, May, 2017

Ruimy, N., Gola, E., and Monachini, M. "Lexicography Informs Lexical Semantics: The SIMPLE Experience." In: Bouillon P. and Busa, F (ed.) Languge of Word Meaning. Cambridge: Cambridge University Press pp 350-362.

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

Wierzbica, A., 1980. Lingua Mentalis: The Semantics of Natural Language. New York: Academic Press.

Wilks, Y.A., Fass. D.C., Guo, C.M. McDonald, J.E., Plate, T. and Slator, B.M. 1990."Providing machine tractable dictionary tools". In Pustejovksy, J. (ed.) Semantics and the Lexicon. Cambridge, Mass: MIT Press.

Wilks, Y.A., Slator, B.M. and Guthrie, L.M. 1996. Electric Words: Dictionaries, Computers, and Meanings. Cambridge, Mass: MIT Press.

Web references

http://www.dictionary.com/

https://en.wikipedia.org/wiki/Ontology

http://wordnetweb.princeton.edu/perl/webwn?s=ontology