LANGUAGE IN INDIA

Strength for Today and Bright Hope for Tomorrow

Volume 14:3 March 2014
ISSN 1930-2940

Malayalam WordNet

Dr. S. Rajendran & Dr. Soman, K.P

==========

Abstract

Work on Malayalam WordNet was initiated in Amrita Vishvavidya Peetam, Coimbatore in December, 2011 as a part of the project entitled "Development of Dravidian WordNet: An Integrated WordNet for Telugu, Tamil, Kannada and Malayalam", funded by Department of Information Technology, MCIT, Govt. of India. The main objective of the project is to build WordNets for Dravidian languages by making use of the already built Hindi WordNet under the project scheme IndowordNet.

Hindi WordNet has been built based on Princeton English WordNet which is a component of EuroWordNet. The main objective of EuroWordNet is to develop an extensive and high quality of multilingual database with WordNets for several languages (mainly European Languages such as French, German, Czech, Italian, etc.) in a cost-effective manner. On similar line, IndoWordNet is being built for Indian languages. Malayalam WordNet is a component of Dravidian WordNet which in turn is the component of IndoWordNet. Malayalam WordNet is an online lexical database. It is useful for many applications of Natural Language Processing.

1 Introduction

Malayalam WordNet aims to capture the network of lexical or semantic relations between lexical items or words in Malayalam. As we know, lexical items are related to one another in the hierarchical dimension as taxonomies (which show hyponymy-hypernymy and meronymy-holonymy relationship) and non-hierarchical dimension as opposites (which include complementaries, antonyms, antipodals, counterparts, reversives and converses) and synonyms (Lyons 1977, Cruse 1986). Also words are related to one another due to their derivational as well as collocational meaning. Componential analysis which studies meanings of lexical items in terms of meaning components or features can help us to capture the above mentioned network of relations in a more systematic way (Nida 1975a).

A database has to be created depicting the lexical items and their meaning relations such as hyponymy-hypernymy (subordination-superordination relationship), meronymy-holonymy (part-whole relationship), synonymy and lexical opposition and the formal relations such as derivation and collocation. The network of relations exist between the lexical items are captured in the Word Net. Such a study can be made use of for various lexical studies as well as application oriented studies like machine translation (in which word-disambiguation is a crucial issue), and machine oriented language learning and teaching.

2 Strategy of the WordNet

According to Miller et al (1993) "Word Net is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory." The organization or Word Net is based on the presumption that there is a mental dictionary or thesaurus in which the words are organized under conceptual fields or semantic domains. The Word Net aims at organizing lexical information in terms of word meanings or concepts rather than word forms. Word Net in this sense resembles a thesaurus more than a dictionary. A thesaurus in its widest contemporary sense is a classification of words by concepts, topics, or subjects. But the Word Net is much more efficient and versatile than the thesaurus whether it is in paper form or available in electronic form. In one sense WordNet is an on-line thesaurus. But its efficiency in bring out the lexical relations exalts it form thesaurus. The present WordNet of Malayalam is aimed to be built on the foundation offered by natural language processing (NLP) taking into account its application in the fields language teaching and language learning, lexicography, translation, both machine and human, and Artificial Intelligence (AI) knowledge representation. The ideas propounded by Miller (1991) and Miller et al (1990) are profusely used in the preparation of WordNet for Malayalam.

WordNet resembles a thesaurus in its broad framework. Its building blocks are synsets; each synset consists of all the words that express a given concept. This facilitates the user of a WordNet to recall the words expressing this concept knowing any one of the words which has lexicalized the same concept. But WordNet is not a list of concepts in the form of synsets. The relations such as hyponymy, meronymy, and entailment link the synsets to one another.

WordNet resembles a traditional dictionary in some respects. For instance, one can find definitions and sample sentences in WordNet for most of its synsets. Information about

morphologically related words also finds its place in WordNet. The goals of WordNet resemble those of a standard dictionary, and the semantics of WordNet is based on the notion of sense that the lexicographers have traditionally used in writing dictionaries. WordNet differs from the dictionaries in their organization. WordNet does not give pronunciation, derivation morphology, etymology, usage notes, or pictorial illustrations. However, WordNet depicts the semantic relation between word senses more transparently and elegantly.

WordNet relies on two commonly accepted relations: the conceptual-semantic relations which link concepts and the lexical relations which link individual words. The mental lexicon tends to build semantic networks with conceptual-semantic relations, whereas those who are focusing on lexical aspects use primarily lexical, word-word relations. Thus WordNet is organized by lexical and semantic relations. Since a semantic relation is a relation between meanings, and since meanings can be represented by synsets, it is natural to think of semantic relations as pointers between synsets. It is characteristic of semantic relations as pointers between synsets. Wordnet does not contain syntagmatic relations linking words from different syntactic categories. 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). The basic semantic relation in WordNet is synonymy. Sets of synonyms (synsets) form the basic building blocks. The notion of synonymy used in WordNet does not entail interchangeability in all contexts. A synset has only a single gloss.

3 Nouns in Wordnet

Nouns are organized in a lexical inheritance system. A typical definition of a noun contains a superordinate term followed by certain distinguishing features. The relation of subordination (or class inclusion or subsumption), which is called hyponymy organizes nouns into a lexical hierarchy. The superordinate relation generates a hierarchical semantic organization of nouns. Synset which contains a group of synonyms representing a concept is the building blocks of noun wordNet. Synonymy is a lexical relation that holds between word forms, whereas the semantic relation holds between lexicalized concepts.

3.1 Hyponymy and Hypernymy

Hyponymy is the relationship that exists between specific and general lexical items, such that the former is included in the latter. The relation that is reverse to hyponymy is hypernymy. The set of items that are hyponyms of same superordinate term or hypernym are co-hyponyms (or coordinates). The hyponymy-hypernymy relation is variously termed as subordination-superordination, subset-superset, etc. The relationship existing between *paSu* 'cow' and *mRIgaM* 'animal' and eruma 'buffalo' and mRIgaM 'animal' is hyponymy and *paSu* and *eruma* are co-hyponyms. *mRIgaM* is the hypernym of *paSu* and *eruma*. Hyponymy is unilateral and asymmetrical.

avaL~ talayil~ mullapuuv cuuTiirikkunnu
she head_LOC flower keep_PRES_she

'She is wearing jasmine on her head'
↓
avaL~ talayil~ puu cuuTiirikkunnu
she head_LOC flower keep_PRES_she
'She is wearing flower on her head'
avaL~ talayil~ puu cuuTiirikkunnu
she head_LOC flower keep_PRES_she
'She is wearing flower on her head'
↓
* avaL~ talayil~ mullapuuv cuuTiirikkunnu
she head_LOC jasmine keep_PRES_she
'She is wearing jasmine on her head'

Since the hyponymy relation is unilateral and symmetrical, the sentence with '*' marker is a wrong claim. Hyponymy shows transitive relation as shown below.

mRIgaM 'animal' ⊃ sastani 'mammal'
sastani 'mammal' ⊃ paSu 'cow'
mRIgaM 'animal' ⊃ sastani 'mammal' ⊃ paSu 'cow'

3.2 Lexical Hierarchy

The conventional dictionaries make use of the hyponymic relations between nouns to represent the meaning. The following illustration depicts it clearly.

tatta 'parrot'
ushNa meekhalayil~ kaaNappeTunna bhaMgiyuLLa paccaniRattooTu kuuTiyatuM kaTTiyuLLa taazhooTTu vaLanjnja cuNTukaL~ uLLatuM paruparutta uLLatumaaya oruyinaM pakshiyaaN.ii var~ggattil~ppeTunna cilayinaM tattakaL~kk manushyassvaraM anukarikkaanuLLa kazhivuNT.
'A mainly tropical bird, typically brightly coloured, with a strong downcurved hooked bill and raucous voice, some kind of which are able to mimic human speech. '
pakshi 'bird'- ushNa raktamuLLatuM muTTayiTunnatuM tuuvalukaL~ uLLatuM muN~ kaikaL~ ciRakukaLaayi ruupaantarappeTTiTTuLLatuM maaya oru jantu var~ggaM.(jiivi).
'bird – a warm-blooded egg-laying animal having feathers and forelimbs modified as wings.'
jantu 'animal'- svantamaayi SvasikkaanuM calikkaanuM kazhivuLLatuM, njaaneendriyangngaL~ uLLatuM selluloosillaatta kooSangngaL~ uLLatumaaya jiivi.
'animal – an organism capable of breathing and voluntary movement and possessing sense organs and cells with noncellulose walls'
jiivi 'organism'- jiivikkunnava
'organism - a living entity'

Each hypernymic relation can be represented by a corresponding hyponymic relation that points in the opposite direction. A lexical hierarchy emerges from this manner of representing

hyponymy and hypernymy relations. Knowledge representations by computer scientist make use of the hierarchies of these sorts widely. The nouns in WordNet depict a lexical inheritance system. Systematic effort has been made in the WordNet to connect hyponyms with their hypernyms (and vice versa).

WordNet presupposes a linguistic knowledge of anaphoric relations; an anaphor can be a hypernym of its antecedent. Each hypernym leads on to a more generic hypernym. Hypernym cannot be represented as a simple relation between word forms. Hypernymy is a relation between lexicalized concepts. It is represented in Wordnet by a pointer between the appropriate synsets. A lexical hierarchy can be reconstituted by a series of synsets which are related by means of hypernymy.

> {tatta, kiiraM} 'parrot' @→ {pakshi, paRava} 'bird' @→ {jantu} 'organism' @→ {jiivi, jiivanuLLava, jiivikkunnava, jantujaalangngaL~} 'living being'

The hyonymy-hypernymy relation can be read as 'IS-A' and 'IS-A-KIND-OF'; For example, *tatta oru pakshi* 'Parrot is a bird' or *tatta oru taraM pakshiyaaN* 'Parrot is a kind of bird'.

3.3 Unique Beginners

The hierarchical structuring of nouns can be assumed to be contained in a single hierarchy. Instead, WordNet divides the nouns into several hierarchies, each with a different unique beginner. The semantic fields or domains (Lehrar 1974) which contain their own stock of vocabulary can be equated with these multiple hierarchies. Unique beginner corresponds roughly to a primitive semantic component in a compositional theory of lexical semantics. There is a list of 25 unique beginners for noun source files of EuroWordNet (Vossen 1998).

{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}	

This way of representing lexical items depicts the ontological structures which captures the lexical inheritance of one item form the other as shown under the title 'lexical inheritance'.

3.4 Distinguishing Features

The hyponymy relation generates the overall structure of nouns hierarchies. But the details given by the features distinguish one concept form another. For example, *tatta mikkavaaRuM* pacca niRavuM *kaTTiyuLLa taazhooTT vaLanjnja cuNTukaLu*M *raNT* ciRakuaLuM *uLLa paruparutta SabddamuNTaakkunnatuM paRakkunnatuM aaya oru pakshiyaaN* 'parrot is a bird that is mostly green in colour and with two wings and raucous voice. It may be possible to associate parrot with at least three different kinds of distinguishing features:

1. Attributes: pacca niRaM 'green colour'
2. Parts: kaTTiyuLLa taazhooTT vaLanjnja cuNTukaL~ 'strong downcurved hooked bill' , raNT ciRakukaL 'two wings'
3. Functions: paruparutta SabddamuNTaakkuM 'raucous voice', paRakkuM 'flies'

3.5 Attributes and Modification

Values of attributes are expressed by adjectives. For example, size and color are attributes of parrot: the usual color of parrot can be expressed by the adjective *pacca* 'green'. The attributes associated with a noun are reflected in the adjectives that can normally modify it.

3.6 Function and Predication

It seems natural to say that the function of a pencil is to write or the function of knife is to cut, but to say that the function of a parrot 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.

katti 'knife – muRikkuka 'cut'
kuzhi 'hole' – kuzhikkuka 'dig'
citraM 'picture – varaykkuka 'draw'
peTTi 'box' – piTikkuka ' hold'

There are also linguistic reasons to assume that a thing's function is a feature of its meaning.

It should be remembered in this context that Pustejovsky who advocates for generative lexicon (Pustejovsky 1995:76, 2001:56) assumes that word meaning is structured on the basis of four generative factors, or qualia roles, that capture how humans understand objects and relations in the world and provide the minimal explanation for the linguistic behaviour of lexical items.

CONSTITUTIVE: the relation between an object and its constituent parts

FORMAL: the basic category that distinguishes the object within a larger domain

TELIC: the object's purpose and function

AGENTIVE: factors involved in the object's origin or "coming into being."

Pustovsky's qualia roles have to be remembered while making definitions for synsets in WordNet.

3.7 Meronymy and Holonymy

Meronymy is the part-whole relation between nouns which is generally considered to be a semantic relation. Meronymy is comparable to synonymy, antonymy, and hyponymy which are the different kinds of relations captured in the WordNet. Reverse of meronymy is holonymy. If X is a meronymy of Y, then Y is said to be a holonym of X (X=meronym; Y= holonym). Many concrete objects like bodies and artifacts are defined by meronymy-holonymy relation. Meronyms are distinguishing features that hyponyms can inherit. Consequently, meronymy and hyponymy are intertwined in complex ways. For example, if *cuNT* 'beak' and *ciRak* 'wing' are meronyms of *bird*, and if *parrot* is a hyponym of bird, then by inheritance, beak and wing must also be meronyms of *parrot*. The fact that parts are hyponyms as well as meronyms complicates the relations between meronymy and hyponymy. For example, {cuNT, kokk 'beak'} is not only a meronym of {pakshi 'bird'}, it is a hyponym of {*taaTi* 'jaw'}, which in turn is a meronym of {talayooT 'skull'} and a hyponym of {*asthi* 'skeletal_structure'}. 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 example, *piTi* 'handle' is a meronym of *katak* 'door' and 'door' is a meronym of *viiT* 'house', yet is sound odd to say *viiTin piTiyuNT* 'The house has a handle' or *viiTinte oru bhaagamaaN piTi*' 'The handle is a part of the house'.

3.8 Antonymy

Antonymy is lexical in nature and is founded on linguistic as well as psycholinguistic principles. Normally the word association test brings together antonymous words.

vijayaM 'victory' – tool~vi 'defeat'
santooshaM – 'happiness' - asantushTi 'unhappiness'

Though semantic opposition is not a fundamental organizing relation between nouns, it is captured in Wordnet for its own merits.

{[purushan~ 'man', strii 'woman',!], vyakti 'person',@...(purushavyakti 'a male person')}
{[, strii 'woman', purushan~ 'man',!], vyakti 'person',@...(striivyakti 'a female person')}

Antonymy is a lexical relation between words, rather than a semantic relation between concepts. The three kind of semantic relations – hyponymy, meronymy, and antonymy –depicts nouns as an interconnected network.

3.9 Binary Opposition

Antonymy, which is often considered opposite to synonymy, relies on the lexical relation, incompatibility. The table gives the typology of binary opposition (Lyons, 1977, vol. 1). There are many kinds of oppositions between words and antonymy is one among the opposite relations. The different types of opposition relations are listed below:

Type	**Example**
Antonymy	*valut* 'big one': *ceRut* 'small one'
Complementarity: Complementarity is a contrastive relation between two lexemes which exhaustively divide some conceptual domain into two mutually exclusive compartments, so that what does not fall into one of the compartments must necessarily fall into the other. There is no 'no-man's land', no neutral ground, no possibility of a third term lying between them.	*pakal~* 'day': *raatri* 'night', *satyaM* 'true': *asatyaM* 'false', *jayaM* 'pass': *tool~vi* 'fail'
Converseness: The pairs which express a relationship between two entities by specifying the direction of one relative to the other along some axis.	*bhar~ttaav* 'husband' vs. *bhaarya* 'wife',
Privative opposition: It is a contrastive relation between two lexemes, one of which denotes some positive property and other of which denotes the absence of that property.	*budhisaali* 'clever' vs. *maNTan* 'fool'
Equipollent opposition: An equipollent opposition is a relation in which each of the contrasting lexemes denotes a positive property.	*aaN* 'male person' vs. *peNN* 'female person'
Reciprocal Social roles	*bhishvaguran~* 'doctor': *roogi* 'patient' , teacher': student, *mutalaaLi* 'boss': *tozhilaaLi* 'servant'
Kinship opposition	acchan~ 'father': makan~ 'son' *amma* 'mother':makaL~ 'daughter'
Temporal opposition: Shows converseness temporally.	*mun~p* 'before': pin~p'after'
Spatial opposition: Shows converseness spatially.	*mukaLil~* 'above': *taazhe* 'below'
Orthogonal Opposition or Perpendicular Opposition	*vaTakk* 'north':*kizhakk* 'east' and *paTinjnjaaR* 'west' *kizhakk* 'east': *tekk* 'south' and *vaTakk* 'north'
Antipodal Opposition or Diametrical Opposition	*vaTakk* 'north': tekk 'south' kizhakk 'east: paTinjnjaaR 'west'

Most of the spatio-temporal postpositions and spatio-temporal adverbs are together considered as spatio-temporal nouns in the present wordNet.

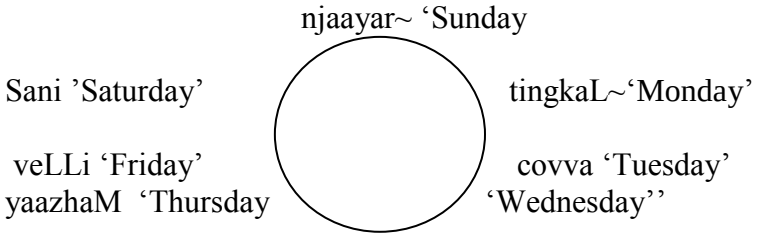
3.10 Multi-member Opposition

There are different types of multi-member sets in a language whose lexical relations can be described as incompatible denoting non-binary contrasts as opposed to binary contrasts. Various kinds of ordering are found in multi-member sets of incompatibles, and such sets may be serially or cyclically ordered (Cruse 1986). The constituents of a serial or cycle may be fixed or overlapping. The fixedly ordered items form a rank. The overlapping items may form a scale.

SERIAL

onn‘one’
↓
raNT ‘two’
↓
muunn ‘three’
↓
naal ‘four’
↓
anjc ‘five’

Cycle



SCALE

SreeshThaM 'excellent'
↓
nallat ‘good'
↓
tRIptikaraM 'satisfactory'
↓
mooSaM 'bad'
↓
vaLareemooSaM 'very bad’

RANK

keeNal~ ‘Colonel’

↓

lephRRanenRR keeNal~ ‘Lieutenant colonel’

↓

meejar~ ‘Major’

↓

kyaapRRan~ ‘ Captain’

↓

lephRRanenRR ‘Lieutenant’

3.11 Summing up of Relations in Nouns

The different types of lexical/semantic relations exhibited in Malayalam wordNet are listed in the following table:

Relations	**Subtypes**	**Example**
Synonymy	-	*pustakaM* to ‘book’ bukk ‘book
Hypernymy-Hyponymy	-	*mRIgaM* 'animal' to sastani 'mammal'
Hyponym-Hypernymy	-	*paSu* 'cow' to sastani 'mammal'
Holonymy-Meronymy	From wholes to parts	*meeSa* 'table' to kaal~ 'leg'
“	From groups to their members	*vakupp* 'department' to aacaaryan~ 'professor'
Meronymy-Holonymy	From parts to wholes	*cakraM* 'wheel' to cumaTuvaNTi 'cart'
“	From members to their groups	*seenaapati (paTattalavan~)* 'captain' to *seena* 'army'
Opposition	Antonymy (gradable opposites)	*nallavyakti (nallavan~)* 'good person' to vRIttikeTTavan~ 'bad person'
“	Complementarity (a item complement another item)	*aaN~* 'male' to peNN 'female'
“	Privative opposition (presence of a feature implies the absence of another)	budhisaali ‘clever’ vs. mantas ‘fool’
	Converse opposition: Spatial opposites showing converseness	miite *(*mukaLil~*)* 'above' vs. taazhe 'below'
“	Converse opposition: Temporal opposites showing converseness.	*mun~p* 'before' to pin~p 'after'
“	Equipollent opposition (both the items have positive features)	*aaN~* 'male' to peNN 'female'

“	Reciprocal Social roles	bhishvaguran~'doctor' to roogi 'patient'
“	Kinship opposition	*acchan*~ 'father' to makan~ 'son'; amma 'mother' to makaL~ 'daughter'
“	Orthogonal Opposition or Perpendicular Opposition	vaTakk 'north' to kizhakk 'east' and paTinjnjaaR'west' kizhakk 'east' to tekk'south' and vaTakk 'north'
“	Antipodal Opposition or Diagonally opposite relation	vaTakk ‘north’ to tekk ‘south’
Multiple opposites	Serial	*onn* ‘one’, raNT, muunn ‘three’, naal ‘four’, and so on.
“	Cycle	*njaayar*~ ‘Sunday’ to *tingkaL*~ ‘Monday ’to *covva* ‘Tuesday’ to *budhan*~ ‘Wednesday’ to *vyaazhaM* ‘Thursday’ to *veLLi* ‘Friday’ Sani ‘Saturday’

4. Adjectives in WordNet

Adjective is the syntactic category which is associated with noun modification. The sole function of adjectives is modification of nouns. But noun, verb, and prepositional phrases do not have the sole responsibility of modifying. Adjectives are organized in WordNet is a unique way which differs from the organization of the other major syntactic categories, noun and verb. WordNet contains:

Descriptive adjectives (Ex. valiya ‘big’, *taatparyamuLLa* ‘interesting’, *saadhyamuLLa* ‘possible’)
Relational adjectives (Ex. adhyakshata presidential, kaar~shika ‘agriculatural’, saMgiita ‘musical’, dvanta ‘dental’)
Reference modifying adjectives (Ex. mun~patte former, aaroopitamaaya alleged)

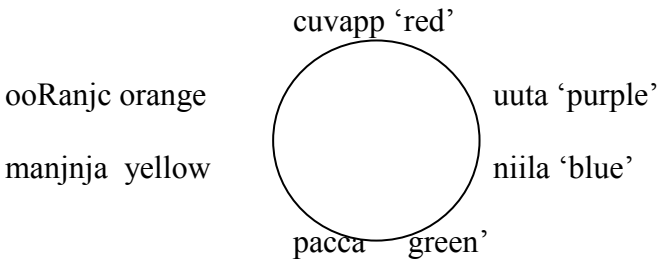
“A descriptive adjective is one that ascribes a value of an attribute to a noun. That is to say, x is Adj presupposes that there is an attribute A such that A(x) = Adj.” (Gross and Miller 1990)

at nalla bhaaramuLLa cumaTaaN ‘that luggage is heavy’

The above sentence presupposes that there is attribute *bhaaraM* ‘WEIGHT’ such that bhaaraM ‘WEIGHT’ *(cumaT* ‘luggage’) = *bhaaraM* ‘heavy’. In the same way *taazhnna* ‘low’ and *uyar~nna* ‘high’ are values of uyaraM ‘HEIGHT’. The wordNet 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. As we propose to keep the referential meanings representing abstract nouns, adjectives and adverbs under the semantic domain 'abstracts', the adjectives will naturally fall under their related abstract nouns. For example, the adjectives viitiyuLLa 'wide' and *iTungngniya* 'narrow' are kept under the semantic domain 'dimension' in which the attribute viiti 'width' is kept. Relating descriptive adjectives with the particular noun they pertain to is known by the term pertainymy.

Colour adjectives: Colour adjectives are the intensely studied ones and are organized differently in WordNet. Colour adjectives can function as nouns as well as adjectives in Malayalam. They can be graded, nominalized and conjoined with descriptive adjectives. But the pattern of direct and indirect antonymy that is observed for other descriptive adjectives does not hold good for colour adjectives. Formation of colour terms is a productive process in Malayalam. New colour terms are coined by the Malayalam speakers to denote different shades of colours (e.g. *elappacca* 'leaf-like green', *tattappacca* 'parrot like green', *rektaccuvappu* 'blood red'). The colours of a spectrum can be visualized as a cycle as shown below.



Colours can be graded between *veLLa* 'white' and *kaRuppu* 'black'. They can be graded by the attribute *iLaM* 'light' (e.g. *iLaM pacca* 'light green', *iLaM cuvapp* 'light red) and *kaTuM* '*dark'* (*kaTum pacca* 'dark green', *kaTum cuvappu* 'dark red')

Quantifiers: Quantifiers are listed under determiners by some linguist; According to Lyons (Lyons 1977, vol.2, 455) quantifier tells us how many entities or how much substance is being referred to. WordNet, therefore, distinguishes quantifiers from determiners. In Malayalam the words such as the following can be classified under quantifiers: ellaa/sakala 'all', aneekaM/pala 'many', atikaM/orupaaT 'much'. kuRacc/cila/mikka 'some', parimitaM/curukkaM/kuRe 'few', raNTu 'both'. In many respects quantifiers resemble descriptive adjectives. Like adjectives, many quantifiers are gradable (e.g. vaLare kuRacc aaTkaL 'very few persons', vaLare atikaM aaTkaL 'very many persons')

Participial adjectives: The adjectival participle forms of verbs function as adjectives in Malayalam. For example the adjectival participle forms such as *viramicca* 'retired' (*viramicca paNTitan*~'retired professor') , *tiLappicca* 'boiled' (*tiLappicca veLLaM* 'boiled water'), *vivaakaM kazhiyaatta* (*vivaakaM kazhiyaatta peNN* 'unmarried woman'), *vivaahaM kazhinjnja* (*vivaakaM kazhinjnja peNN* 'married woman'), *veer~prinjnja veer~prinjnja dampati* 'divorced couple'), etc. can function as adjectives. The adjectivalizers (or adjectival suffixes) such as aaya and uLLa are basically the adjectival participle forms of the verbs *aakuka* 'become' and *uL* 'be'

respectively. They form adjectives when suffixed to nouns (e.g. *miTukkan'* clever male person' + *aaya* > *miTukkanaaya* 'clever' *sundari* 'beautiful female person' + *aaya* > *sundariyaaya* 'beautiful', *azhaku* 'beauty' + *uLLa* > *azhakuLLa* 'beautiful', *kaTTi* 'thickness' + *uLLa* kaTTiyuLLa 'thick').

4.1 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. 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 bhaaramuLLa 'heavy' is *bhaaraM kuRanjnja* 'light' that expresses a value at the opposite pole of the *bhaaraM* WEIGHT attribute. This binary opposition is to be represented in Malayalam WordNet.

4.2 Gradation and Non-gradation in Adjectives

Distinction is drawn between gradable and non-gradable adjectives. The first is referred as antonyms and the second one as complementaries by Lyons. The essence of a pair of complementaries is that between them they exhaustively divide some conceptual domain into two exclusive compartments, so that what does not fall into one of the compartments must necessarily fall into the other. There is no 'no-man's-land', no neutral ground, no possibility of a third term lying between them. It has been claimed that complementary adjectives are not normally gradable, that is to say, they are odd in the comparative or superlative degree or when modified by intensifiers such as *tiivramaaya* 'extremely', *mitamaaya* 'moderately' and *neeriya* 'slightly'. Antonymy is expressed by pairs such as niiNTa 'long'/kuRukiya'short', *veegattil~* 'fast'/*patukke* 'slow', *laLitamaaya* 'easy'/*kaThinamaaya* 'difficult', *nalla* 'good'/*mooSamaaya* 'bad', *cuuTuLLa* 'hot'/*kuLiruLLa* 'cold'. They are fully gradable. The members of a pair denote degree of some variable property such as length, speed, weight, accuracy, etc. The terms of a pair do not strictly bisect a domain: there is a range of values of the variable property, lying between those covered by the opposed terms, which cannot be properly referred to by either term. The complementaries and antonyms of Lyon are otherwise called as contradictory and contrary terms respectively. 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. For example, jiivanuLLa 'living' and jiivanillaatta 'non-living' are contradictory terms as at *jiivanuLLa jantu* 'it is a living creature' necessarily implies at jiivanillaatta jantu 'it is not a non-living creature'. But taTicca 'fat' and melinjnja 'thin' are contrary terms because *maala taTicca peNN kuTTiyaaN* 'Mala is a fat girl' and *maala melinjnja peNN kuTTiyaaN* 'Mala a thin girl' cannot both be true, although both can be false if maalaa 'Mala' is of average weight. Contraries are gradable adjectives, whereas contradictions are not. Gradation must also be considered as a semantic relation to organize adjectives. The following data will exemplify the gradation found among adjectives:

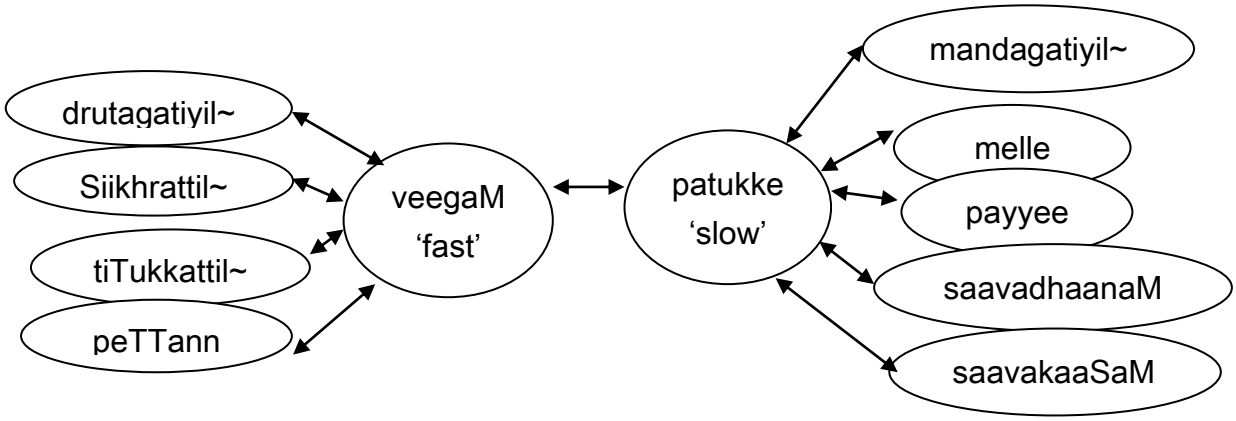
nalla cuuTuLLa 'very hot'
cuuTuLLa ' hot'

iLaM cuuTuLLa ‘warm’
taNutta ‘cold’

Word Net has to account for the gradation found among adjectives.

4.3 Similarity in Adjective

The adjectives lacking antonyms are similar in meaning to adjectives that do have antonyms. Adjectives are organized in clusters of synsets associated by semantic similarity to a focal adjective that relates the cluster to a contrasting cluster at the opposite pole of the attribute. In WordNet direct antonyms are represented by antonymy pointer !; indirect antonyms are inherited through similarity, which is indicated by the similarity pointer & meaning ‘IS SIMILAR TO’. The following figure illustrates this.



In the figure *veegam* ‘fast’ and *patukke* ‘slow’ are direct antonyms; others are related to the directly antonymous words by ‘similarity’. Overwhelmingly, association data and co-occurrence data indicate that valiya ‘big’ and ceRiya ‘little’ are considered as a pair and uyar~nna ‘high’ and taazhnna ‘low’ are considered as a pair. These pairs demonstrate that antonymy is a semantic relation between words rather than concepts. Polysemy is found among adjectives as a limited number of adjectives are used to attribute a considerable number of nouns.

4.4 Markedness in Adjectives

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 uyar~nna ‘high’/ taazhnna ‘low, praayamuLLa ‘old’/yavvanamaaya ‘young’, niiLamuLLa ‘long’/kuRukiya ‘short’, visthaaramuLLa ‘wide/iTungngiya ‘narrow’. We measure things by uyaraM ‘height’ rather than uyaraM kuRanjnja ‘shortness’. While asking questions about ‘height’, we say *at yatra uyaramuLLa tuuNaaN* ‘How high that pillar is?’rather than *at yatra uyaramkuRanjnja tuuNaaN* ‘How short that pillar is?’. A question ‘How short is X?’ is felt to contain the assumption that X is short, while no equivalent assumption is present in ‘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, *uyaramuLLa* 'high' is the unmarked term; the secondary member, *uyaraM kuRanjnja* 'short' is the marked one. They are related to the attribute noun *uyaraM* 'height'. WordNet has to capture the relation between marked and unmarked terms and their cross reference to their variable property.

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

uyaraM : taazhcca 'high': 'low', vayassaaya: ceRuppamaaya 'old': 'young', niiLamuLLa: kuRukiya 'long: short', visthaaramuLLa: iTungngiya 'wide': 'narrow'

4.5 Polysemy and Selectional Preferences

Polysemy is found among adjectives 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 samayaM 'good time'
nalla naaNayaM 'good coin'
nalla suhRItt 'good friend'
nalla cerupp 'good chappal'

Adjectives are choosy about the nouns they modify. The general rule is that if the referent denoted a noun does not have the attribute whose value is expressed by the adjective, then the adjective-noun combination requires a figurative or idiomatic interpretation (Gross and Miller 1990). For example, a 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. The selectional preferences of adjectives should be captured in the WordNet by organizing the adjectives under abstracts.

5 Adverbs in WordNet

Adverbs may be morphologically complex or simple in Malayalam. Among those that are morphologically complex, some can readily be considered the result of a derivational process that is still productive, and some the result of a process that is no longer so. The most productive process by which adverbs are derived from nouns (and occasionally from adjectives) by the suffixation of *–aayi* (e.g. bangngiyaai 'beautifully', gambhiiramaayi 'grand manner', taazhmayaayi 'humbly', viSadamaayi 'in detail', aadyamaayi 'for the first time').

Semantically adverbs can be classified into three broad categories: spatial adverbs (aviTe 'here', eviTe 'where'), temporal adverbs (ippozh~ 'now', eppoozh~ 'when', innu 'today', innalee 'yesterday') and manner adverbs (patukke 'slowly', uRakke'loudly', veegam 'quickly').

In Malayalam WordNet, adverbs derived from nouns are linked to their nominal senses by means of a pointer meaning 'DERIVED FROM'. The semantic organization of adverbs is simple and straightforward. There is no tree structure, as for nouns and verbs; nor is there a cluster structure as for adjectives. Synonymy and sometimes antonymy are recognized. All adverbs are listed individually in a single adverb file in WordNet.

6 Verbs in WordNet

Verbs are arguably most important lexical and syntactic category of a language. The verb provides the relational and semantic framework for its sentence. Its predicate-argument structure (or subcategorization frame) specifies the possible syntactic structures of the sentence in which it can occur.

6.1 Organization of Verbs in WordNet

In EuroWordNet verbs are grouped under 15 semantic domains (Vossen 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.)

6.2 Unique Beginners

The verbs are divided into certain number of semantic domains so as to organize them in the verb lexicon. It is difficult to think of a single root verb or "unique beginner" that could head the entire verb lexicon. The above mentioned semantic domains act as unique beginners for verbs in the present WordNet.

6.3 Polysemy in Verbs

The verbs are fewer in number than nouns in Malaylam 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. The polysemous nature of the verbs suggests that verb meanings are more flexible than noun meanings. Verbs change their meaning depending on the kinds of noun

arguments with which they co-occur, whereas the meanings of nouns tend to be more stable in the presence of different verbs.

6.4 Componential Features of Verbs

Verbs can be paraphrased in terms of finer semantic features. The decompositional nature of verbs has been exploited for the interpretation of verbs denoting complex events in terms of verbs denoting simple events. For example the verb *kolluka* 'kill' can be decomposed into 'cause not to become alive'. The verb *eRiyuka* '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. *ooTuka* 'run' vs. *ooTikkuka* 'cause to run') is understood by the 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.

6.5 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 *kazhikkuka* 'eat' and tinnuka "eat'. avan~ cooR kazhikkunnu 'He is eating the meals' can entail avan~ cooRu tinnunnu 'He is eating the meals. The relation existing between kazhikkuka and tinnuka is synonymy and kazhikkuka and tinnuka are synonyms, at least in this context. Truly synonymous verbs are difficult to find and mostly quasi synonymous verbs are found in Malayalam. The existence of a simple and a parallel compound form (noun + verbalizer) causes synonymy (quasi synonymy) in verbal system of Malayalam.

kolluka 'kill' and kolaceyyuka 'murder'

anVeeshikkuka 'enquire' and vicaaraNaceyyuka 'investigate'

The synonymous expressions of many verbs show that they are manner elaborations of more basic verbs. For example, vitaraNaM cceyyuka 'distribute' can be considered as an elaboration of the basic verb koTukkuka '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.

6.6 Lexical Entailment and Meronymy

Lexical entailment refers to the relation that holds between two verbs *V1* and *V2* when the statement *Someone V1* entails *Someone V2* (Miller, 1991:233). For example, kuur~kkaM valikkuka 'snore' lexically entails uRangnguka 'sleep', because the sentence *avan~ kuur~kkaM*

valikkunnu ‘he is snoring’ implies avan uRangngunnu ‘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, uRangnguka need not entail svapnaM kaaNuka.

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 paacakaM ceyyuka 'cook' is a complex activity involving a number of sub-activities. Consider the relation between the verbs *vaangnguka* 'buy' and koTukkuka '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 kuur~kkaM valikkuka 'snore', svapnaM kaaNuka 'dream', and uRangnguka '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 for snoring or dreaming is a proper part of the time you spend for sleeping. And it is true that when you stop sleeping you also necessarily stop snoring or dreaming. The relation between pairs like vaangnguka 'buy' and koTukkuka pay' and kuur~kkaM valikkuka 'snore' and uRangnguka 'sleep' are due to the temporal relations between the members of each pair. The activities can be simultaneous (as in the case of vaangnguka 'buy' and koTukkuka'pay' or one can include the other (as in the case of kuur~kkaM valikkuka'snore' and uRangnguka 'sleep').

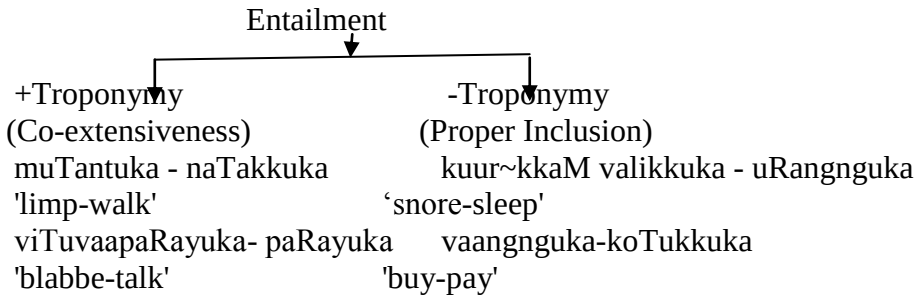
6.7 Hyponymy among Verbs

Some verbs seem more generic than others. For example, koTukkuka 'give' describes a wider range of activities than vitaraNaM ceyyuka 'distibute'. 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. The analysis of verbs of motion in Malayalam reveals the fact that the semantic component such as +DIRECTION (eg. kayaRuka 'climb up' vs iRangnguka 'climb down'), +MANNER (e.g. vazhuti viizhuka 'slip down' vs viizhuka 'fall') + CAUSE (eg. ooTuka ‘run’ vs. ooTikkuka ‘cause to run’, +SPEED (e.g. izhayuka 'crawl' vs ooTuka '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, muTantuka‘to walk unevenly’ is a troponym of naTakkuka ‘walk’ as the former entails the latter.

6.8 Troponymy and Entailment

Troponymy is a particular kind of entailment in that every troponym *V1* of a more general verb V2 also entails V2 (Miller, 1991). Consider for example the pair muTantuka 'limp' and naTakkuka 'walk'. The verbs in this pair are related by troponymy: muTantuka is also naTakkuka in a certain manner. So muTantuka is a troponym of naTakkuka. The verbs are also in entailment relation: the statement avan~ muTantunnu 'he is limping' entails avan~naTakkunnu 'he is walking'.

In contrast with pairs like muTantuka 'limp' *and* naTakkuka 'walk', a verb like kuur~kkaM valikkuka 'snore' entails and is included in uRangnguka 'sleep', but is not a troponym of uRangnguka. Similarly vaangnguka 'buy' entails koTukkuka 'give', but is not a troponym of koTukkuka 'give'. The verbs in the pairs like kuur~kkaM valikkuka snore' and uRangnguka '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. The following tree diagram adopted from Fellbaum (1990) depicts the two categories of lexical entailment that have been identified so far:



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 unique beginner.

6.9 Opposition Relations and Entailment

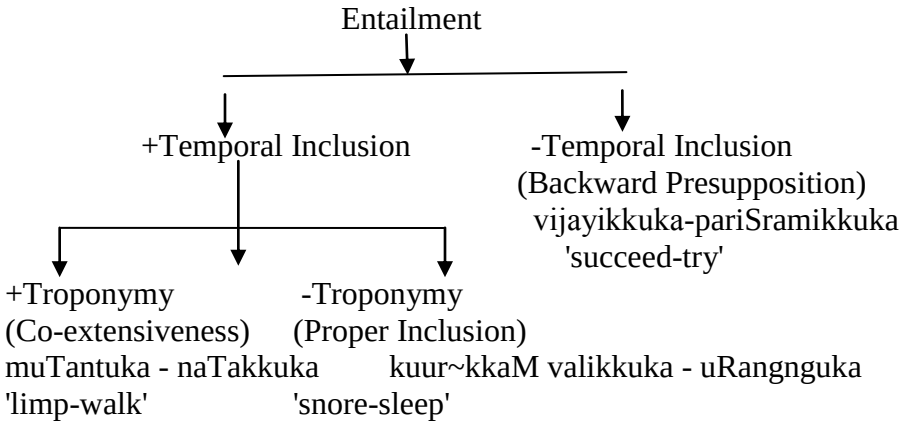
Opposition relations are psychologically significant not only for adjectives, but also for verbs. It is found that after synonymy and troponymy, opposition relations are the most frequently coded semantic relations in building database for verbs. The semantics of opposition relations among verbs is complex. As for as Malayalam 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~ avaLe konnu** is in converse relation

with the passive expression avaL~avanaal~ kollappeTTu. Thus active-passive pairs of transitive verbs in Malayalam show converse opposition. The relation between the verbs *vaangnguka* 'buy' and *vil~kkuka* '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 *eththicceeruka* 'arrive' and *puRappeTuka* 'reach', *varuka* 'come':and *pookuka* 'go' is directional opposition. Under this category are the verb pairs such as *uyaruka* 'rise' and *taazhuka* 'go down', *kayaRuka* 'ascend' and *iRangnguka* 'descend'. There are other oppositions with reference to change of state, manner, speed, etc. as exemplified below:

nir~mmikkuka 'build' : poLikkuka 'demolish'
keTTuka 'tie' : azhikkuka 'untie'
sammatikkuka 'agree' : viyoojikkuka 'disagree'
valikkuka 'inhale' : viTuka 'exhale'
naTakkuka 'walk' : ooTuka 'run'

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.

The componential analysis of verbs shows that many verb pairs in an opposition relation also share an entailed verb. For example the pair vijayikkuka 'succeed' and paraajayappeTuka 'fail' entails pariSramikkuka 'try'. "A verb V1 that is entailed by another verb V2 via backward presupposition cannot be said to be part of V2. Part-whole statements between verbs are possible only when a temporal inclusion relation holds between these verbs" (Fellbaum, 1990). On the basis of temporal inclusion, the set of verbs related by entailment can be classified exhaustively into two mutually exclusive categories as shown in the following tree diagram adopted from Fellbaum (1990):



(Three kinds of entailment)

6.10 Causation and Entailment

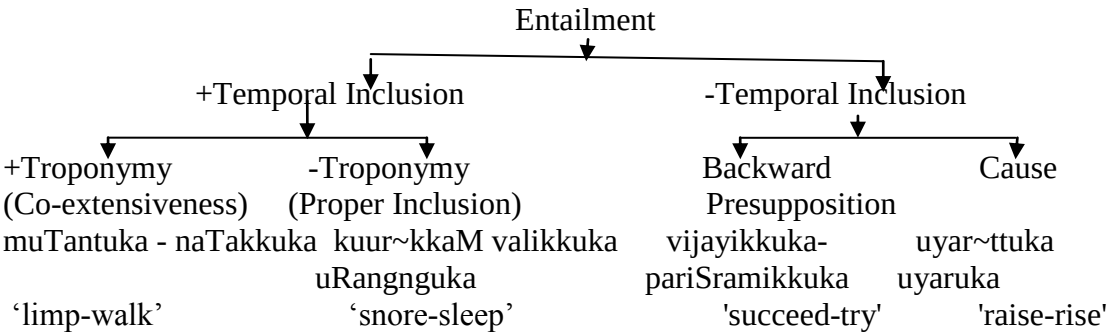
The causative relation exists between two verbal concepts: one is causative (e.g. koTukkuka 'give') and the other is resultative (e.g. kiTTuka 'get'). Causation can be considered

as a specific kind of entailment: if *V1* necessarily causes *V2*, then *V1* also entails *V2* (Fellbaum 1990).

puRattaakkuka 'expel' entails viTTupookuka 'leave'

uyar~ttuka ‘raise’ and uyaruka ’rise’ (temporal inclusion)

We have distinguished four different kinds of lexical entailment that systematically interact with the semantic relations mapped in WordNet. These four kinds of entailment can be related as shown in the following tree (Fellbaum (1990):



6.11 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 the WordNet has to be explored for the better understanding of verb net.

6.12 Summing up of VerbNet

The following table sums up the lexical/semantic relations to be captured in the VerbNet.

Relations	**Definition/sub types**	**Example**
Synonymy	Replaceable events	mayangnguka ‘sleep’ → uRangnguka ‘sleep’
Meronymy- Hypernymy	From events to superordinate events	*paRakkuka* 'fly' → yaatraceyyuka 'travel'
Troponymy	From events to their subtypes	*naTakkuka* ‘walk’→ muTanthuka 'limp'
Entailment	From events to the events they entail	*kuur~kkaM valikkuka* 'snore' *uRangnguka* 'sleep'
“	From event to its cause	uyar~ttuka ‘raise’ → uyaruka‘rise’
“	From event to its presupposed event	jayikkuka ‘succeed’ → pariSramikkuka ‘try’
“	From even to implied event	kolluka ‘murder’ →

		marikkuka ‘die’
Antonymy	Opposites	kuuTuka 'increase' → kuRayuka 'decrease'
“	Conversensess	vil~kkuka ‘sell’ → vaangnguka ‘buy’
“	Directional opposites	puRappeTuka ‘start’ → etticceeruka ‘reach’

7 Designing and Implementing WordNet

The word net automatically inherits the all the powers of a thesaurus. It also resembles an on-line dictionary as it provides meanings for lexical items. Being superior to these two tools, word net provides much more information that has been loaded in an on-line thesaurus as well as in an on-line dictionary. The task of developing the on-line database can be conveniently divided into two interdependent tasks (Beckwith and Miller 1990). These tasks bear a vague similarity to the traditional tasks of writing and printing a dictionary:

- To write the source files that contain the basic lexical data - the contents of those files are the lexical substance of WordNet.
- To create a set of computer programs that would accept the source files and do all the work leading ultimately to the generation of a display for the user.

The WordNet system is divided into four parts based on the specific tasks assigned to them:

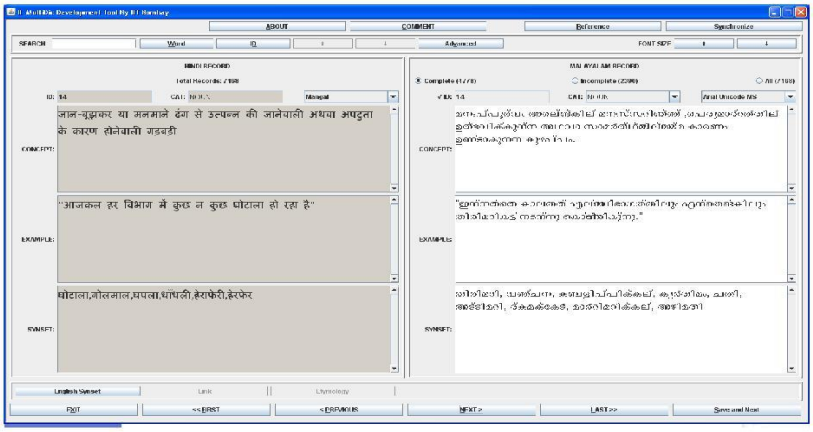
- Lexical resource system
- Compiler system
- Storage system
- Retrieval system

Lexical resource system contains source files. Lexicographers write WordNet’s source files. They are the products of a detailed relational analysis of lexical semantics: a variety of lexical and semantic relations are used to represent the organization of lexical knowledge. The Compiler System primarily compiles the lexical resource files into a database format and sends it to storage system to facilitate machine retrieval of the information in the WordNet. Storage system works as an intermediary between Compiler System and Retrieval System. The cooked database of the compiler system is stored in the Storage System for retrieval.

The primary focus in WordNet construction is to ensure the degree of precision that is called for in a given information search and retrieval system and to eliminate any redundancy in the codification of the hierarchies. An interface is required in order to give a user access to information in the database. Interfaces enable end users to retrieve the lexical data and display it via window-based tool or the command line. The interface provides a user with a variety of ways to retrieve and display lexical information. Different interfaces can be created to serve the purpose of different users, but all of them will draw on the same underlying database, and may use the same software functions that interface to the database files.

8. Malayalam WordNet Development Using Expansion Approach

Malayalam wordNet is being built using expansion approach (Vossen 1998). In this approach synsets are created by referring to the existing WordNet of the chosen language. Hindi is used as a source language to create synsets of Malayalam. A synset linkage tool provided by Indian Institute of Technology, Bombay, is used to create synsets in Malayalam. This synset linking tool provides graphical user interface which shows Hindi synset on the left side and provides interface to enter Malayalam synset on the right hand side.



Lexicalization of concepts varies across languages, which leads to synsets that are present in one language but not in another language. In the present project the lexical items are divided into six categories: (1) Universal, (2) Pan-Indian, (3) In-family, (4) Language specific, (5) Rare and (6) Synthesized. The lexical items covered consist of noun, verb, adjectives and adverbs. The main objective of the project is word sense disambiguation. For that sense marking will be done at the next stage of the project. The sense making will be done on the corpus by using sense IDs as tags. This will enable word sense disambiguation in the text.

9 Conclusion

The theme of lexical semantics, computational lexicography, and computational semantics are expanding rapidly. The availability of machine-readable resources and newly developed tools for analyzing and manipulating lexical entries makes it possible to build a massive WordNet for a language. In present state of affairs, it is quite feasible to build an efficient WordNet for Malayalam. Linkage of Hindi WordNet with Malayalam and other chosen Indian languages creates a multilingual resource for Indian languages which is useful for many NLP applications. However, variation in the lexicalization of the concepts across languages poses a major challenge in WordNet linking.

==========

References

Aitchison, J. 1987. Words in the Mind: An Introduction to the Mental Lexicon. Oxford: Basil Blackwell.

Atkins, B.T.S. and A. Zampolli. 1994. Computational Approaches to the Lexicon. Oxford: Oxford University Press.

Bhattacharyya, P. 2010. IndoWordNet. Proceedings of the 7th International Language Resources and Evaluation Conference (LREC-10), Valletta, Malta, May 19-21, 2010.

Beckwith, R. and Miller, G.A. 1990. 'Implementing a lexical network.' In: International Journal of Lexicography 3 (4):302 - 312.

Bierwisch, M. 1967. 'Some semantic universals of German adjectives.' Foundations of Language 3:1-36.

Bierwisch, M. 1989. 'The semantics of Gradation'. In Bierwsch, M. and Lang, E. (eds). Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation. Berlin: Springer-Verlag.

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

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

Calzolari, N. 1988. 'The Dictionary and the Thesaurus can be combined. In Evens, M. (ed.). Relational Models of the Lexicon: Representing Knowledge in Semantic Networks. Cambridge: Cambridge University Press.

Carter, R. 1976. 'Some constraints on possible words.' Semantikos 1:27-66.

Chafe, W. 1970. Meaning and Structure of Language. Chicago: University of Chicago Press.

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

Dixon, R.M.W. 1982. Where have all the Adjective gone? Berlin: Mouton Publishers.
Fellbaum, C. 1990. 'English verbs as a semantic net.' In: International Journal of Lexicography 3 (4):278 - 301.

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

Graside, Roger, Geoffrey Leech & Geoffrey Samson (eds.) 1987. The Computational Analysis of English: A Corpus based Approach. London: Longman

Gross, D. and Miller, K.J. 1990.'Adjectives in WordNet.' In: International Journal of Lexicography 3 (4):265 - 277.
Gruber, J. 1976. Lexical Structures in Syntax and Semantics. New York: North Holland.

Guckler, G. 1983. Appendix: B: 'A Computer-based Monolingual Dictionary: A Case Study.' In R.R.K. Hartmann (ed.) 1983. Lexicography: Principles and Practice. London: Academic Press Inc.

Hudson, R. 1995. Word Meaning. London and New York: Routledge.
Jackendoff, R. 1972. Semantic Interpretation in Generative Grammar. Cambridge, Mass.: MIT Press.

Jackson, H. 1988. Word and their meaning. London: Longman.
Jones, K.S. 1986. Synonymy and Semantic Classification. Edinburgh: Edinburgh University Press.

Justeson, J.S., and Katz, S..M. 1991. Co-occurrences of Antonymous Adjectives and their Contexts. Computational Linguistics 17.1-19.

Katz, J.J. 1972. Semantic Theory. New York: Harper and Row.

Katz. J.J. and Fodor, J. 1963. 'The Structure of Semantic Theory.' Language 39:170-210.

Leech, G.N. 1974. Semantics. Harmondsworth: Penguin.

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

Levi, J.N. Syntax and semantics of complex nominals. New York: Academic Press.

Levin, J.N. 1989. 'Towards a Lexical Organization of English verb'. Ms., Evanston: Northwestern University

Lyons,J.1977. Semantics, 2 volumes. New York: Cambridge University Press.

Lyons, J. 1995. Linguistic Semantics: An Introduction. Cambridge: Cambridge University Press.

Martin, W.J.R., B.P.F.Al and P.J.G.Van Sterkenburg. 1983. 'On Processing of A text Corpus.' In

R.R.K. Hartmann (ed.). 1983. Lexicography: Principles and Practice. London: Academic Press Inc.

McCawley, J.D. 1968. 'Lexical Insertion in a Transormational Grammar without Deep Structure.' Darden, B.J., Bailey C-J.N, Davison (eds.)1968. Papers from the Fourth Regional Meeting. Chicago Ill.: Department of Linguistics, University of Chicago, 71-80.

Miller, G.A.1990. 'Nouns in WordNet: a lexical inheritance system.' In: International Journal of Lexicography 3 (4): 245 - 264.

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

Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K.J.1990. 'Introduction to WordNet: an on-line lexical database.' In: International Journal of Lexicography 3 (4):235 - 244.

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

-----1975.b. Exploring Semantic Structure. The Hague: Mouton

Pandey. M.K. 1995. 'An Electronic Thesaurus: Theoretical Premise'. In Francis Ekka et al (ed.) 1995. Indian Congress of Knowledge and Language vol. I. Mysore: CIIL.

Pustejovksy, J. 1995. The Generative Lexicon. Cambridge, Massachusetts: The MIT Press.

Rajendran S. 2002. Preliminaries to the preparation of a Word Net for Tamil. Language in India 2:1, www.languageinindia.com

----, 2003. Dravidian WordNet: A Proposal. R.M.Sundaram et al (edited), Facets of Language, Thanjavur, 2003, 467-497.

----, 2009. Dravidian WordNet. In: Proceedings of Tamil Internet Conference 2009. Cologne, Germany, October, 2009

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

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

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

Rajendran, S, S Arulmozi, B Kumara Shanmugam, S Baskaran and S Thiagarajan. 2002. Tamil WordNet. In: Proceedings of the First Global WordNet Conference, CIIL, Mysore, 22-25, 2001.

Vossen, P. 1998. EuroWordNet: a multilingual database with lexical semantic networks. London: Kluwer Academic Publishers.

Dr. S. Rajendran
Professor of Linguistics
Department of Computational Engineering and Networking
Amrita Vishwa Vidyapeetham
Ettimadai
Coimbatore 641105
Tamilnadu
India
raj_ushush@yahoo.com

Dr. Soman, K.P.
Professor and Head
Department of Computational Engineering and Networking
Amrita Vishwa Vidyapeetham
Ettimadai
Coimbatore 641105
Tamilnadu
India
Kp_soman@amrita.edu