**PRELIMINARIES TO THE PREPARATION OF
A WORDNET FOR TAMIL**

**S. Rajendran, Ph.D.**

**1. INTRODUCTION**

The main objective of the project entitled **WORDNET FOR TAMIL** is to capture the network of lexical relations between lexical items in Tamil. 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. 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 net work of relations in a more systematic way.

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. Programs have to be written to capture the net work of relations existing between the lexical items and a user friendly interface has be set up to make use of the Word Net for various purposes. 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. STRATEGIES TO BE ADOPTED IN WORD NET**

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 organised under conceptual fields or semantic domains. The Word Net aims to organize 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 (see Rajendran, 2001). But the Word Net, that is much more efficient and versatile than the paper thesaurus, is available in electronic media. In one sense, Word net is an on-line thesaurus. But its efficiency in bringing out the lexical relations exalts it to a higher position than the thesaurus.

The present **Word Net of Tamil** is aimed to be built on the foundation offered by natural language processing (NLP), taking into account its application in the fields of language teaching and language learning, lexicography, translation, both machine and human, and AI knowledge representation. The ideas propounded by Miller (1991) and Miller, et al. (1993) will be profusely used in the preparation of word net for Tamil.

Rajendran (1978, 1983) has analysed Tamil vocabulary based on componential analysis of meaning that can be used as a source material in the preparation of Word Net for Tamil. A paper thesaurus (Rajendran, 2001) and an electronic thesaurus prepared by Rajendran are available for ready reference.

**3. WORDNET FOR TAMIL**

There are at least two issues - linguistic and computational - to be encountered in preparing the word net for Tamil.

**I. Linguistic issues comprise of the following items:**

1. Classification of lexical items into semantic domains in a hierarchical fashion.
2. Establishing semantic domains and sub domains based on distinguishing semantic or componential features.
3. Selection of lexical items and assigning them to semantic domains.
4. Arrangement of lexical items under terminal domains.
5. Establishing the semantic relation between lexical items based on the lexical relations such as synonymy, compatibility, incompatibility (antonymy, etc.), hyponymy, hypernymy, meronymy, holonymy, troponymy, and entailment.
6. Relating basic meaning with derivational meaning.
7. Bringing out the network of relations existing between the lexical item which include co-occurrence and collocation.

**II. Computational issues comprise of the following items:**

1. Using computer corpus for preparing Word Net.
2. Semi-automatic way of classifying the words.
3. Creation of a database for building Word Net.
4. Designing a user-friendly interface that can help the user to get the needed information.

**3.1. LINGUISTIC ISSUES**

The above listed seven linguistic issues can be condensed into the following three headings:

1. Problem of lexical analysis.
2. Problem of lexical relations.
3. Problem of lexical organization.

**3.1.1. PROBLEMS OF ANALYSIS**

Structural approach to semantics (Lyons, 1977, 1995) comes handy in analysing lexical items of Tamil for building the word net. Lyons (1995: 102) opines,

Looked at form a semantic point of view, the lexical structure of a language - the structure of vocabulary - can be regarded as a network of sense-relations: it is like a web in which each strand is one such relation and each knot in the web is a different lexeme.

The semantic structure of vocabulary of a language can be studied in a precise and systematic way by means of componential analysis of meaning to which the theory of semantic field greatly leans on.

The assumption of lexical field analysis or semantic field/domain analysis is that lexemes can be grouped together into **lexical fields** on the basis of shared meaning and that most, if not all the vocabulary of a language, can be accounted for in this way. The description of meaning, the definition of lexemes, is then undertaken within each lexical field and involves defining each lexemes in relation to the other lexemes in the field.

Probably the most famous attempt to group vocabulary by lexical fields is the ***Roget's Thesaurus of English Words and Phrases*** by Peter Roget, first published by Longman in 1852, and appearing in many editions since.

The problem of lexical field analysis comprises of the following issues:

1. Grouping lexical items in terms of semantic domains or fields.
2. Establishing semantic domains and sub domains.
3. Finding out superordinate terms for a set of lexical items or titles for sub domains.
4. Finding out the distinguishing or componential features which differentiate one lexical item form the other.
5. Finding out componential features which establish different kinds of lexical relationships between lexical items.

Nida (1975a) who was concerned with the preparation of a thesaurus dictionary for Greek gives a detailed theoretical frame work for structuring lexical items in terms of components of meaing. He 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. Supernaturalogical, C. Sensory, D. Emotive,
E. Intellection, G. Communication, G. sociation, H. Control, I. Movement, J. Impact, K. Transfer, L. Complex activities, involving a series of movements or actions

**III. Abstracts**

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

**IV. Relationals**

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

This classification is based on referential meanings and it is not possible to obtain one to one correspondence between the classes of semantic domains and the grammatical classes. There are, of course certain parallel between them,

since on some level of the deep structure entries tend to be represented by nouns, events by verbs, abstracts by qualifiers, and relationals by a number of different features: particles, affixes of case, word order, etc. (Nida, 1975.a:176).

Nida's universal semantic classification can be adopted for Tamil without much drastic changes, though one may come across a number of problems while doing so. It is proposed here to follow Nidas classification of referential meaning for schematizing the architecture of word net for Tamil. As the meanings are realized in terms of lexical items, the surface categorization is taken into account while capturing the underlying conceptual organization.

**3.1.2. PROBLEM OF LEXICAL RELATIONS**

Understanding the network of lexical relations existing between words, broadly speaking, is the prerequisite for the preparation of a word net for Tamil. The important lexical relations that have to be studied are the following:

1. Synonymy
2. Hyponymy and hypernymy
3. Compatibility
4. Incompatibility
5. Meronymy and holonymy
6. Troponymy and entailment
7. Morphological relations
8. Pertainymy

The first 6 relations are sense relations, so can be grouped semantic relations and the last two relations are formal relations. These lexical relations help us to organize and group lexical items into semantic domains, sub domains and lexical sets.

A word acquires its referential meaning as a member of a semantic domain by the common features it shares with other members in that domain, and by having contrasting features that separate it from other members of the domain. It is the semantic relations among words, such as synonymy, hyponymy, compatibility and incompatibility, which help one to classify and organize words in terms of semantic features or components in a hierarchical or orderly fashion. Componential analysis of meaning is a welcome thing to achieve this mission of relating and classifying lexical items by semantic features.

**3.1.2.1. 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. For example, the relation existing between *nduul* 'book' and *puttakam* 'book' is synonymy and *nduul* and *puttakam* are synonyms. (Note that *nd* is a single sound, and is used to refer to the dental nasal in Tamil.) Generally, finding the relation of synonymy among lexical items goes by our intuition about the meanings of the lexical items concerned.

The following could be our two features of our intuition in truying to find the synonymous relation:

1. There are pairs of lexical items or groups of lexical items which show similarity in terms of meaning which we call synonymy.
2. There are pairs of lexical items which are comparatively more similar than other pairs of lexical items.

It is difficult to characterize synonymy. We can solve the problem in two ways:

1. In terms of necessary resemblance and permissible difference.
2. Contextually, by means of diagnostic frames.

The synonyms must not only manifest a high degree of semantic overlap, they must also have a low degree of implicit contrastiveness. Synonyms show bilateral and symmetric relation. Usually, denying one member of a pair of synonyms implicitly denies the other too.

ndii andta puttakatt-aip paTi-tt-aay-aa?
you that book\_ACC read\_PAST\_you\_INTR
'Did you read that book?'

\*illai. ndaan andta nduul-aip paTi-tt-een.
no I that book\_ACC read\_past\_I
\*'No, I read that book.'

In certain types of expression, synonyms occur together.

avaL avan-aik konRuviT-T-aaL. ataavatu kolai ceytuviT-T-aaL
she he\_ACC kill\_PAST\_she. that is murder do\_PAST\_she
'She killed him, that is she murdered him.'

The synonyms are identical in respect of 'central' semantic traits, but differ, if at all, only in respects of 'minor' or 'peripheral' traits. Within the class of synonyms some pairs of items are more synonymous than others. This raises the possibility of a scale of synonymity of some kind.

**3.1.2.2. HYPONYMY AND HYPERONYMY**

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 *pacu* 'cow' and *vilangku* 'animal' and *erumai* 'buffalo' and *vilangku* 'animal' is hyponymy and *pacu* and *erumai* are co-hyponyms. *vilangku* is the hypernym of *pacu* and *erumai*. Hyponymy is unilateral and asymmetrical. (Note that *ng* stands for the velar nasal sound.)

avaL talai-yil mallikai vaittiru-kkiR-aaL
she head\_LOC flower keep\_PRES\_she
'She is wearing jasmine on her head.'

avaL talai-yil puu vaittiru-kkiRaaL
she head\_LOC flower keep\_PRES\_she
'She is wearing flower on her head.'

avaL talai-yil puu vaittiru-kkiR-aaL
she head\_LOC flower keep\_PRES\_she
'She is wearing jasmine on her head.'

\*avaL talai-yil mallikai vaittiru-kkiR-aaL
she head\_LOC jasmine keep\_PRES\_she
'She is wearing jasmine on her head.'

Hyponymy shows transitive relation.

vilangku 'animal' > paaluuTTi 'mammal'
paaluuTTi 'mammal' > pacu 'cow'
vilangku 'animal' > paaluuTTi 'mammal' > pacu 'cow'

**3.1.2.3. COMPATIBILITY**

The term **compatibility** (Lyons, 1977 vol.1) denotes the semantic relation existing between lexical items that 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. Take for example, the words *ndaay* 'dog' and *cellappiraaNi* 'pet'. A dog could be a pet, but neither all pets are dogs nor all dogs are pets. The relationship existing between *ndaay* and *cellappiraaNi* is compatible.

**3.1.2.4. INCOMPATIBILITY**

**Incompatibility** (Lyons, 1977 vol. 1) refers to sets of items where the choice of one item excludes the use of all the other items from that set. The relation existing between puunai and naay is incompatible. Both come under the superordinate term *vilangku* '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 hyponymic 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.

**3.1.2.4.1. BINARY OPPOSITION**

Antonymy, which is often considered opposite to synonymy, relies on the lexical relation, incompatibility. The table below gives the typology of binary opposition (Lyons, 1977, vol. 1).

|  |  |
| --- | --- |
| Type | Example |
| 1.1. Antonymy | peritu 'big' vs. ciRitu 'small' |
| 1.2. Complementarity | aaNmai 'manliness' vs. peNmai 'feminity' |
| 2.1. Privative opposition | ahRiNai 'irrational' vs. uyartiNai 'rational' |
| 2.2. Equipollent opposition | aaN 'male' vs. peN 'female' |
| 3.1. Reciprocal Social roles | maruttuvar 'doctor' vs. ndooyaaLi 'patient' |
| 3.2. Kinship Relations | appaa 'father' vs. makan 'son' ammaa 'mother' vs. makaL 'daughter' |
| 3.3. Temporal Relations | munnar 'before' vs. pinnar 'after' |
| 3.4. Spatial Relations | melee 'above' vs. kiizee 'below' |
| 3.5. Complex Relations | vaangku 'buy' vs. vil 'sell' |
| 4.1. Directional Opposition | vandtuceer 'arrive vs. puRappaTu 'depart' vaa 'come' vs. poo 'go' |
| 4.2. Orthogonal Oppositionor Perpendicular Opposition | vaTakku 'north' vs. kizakku 'east' and meeRku 'west' kizakku 'east' vs. teRku 'south' and vaTakku 'north' |
| 4.3.Antipodal Opposition or Diametrical Opposition | vaTakku 'north' vs. teRku 'south' kizakku 'east vs. meeRku 'west' |

**3.1.2.4.2. 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. 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**

↓

onRu 'one'

↓

iraNTu 'two'

↓

muunRu 'three'

↓

ndaanku 'four'

↓

aindtu 'five'

**CYCLE**

civappu 'red'

aaranjcu 'orange'

uutaa 'purple'

manjcaL 'yellow' ndiilam 'blue

paccai 'green'

**SCALE**

↓

mikandanRu 'excellent'

↓

ndanRu 'good'

↓

tirupti 'satisfactory'

↓

moocam 'bad'

↓

mika moocam 'very bad

**RANK**

↓

onRu 'one'

↓

iraNTu 'two'

↓

muunRu 'three'

↓

ndaanku 'four'

↓

aindtu 'five'

**3.1.2.5. MERONYMY AND HOLONYMY**

Meronymy (part-whole relation) also plays an important role in the hierarchical arrangement of lexical items. Cruse (1986:159) uses the term **meronymy** to denote "the semantic relation between a lexical item denoting a part and that denoting the corresponding whole". The relation that is reverse to meronymy is **holonymy** (whole-part relation). Meronymic relation is transitive (with qualification) and asymmetrical (Cruse, 1986) and can be used to construct a part hierarchy (with some reservations, since a meronym can have many holonyms). The division of the human body into parts can serve as a prototype for all part-whole hierarchies:



Meronymy plays an important role in the hierarchical arrangement of nouns. The part-whole relationship forms the basis for the classification vocabulary into different lexical fields.

The pat-whole relationship which holds between individual lexemes and the lexical filed within which they are interpreted, is identical with, or at least similar to, the part-whole relationship which holds between lexical fields and the totality of the vocabulary (Lyons, 1977, vol.1:253)

**3.1.2.6. TROPONYMY AND ENTAILMENT**

The semantic relation existing between lexical items in which the semantic composition of one includes the other is called entailment. This is discussed elaborately under the subheading 'organization of events.'

**3.1.2.7. MORPHOLOGICAL RELATIONS**

The derivational and inflectional meaning relations have to be taken care of under morphological relations. The relation between *vaa* 'come' and *vandtaan* 'he came', *vandtavan* 'he who came', *varavu* 'income' and *varukai* 'arrival' has to be included under morphological relations. Also the derivation of adjective, *azakaana* 'beautiful', and adverb, *azakaaka* 'beautifully', from *azaku* beauty' is a matter of morphological relation. Morphological relations existing between lexical items comes under formal relations, differentiating them form pure semantic relations.

**3.1.2.8. PERTAINYMY**

Relating descriptive adjectives with the particular nouns they pertain to is known by the term **pertainymy** (see under the subheading 'organizations of adjectives').

**3.1.3. PROBLEM OF LEXICAL ORGANIZATION**

The problem of lexical organization involves the following issues:

1. Bringing out the net work of lexical relations existing between lexical items.
2. Arrangement of lexical items.
3. Relating basic meaning with derived meaning.
4. Tackling co-occurrence and word associations.

**3.1.3.1. ORGANIZATION OF ENTITIES**

Nida's (1978) classification of entities is given above. Rajendran (1983, 2001) has elaborately studied the entities and made a finer classification of the entities in terms of componential analysis. The entities are represented as nouns in the surface level or formal level. It is proposed to make use of Rajendran's classification for building the word net for nouns. Relations pertaining to entities can be captured by the lexical relations such as synonymy, hyponymy, compatibility, incompatibility and meronymy which have been elaborately discussed in the previous sections. The following table sums up the lexical relations of entities to be captured in the word net.

|  |  |  |
| --- | --- | --- |
| Relations | Subtypes | Example |
| Synonymy | - | puttakam 'book' to nduul 'book |
| Hypernymy-Hyponymy | - | vilangku 'animal' to paaluuTTi 'mammal' |
| Hyponym-Hypernymy | - | pacu 'cow' to paaluuTTi 'mammal' |
| Holonymy-Meronymy | From wholes to parts | meecai 'table' to kaal 'leg' |
| " | From groups to their members | tuRai 'department' to peeraaciriyar 'professor' |
| Meronymy-Holonymy | From parts to wholes | cakkaram 'wheel' to vaNTi 'cart' |
| " | From members to their groups | paTaittlaivar 'captain' to paTai 'army' |
| Opposites | Antonymy (gradable opposites) | ndallavan 'good person' to keTTavan 'bad person' |
| " | Complementarity (a item complement another item) | aaN 'male' to peN 'female' |
| " | Privative opposition(presence of a feature implies the absence of another) | ahRiNai 'irrational' to uyartiNai 'rational' |
| " | Equipollent opposition(both the items have positive features) | aaN 'male' to peN 'female' |
| " | Reciprocal Social roles | vaittiyar 'doctor' to ndooyaaLi 'patient' |
| " | Kinship Relations | appaa 'father' to makan 'son';ammaa 'mother' to makaL 'daughter' |
| " | Temporal Relations | munnar 'before' to pinnar 'after' |
| " | Orthogonal Opposition orPerpendicular Opposition | vaTakku 'north' to kizakku 'east' and meeRku 'west' kizakku 'east'to teRku 'south' and vaTakku 'north' |
| " | Antipodal Oppositionor Diagonally opposite relation | vaTakku 'north' to teRku 'south' |
| Multiple opposites | Serial | onRu 'one', iraNTu, muunRu 'three', ndaanku 'four', and so on. |
| " | Cycle | njaayiRu 'Sunday' to tingkaL 'Monday 'to cevvaay 'Tuesday' to putan 'Wednesday'to viyaazan 'Thursday' to veLLi 'Friday' cani 'Saturday' |
| Compatibility | - | ndaay 'dog'to cellappiraaNi 'pet' |
| Lexical association | - | - |
| Collocation | - | kaakam'crow' to karai 'cry' (in sentence kaakam karaiyum 'The crow caws')cingkam to karji (as in sentence cinkam karcikkum 'The lions roars' |
| Morphological relations | - | paTi 'study' to paTittavan 'educated man' |

**3.1.3.2. ORGANIZATION OF EVENTS**

The semantic domain EVENTS comprises of verbs and the abstract nouns derived from them. Nida's (1976b) tentative classification of events into twelve semantic domains based on componential analysis has been given already. Events are mostly realized in the surface level as verbal forms. Rajendran (1978) classified verbs into 31 groups out of which nine are major important semantic domains. The important semantic domains identified by him based on componential analysis of verbs are:

1. Verbs of movement (i.e. change of position).
2. Verbs of transferring (change of possession).
3. Verbs of change of state (change of shape, condition, etc)
4. Verbs of impact.
5. Verbs of senses.
6. Verbs of emotion.
7. Verbs of intellection.
8. Verbs of communication and calling.
9. Verbs of association.

Each major domain is divided into sub domain by taking into account distinguishing semantic component. The classification may need second look to make it more user-friendly. Even though verbs do not show hierarchical ordering, a quasi-hierarchical ordering is possible by taking into account certain pertinent distinguishing features. For wider coverage of verbs, it is proposed to follow the twelve-way classification of verbs by Nida (1975a) and this tentative classification is liable to change to accommodate more verbs.

**3.1.3.2.1. POLYSEMOUS NATURE OF VERBS**

The verbs are fewer in number than nouns in Tamil and at the same time verbs are more polysemous in nature than nouns. The semantic flexibility of verbs makes the lexical analysis of verbs difficult. A look at the Tamil corpus by listing the contexts in which a particular verb is used will reveal the polysemous behaviour of verbs. The polysemy will be captured in line with Nida (1978). The following table shows the different senses in ooTu is used (Rajendran, 1978).

|  |  |
| --- | --- |
| **Different senses** | **Examples** |
| **I. Movement** |  |
| 1. run as animals | avan pattu mail tuuram ooTinaan 'He ran for about ten miles' |
| 2. run as a liquid in a channel, river, tube, vessel, etc. (Note: poo 'go' can replace ooTu in this context.) | aaRRil veLLam ooTukiRatu 'The water is running in the river' |
| 3. work as a machine (the movement of which can be seen from the movement of wheels). | kaTikaaram ooTukiRatu 'The clock is running' |
| 4. run as vehicles (generic locomotion); ply. | rayil taNTavaaLattil ooTukiRatu 'Train moves on tracks' kappal taNNiirril ooTukiRatu 'Ship moves in water' cennaiyilirundtu kanniyaakumaarikku bas ooTukiRatu 'Buses are playing between Chennai and Kanyakumari' |
| 5. escape | avan miinaip piTikkap poonaan, aanaal atu ooTiviTTatu 'He tried to catch the fish, but it ran away' avan vaNNattuppuucciyaip piTikkap poonaan, aanaal atu ooTiviTTatu 'He tried to catch the butter-fly, but it ran away' |
| 6. elope (Note: The compound ooTippoo 'having run go' also gives the meaning 'elope away') | avaL avan kuuTa ooTiviTTaaL 'She eloped away with him' |
| **II. Abstract Movement** |  |
| 1. run or go on as a performance, a business, an organization, life, etc. (Note: ndaTa 'walk; happen' can be used in the place of ooTu in all these contexts. Running of dance or drama performance cannot be denoted by the verb ooTu.) | anta tiyeeTTaril oru cinimaa ooTukiRatu 'A cinema is running in that theatre' viyaapaaram ndanRaaka ooTukiRatu 'The business is going on well' kampani ndanRaaka ooTukiRatu 'The company is running well' vaazkkai eppaTiyoo ooTukiRatu 'The life is going on somehow.' |
| 2. pass quickly as time. | ndaan inkee vandtu muunRu varuTankaL ooTiviTTana 'Three years have passed after my coming over here.' |
| 3. by capable of comprehending, doing work, etc. (Note: poo 'go' is synonymous to ooTu in this context. Note: The verb in this context receives a dative-subject. The verb vaa 'come' can also be used in the place of ooTu. ooTu in this context when compounded with the negative auxiliaries illai and maaTu gives the meaning 'be incapable or paralyzed'. | enakku kaNakku ooTum 'I can comprehend mathematics' avanukku kottaveelai ooTum 'He can do masonry' enakku kaNakku ooTavillai 'I could not comprehend mathematics' enakku veelai ooTamaaTTeen enkiRatu 'I am unable to work' avanaip paarttatum enakkuk kaiyum kaalum ooTavillai '(As soon as I saw him I could not operate my hands and legs) I was paralyzed by seeing him' enakku onRumee ooTavillai 'I could not do anything (I am inactive). |

For each sense/meaning the related synonymous, hyponymous, antonymous and other lexically related items will be given in the word net.

**3.1.3.2.2. 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 kol 'kill' can be decomposed into 'cause not to become alive'. The verb eRi 'throw' can be decomposed into 'cause an object to move away from one's possession by force'. The decompositional nature of verbs reveals the entailment relation existing between verbs. For example, the entailment of simple verb under causative verb (ex. ooTu 'run' vs. ooTTu 'cause to run') is understood by decompositional nature of verbs. The decompositional features of verbs can be captured by the componential analysis of verbs into finer semantic components (Leech, 1974). All types of lexical relations such as synonymy, entailment, hyponymy and troponymy and sentential properties such as presupposition, inconsistency, tautology, contradiction, and semantic anomaly can be mapped clearly if verbs are decomposed into componential features.

**3.1.3.2.3. SYNONYMY AMONG VERBS**

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

kol 'kill' and kolai cey 'murder'
vicaari 'enquire' and vicaaraNai cey 'investigate'

The synonymous expressions of many verbs show that they are manner elaborations of more basic verbs. For example, viniyooki 'distribute' can be considered as an elaboration of the basic verb koTu 'give'. The more effective way of depicting the lexical and semantic relations among verbs is to establish these relations in terms of different senses of each verb.

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

The entailment relation between verbs discussed above is similar to meronymy found between nouns, but meronymy is more suitable to nouns than to verbs. Fellbaum and Miller (1990) argue that, first, verbs cannot be taken as parts in the same way as nouns, because the parts of verbs are not analogous to the parts of nouns. Most nouns and noun parts have distinct, delimited referents. The referents of verbs, on the other hand, do not have the kind of distinct parts that characterize objects, groups, or substances. Componential analyses have shown that verbs cannot be broken into referents denoted solely by verbs. It is true that some activities can be broken down into sequentially ordered sub-activities, say for example camai 'cook' is a complex activity involving a number of sub-activities. Consider the relation between the verbs vdangku 'buy' and koTu 'pay'. Although neither activity is a discrete part of the other, the two are connected in that when you buy something, somebody gives it to you. Neither activity can be considered as a sub-activity of the other. Consider the relations among the activities denoted by the verbs kuRaTTaiviTu 'snore', kanavukaaN 'dream', and uRanku 'sleep'. Snoring or dreaming can be part of sleeping, in the sense that the two activities are, at least, partially, temporally co-extensive; the time that you spend snoring or dreaming is a proper part of the time you spend sleeping. And it is true that when you stop sleeping you also necessarily stop snoring or dreaming. The relation between pairs like vangku 'buy' and koTu 'pay' and kuRaTTaiviTu 'snore' and uRagnku 'sleep' are due to the temporal relations between the members of each pair. The activities can be simultaneous (as in the case of vaangku 'buy' and koTu 'pay' or one can include the other (as in the case of kuRaTTaiviTu 'snore' and uRangku 'sleep').

**3.1.3.2.5. HYPONYMY AMONG VERBS**

Some verbs seem more generic than others. For example, koTu 'give' describes a wider range of activities than viniyooki '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 Tamil (Rajendran, 1978) reveals the fact that the semantic component such as +DIRECTION (eg. eeRu 'climb up' vs iRanku 'climb down'), +MANNER (eg. ndazuvu 'slip down' vs vizu 'fall') + CAUSE (eg. ooTu 'run' vs. ooTTu 'cause to run', +SPEED (e.g. uur 'crawl' vs ooTu 'run) added to the common semantic component +MOVE establish co-hyponymous relation found among verbs of motion. Miller (1991) makes use of the term troponymy to establish this type of relation existing between verbs. "When two verbs can be substituted into the sentence frame To V1 is to V2 in a certain manner, then V1 is a troponym of V2" (Miller, 1991:228). For example, ndoNTu 'to walk unevenly' is a troponym of ndaTa 'walk' as the former entails the latter.

**3.1.3.2.6. 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 ndoNTu 'limp' and ndaTa 'walk'. The verbs in this pair are related by troponymy: ndoNTu is also ndaTa in a certain manner. So ndoNTu is a troponym of ndaTa. The verbs are also in entailment relation: the statement avan ndoNTukiRaan 'he is limping' entails avan ndaTakkiRaan 'he is walking'.

In contrast with pairs like ndoNTu 'limp' and ndaTa 'walk', a verb like kuRaTTaiviTu 'snore' entails and is included in tuungku 'sleep', but is not a troponym of tuungku. Similarly vaangku 'buy' entails koTu 'give', but is not a troponym of koTu 'give'. The verbs in the pairs like kuRaTTaiviTu 'snore' and tuungku 'sleep' are related only by entailment and proper temporal inclusion. It can be generalized that the verbs related by entailment and proper temporal inclusion cannot be related by troponymy. If the activities denoted by two verbs are temporally co-extensive, they can be linked by troponymy. Troponymy represents a special kind of entailment. The following tree diagram adopted from Fellbaum (1993) 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. More over, within a semantic domain, not all verbs can be grouped into a single hierarchy, under a single unique beginner.

**3.1.3.2.7. OPPOSITION RELATIONS AND ENTAILMENT**

Opposition relations are psychologically significant not only for adjectives, but also for verbs. It is found that after synonymy and troponymy, opposition relations are the most frequently coded semantic relations in building database for verbs. The semantics of opposition relations among verbs is complex. As for as Tamil is concerned there is no morphologically derived opposite verbs. Some of the oppositions found among nouns are absent in verbs. A number of binary oppositions have been shown by the verbs that include converseness, directional, orthogonal, and antipodal oppositions. Active and passive forms of transitive verbs can be taken as showing converse opposition. avan avaLaik konRaan is in converse relation with the passive expression avaL avanaal kollappaTTaaL. Thus active-passive pairs of transitive verbs in Tamil show converse opposition. The relation between the verbs vaangku 'buy' and vil 'sell' is rather more complex. The lexical items that are directionallly opposite are in direcional opposition. The relationship which hold between the pairs such as vandtuceer 'arrive' and puRappaTu 'reach', vaa 'come':and poo 'go' is directional opposition. Under this category are the verb pairs such as uyar 'rise' and taaz 'go down', eeRu 'ascend' and iRangku 'descend'. There are many other oppositions with reference to change of state, manner, speed, etc. as exemplified below:

|  |  |
| --- | --- |
| kaTTu 'build' | iTi 'demolish' |
| kaTTu 'tie' | aviz 'untie' |
| ottukkoL 'agree' | maRu 'disagree' |
| uLLizu 'inhale' | veLiviTu 'exhale' |
| ndaTa 'walk' | ooTu 'run' |

Not only the opposing features, even the presence or absence of a feature can also keep two items in opposition relation. These contrasting or distinguishing features can be arrived at by componential analysis of verbs (Rajendran, 1978).

The componential analysis of verbs shows that many verb pairs in an opposition relation also share an entailed verb. For example the pair jeyi 'succeed' and tool 'fail' entails muyal '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, 1993). 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 (1993):



**Entailment (Three kinds of entailment)**

**3.1.3.2.8. CAUSATION AND ENTAILMENT**

The causative relation exists between two verbal concepts: one is causative (e.g. koTu 'give) and the other is resultative (e.g. peRu 'get'). Causation can be considered as a specific kind of entailment: if V1 necessarily causes V2, then V1 also entails V2.

veLiyeeRRu 'expel' entails veLiyeeRu 'leave'
uyarttu 'raise' and uyar 'rise' (temporal inclusion)

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



**3.1.3.2.9. SNTACTIC 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 word net has to be explored for the better understanding of verb net.

**3.1.3.2.10. SUMMING UP VERBNET**

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

|  |  |  |
| --- | --- | --- |
| **Relations** | **Definition/sub types** | **Example** |
| Synonymy | Replaceable events | tuungku 'sleep' → uRangku 'sleep' |
| Meronymy- Hypernymy | From events to superordinate events | paRa 'fly' → pirayaaNi 'travel' |
| Troponymy | From events to their subtypes | ndaTa → ndoNTu 'limp' |
| Entailment | From events to the events they entail | kuRaTTaiviTu 'snore' muyal 'try' tuungku 'sleep' |
| " | From event to its cause | uyar 'rise' → uyarttu 'raise' |
| " | From event to its presupposed event | vel 'succeed' → muyal 'try' |
| " | From even to implied event | kol 'murder' → iRa 'die' |
| Antonym | Opposites | kuuTu 'increase' → kuRai 'decrease' |
| " | Conversensess | vil 'sell' → vaangku 'buy' |
| " | Directional opposites | puRappaTu 'start' → vandtuceer 'reach' |

**3.1.3.3. ORGANIZATION OF ABSTRACTS**

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

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

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

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

Rajendran (2001) has classified abstracts in which adjective forms a part into 38 sub domains by taking into account the componential features of meaning and classification of Nida (1978) and Dixon (1982). The items are represented mainly in their nominative forms and the adjectival and adverbial forms derived from them or related to them by componential analysis of meaning are listed along with them. The lexical sets are built taking into account the above mentioned classification and the morphological relation takes into account the derivative relation between nouns, adjective and adverbs.

Fellbaum, et al (1993) describes in detail the organization of adjectives in word net. Languages exhibit a common phenomenon by providing some means of modifying or elaborating the meanings of nouns. They may show difference in the syntactic structure by means of which such modification is made. Tamil syntax exhibits different of ways to express the qualification of a noun. For example, if a speaker is not satisfied with the word ndaaRkaali 'chair', he may make use of modifiers such as periya 'large' and vacatiyaana 'comfortable' to denote the object he has in mind more accurately. Words belonging to other syntactic categories such as relative participle form of verbs and nouns can function as adjectives.

**Past participle form as adjectives**
iruNTa viiTu 'dark house'
mangkiya oLi 'dim light'
varaNTa ndilam 'dry land'
**Nouns as adjectives**
talaimai atikaari 'chief officer'
tiruTTu paNam 'illegal money'
iNai aayvaaLar 'co-inverstigator'

Nouns phrases as well as clauses can modify a noun.

avanuTaiya taattaavin ndaaRkaali
'his grandfather's chair'
ndeeRRu kaTai-yil vaangk-iy-a ndaaRkaali
yesterday shop\_LOC buy\_PAST\_RP chair
'the chair which was bought from the shop yesterday'

Adjectives have been regarded as forming a distinct category on the basis of their morphosyntactic, semantic and also functional characteristics. The primary function of adjective as a syntactic category is noun modification. The sole function of adjectives is modification of nouns and thus differs from nouns or prepositional/postpositional phrases whose primary function is not modification. The organization of adjectives differs considerably form that of the other major syntactic categories, noun and verb. The adjective domain contains mostly adjectives, although some nouns and relative participial forms of verbs that function frequently as modifiers have to be incorporated as well.

Though adjectives can be established as a separate grammatical category in Tamil, traditional grammarians have taken it partly as verbs and partly as nouns. There are many bound forms or root forms that are purely adjectival in their character. There are certain inherent characteristics of adjectives that cannot be stated as derived characteristics or features. The relative participle forms of the verbs that come before nouns are also adjectival in their function. So mapping adjectives in word net is a challenging problem for Tamil. Many adjectives in Tamil have their origin in nouns and so the adjectives will be morphologically to their respective nominative forms. For, example azakaana 'beautiful' has to be related to the nominative form azaku 'beauty'.

**3.1.3.3.1. COMPONENTIAL FEATURES OF ADJECTIVES**

The distinguishing feature of the adjective is the fact that it modifies a noun. Another distinctive feature of adjectives is their ability to suggest characterizing qualities. Characterizing and state-denoting adjectives frequently refer to features that are perceived as variable in degree. In that respect the adjectives become gradable. A gradable adjective can be modified by adverbs of degree and occur in comparative and superlative constructions. Adjectives need to be distinguished into two types: descriptive and relational. Descriptive adjectives assign to their head nouns values of bipolar attributes. They are organized in terms of binary oppositions (antonymy) and similarity of meaning (synonymy). Descriptive adjectives that do not have direct antonyms are said to have indirect antonyms by virtue of their semantic similarity to adjectives that do have direct antonyms. Cross references have to be established between descriptive adjectives expressing a value of an attribute and the noun by which that attribute is lexicalized. Reference-modifying adjectives have special syntactic properties that distinguish them from other descriptive adjectives. Relational adjectives can be presumed as stylistic variants of modifying nouns and so are morphologically related to the nouns concerned.

**3.1.3.3.2. DESCRIPTIVE ADJECTIVES**

A descriptive adjective is one that ascribes a value of an attribute to a noun.

atu kanamaana cumai 'that luggage is heavy'

The above sentence presupposes that there is attribute eTai 'WEIGHT' such that eTai (cumai 'luggage') = kanam 'heavy'. In the same way taazndta 'low' and uyarndta 'high' are values of HEIGHT. The word net hast 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 akalamaana 'wide' and kuRukalaana 'narrow' are kept under the semantic domain 'dimension' in which the attribute akalam 'width' is kept. Relating descriptive adjectives with the particular noun they pertain to is known by the term pertainymy.

**3.1.3.3.2.1. 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 kanamaana 'heavy' is ileecaana 'light' that expresses a value at the opposite pole of the WEIGHT attribute. This binary opposition is to be represented in Tamil word net.

**3.1.3.3.2.2. 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 mikamika 'extremely', mika 'moderatley' or konjcam 'slightly'. Antonymy is expressed by pairs such as ndiiNTa 'long'/kuTTaiyaana 'short', viraivaana 'fast'/metuvaana 'slow', culapamaana 'easy'/kaTinamaana 'difficult', ndalla 'good'/keTTa 'bad', cuuTaana 'hot'/kuLirndta '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, uyiruLLa 'living' and uyiraRRa 'non-living' are contradictory terms as atu uryiruLLa jandtu 'it is a living creature' necessarily implies atu uyiraRRa jandtu alla 'it is not a non-living creature'. But kuNTaana 'fat' and melindta 'thin' are contrary terms because maalaa kuNTaana peN 'Mala is a fat girl' and malaa melindta peN 'Mal is 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:

kotikkiRa 'very hot'
cuuTaana ' hot'
iLanjcuuTaana 'warm'
kulirndta 'cold'

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

**3.1.3.3.2.3. MARKED AND UNMARKED DISTINCTION 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 uyaramaana 'high'/kuTTaiyaana 'low, vayataana 'old'/iLamaiyaana 'young', ndiiLamaana 'long/kuTTaiyaana 'short', akalamaana 'wide/kuRukalaana 'narrow'. We measure things by uyaram 'height' rather than kuTTai 'shortness'.

While asking questions about uyaram 'height', we say atu evvaLavu uyaramaana tuuN 'How high that pillar is?'rather than atu evvaLavu kuTTaiyaana tuuN 'How short that pillar is?'. A question X evvaLavu kuTTaiyaanatu 'How short is X?' is felt to contain the assumption that X is short, while no equivalent assumption is present in X evvaLavu uyaramaanatu 'How high is X?' That is, if the two antonyms contrast with reference to a scale of measurement, the unmarked one is capable of referring to a point on that scale, thereby neutralizing the contrast. Thus the primary member, uyaramaana 'high' is the unmarked term; the secondary member, kuTTaiyaana 'short' is the marked one. They are related to the attribute noun uyaram 'height' Word net has to capture the relation between marked and unmarked terms and their cross reference to their variable property.

**3.1.3.3.2.4. POLYSEMY IN ADJECTIVES**

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 ndalla 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.

ndalla kalam 'good time'
ndalla ndaaNayam 'good coin'
ndalla ndaNpan 'good friend'
ndalla ceruppu 'good chappal'

Adjectives are choosy about the nouns they modify. The general rule is that if the referent denoted a noun does not have attribute whose value is expressed by the adjective, then the adjective-noun combination requires a figurative or idiomatic interpretation. For example, a road can be long because roads have LENGTH as an attribute, but stories do not have LENGTH, so ndiiNTa 'long' does not admit literal readings. The selectional preferences of adjectives should be captured in the word net by organizing the adjectives under abstracts.

**3.1.3.3.2.5. SYNTAX OF ADJECTIVES**

In Tamil the premodifying form of the adjective is different from the postmodifying form, as the premodifying form is a real adjective whereas the postmodifying form is a pronominalized noun. For example, for the adjective form ndalla 'good' which comes before a noun, the predicative forms are the pronominalized forms, ndallavan 'good male person', ndallavaL 'good female person' ndallavar 'good person', ndallatu 'good thing'.

ndalla pen 'a good woman'
peN ndallavaL 'the woman is good'

Word net should account for the relation between the pure adjectives and their derived nominal forms.

The predicative form can be of the following structure too:

N+ aaka + iru

aaka is an adverbial equivalent of the adjectival formative suffix ana. iru is a be-verb. The structure helps to relate a simple adjectival form to its predicate form. The following example will illustrate the point.

avaL azakaana ciRumi
she beauty\_become\_RP girl
'she is a beautiful girl'

ciRumi azakaanavaL 'the girl is beautiful'
ciRumi azakaaka irukkiRaaL 'the girl is beautiful

**3.1.3.3.3. REFERENCE-MODIFYING AND REFERENT-MODIFYING ADJECTIVES**

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

ndeRRaiya kurravaaLikaL inRaiya mandtirikaL
yesterday's criminals are today's ministers

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

ndeRRaiya 'past' vs. innaaLaiya 'present'
mundtaiya 'past' vs. inRaiya 'present'.

**3.1.3.3.4. COLOUR ADJECTIVES**

Colour terms need to be organized differently than other adjectives. They can be both nominal as well as adjectival. As adjectives they can be graded and conjoined with other descriptive adjectives. But they differ form the descriptive adjectives as the pattern of direct and indirect anotonymy does not hold good for colour adjectives.

**3.1.3.3.5. RELATIONAL ADJECTIVES**

Relational adjectives include of a large and open class of adjectives. Relational adjectives can be defined by using the phrase 'of, relating/pertaining to or associated with some noun', and they play a role similar to that of a modifying noun (Levi,1978). For example, cakootara 'fraternal', as in cakootra paacam 'fraternal love' relates to cakootran 'brother', and poruLaataara 'economical', as in poruLaataara eRRa taazvu 'economical difference', is related to poruLaataaram 'economics'. As for as Tamil is concerned noun form is used mostly in the place of relational adjective in English. For example,

icaik karuvi 'musical instrument'
paR cuttam 'dental hygiene'

Since relational adjectives do not have antonyms, they cannot be incorporated into the clusters that characterize descriptive adjectives. And because their syntactic and semantic properties are a mixture of those of adjectives and those of nouns used as noun modifiers, rather than attempting to integrate them into either structure Tamil word net will maintain a separate file of relational adjectives with cross references to the corresponding nouns.

**3.1.3.4. ORGANIZING RELATIONS**

Nida classifies relationals into a few types and he has listed few of them (Nida, 1978: 186)

A. Spatial: 'up', 'down', 'before', 'behind', 'through', etc.
B. Temporal: 'when', 'while', 'during' 'since', etc.
C. Deictic: 'this', 'that', 'former', 'latter', 'the' (definite), 'a' (idefinite) etc.
D. Logical: 'since', 'because', 'in order that', 'whereas', 'although', 'moreover', 'therefore', 'however', 'but', 'and', etc.

Etc.

Following his footsteps, Rajendran (2001) has also classified relationals in Tamil in terms of the above mentioned semantic sub-domains. The postpositions functions as predicates in the logical sense linking a noun with another noun by means of relations which include spatial and temporal relations. The Tamil word net proposes to capture these relations.

**3.2. COMPUTATION ISSUES**

The computational issues listed in the section two can be clubbed into two:

1. The problem of collecting lexical information
2. The problem of designing and implementing word net

**3.2.1. THE PROBLEM OF COLLECTING LEXICAL INFORMATION**

The usefulness of lexical relations in linguistic, psycholinguistic, and computational research has led to a number of efforts to create large electronic databases of such relations. Efforts to create such databases have, in general, followed one of two basic approaches: mining information from existing dictionaries and thesauri, and handcrafting a database from scratch. Corpus can be made use of for the creation of database for making a word net. A tagged corpus can help us to prepare a word net more effectively. Dictionaries can be used as a secondary source. The paper thesaurus prepared by Rajendran (2001) can be used to create the database for the preparation of word net. The hierarchical classification of vocabulary of Tamil is available in the above mentioned paper thesaurus. If dictionaries and thesaurus are available in the electronic media, they can be used for the purpose of developing a database for word net.

**3.2.1.1. COMPUTER CORPUS**

A computer corpus, as we understand, is a large body of naturally occurring computer-readable texts or text extracts used for research, and especially for the development of natural language processing. Any work related to lexicon presumes a computer corpus these days. Using a computer corpus for lexical work makes things easy. Also the corpus helps in making any decision with maximum perfection. A corpus annotated for grammatical categories and semantic information will be a useful tool in word net making. Automatic tagging for grammatical information is feasible, but automatic tagging for semantic information needs a knowledge base, the building up of which is a difficult task. Building a separate corpus for the preparation of word net is a welcome thing, though one can make use of an already available Tamil corpus, for example, that which is prepared by the CIIL, Mysore, under their project **Development of Corpora of Texts of Indian Languages in Machine Readable Form**.

**3.2.1.2. SEMANTIC PROCESSOR**

What is needed form a corpus for the thesauric classification of lexical item into semantic domains is semantic information in terms of the semantic relations such as synonymy, hyponymy, compatibility and incompatibility. The lexical relations like meronymy and metonymy have also to be taken into consideration for a wider perspective.

TEXTS → PROCESSING → SEMANTIC INFORMATION

Apart from the conceptual information or information concerning the possible relations between meanings of lexical items, the following information are also welcome:

1. Definitional information or information concerning the internal meaning structure of lexical items
2. Contextual information or information concerning the meaning of a lexical item with reference to the context in which it is used
3. Collocational information or information concerning the meaning of a lexical items in sequence

Automatic tagging, annotating and coding the text for the above mentioned semantic information is a difficult task. This can be partially achieved through an efficient semantic parser which is constructed to bring out the network of meaning relations existing between lexical items. A morphological analyzer and syntactic parser can help to categorize the lexical items in terms of their grammatical functions in sentences. But semantic information requires intelligence on the part of computer, which can be achieved partially incorporating artificial intelligence.

**3.2.2. PROBLEM OF DESIGNING AND IMPLEMENTING WORDNET**

Designing and implementation of word net are the two major tasks assigned to computer scientists. To achieve them they have to work in collaboration with lexicographers and linguists. Once the lexicographers complete their work of collecting the data required for building the word net, the job will be handed over to computer scientists. The semantic information collected on lexical items form the basic building blocks for the computer scientist to construct word net. As the words and meanings are related to one another and mapped as such in word net, it is but natural that the word net gives the impression of an on-line thesaurus. 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 componential analysis of meaning of Tamil vocabulary will be the input for the designer and implementers of word net. The componential analysis will be manipulated to map the word net in terms of different types of lexical sets such as synonym sets, sets of lexical oppositions, multi-member sets (such as scale-sets, serial sets, cyclic sets and rank-sets), hypernym-hyponym sets, meronym-holonym sets, troponym sets, entailment sets, and other relevant lexical sets. The relations between hypernym and hyponym, meronym and holonym, and binary oppositions such as directional opposition, antipodal opposition, orthogonal opposition, converseness, complementarity, etc. and multi member oppositions will be captured in word net by making use of the componential features of lexical items.

The componential analysis captures the way the native speakers of Tamil distinguish one lexical item form the other and relate them in terms of certain meaning relations. In an attempt to model the lexical knowledge of a native speaker of Tamil, word net will be been given detailed information about relations existing between word forms and word meanings. They are the product of a detailed semantic analysis of Tamil vocabulary (Rajendran, 1978, 1983, 2001). The structuring of these relations and the architectural skills involved in building such a system distinguish word net from the conventional way of building a thesaurus or a dictionary in the background of lexicography.

The task of developing the on-line database can be conveniently divided into two interdependent tasks (Beckwith, Miller and Tengi,1993). These tasks bear a vague similarity to the traditional tasks of writing and printing a dictionary:

1. To write the source files that contain the basic lexical data - the contents of those files are the lexical substance of word net.
2. 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 word net system will divided into four parts based on the specific tasks assigned to them:

1. Lexical resource system
2. Compiler system
3. Storage system
4. Retrieval system

**3.2.2.1. LEXICAL RESOURCE SYSTEM**

Word net captures a fairly conventional set of lexical relations, the base set that usually appear in the lexical semantics literature (Cruse, 1989). Two kinds of lexical relations can be recognized: formal and semantic. Formal relations hold between word forms; semantic relations hold between word meanings. In the lexical resource system each lexical item is analysed in terms of their componential features and are arranged hierarchically and ontologically exploiting their componential features. As a result, the lexical items will be grouped into many semantic domains.

If we follow Nida (1976), the lexical items will fall automatically into four major domains: entities, abstracts, event and relationals. Exploiting the superordinate componets the lexical items can be grouped and sub grouped till the terminal domains end up consisting of items of synonymy and multiple oppositions. A single lexical item with contrasting components will fall in different semantic domains. The word net scheme distinguishes between 'lexical items' and 'senses'.

For example, the lexical items paTi would appear in the sense of 'read' under verbs of comprehension and in the sense of 'steps' under parts of a building. Sense distinctions in word net are explicit, but they are not numbered in the convention of the usual dictionaries. On the contrary, the word net notion of sense distinction is an indirect by-product of network connectivity. Word forms will be represented in their familiar orthography. The meaning of each lexical item will be represented by meaning components. From the meaning componets it will be possible to get the meaning in definition. From the componential representation of meaning of the lexical items, the meaning relations between them will be extracted and will be represented as lexical sets like synonym sets, hyponym-hypernymy sets, meronymy-holonymy sets, cyclic sets, serial sets, scale sets, rank sets, and so on. Tamil wordNet will organize entities, events, abstracts and relationals into lexical sets which will be further arranged into a set of lexicographer's source files by syntactic category and other organizational criteria. Nouns, verbs, adjectives and adverbs are grouped according to semantic fields.

The lexical relation holding between forms will be captured by cross reference pointers. Cross reference relates a word form in a set and to a related form in another set. Lexical relations exist between relational adjectives and the nouns to which they are related to, and between adverbs and the adjectives form which they are derived. The semantic relation between adjectives and the nouns for which they express values are encoded as attributes. The semantic relation. between noun attributes and the adjectives expressing their values are also encoded. Antonyms are also lexically related. Synonymy of word forms is implicit by inclusion in the same synonym set. Meronymy can be further specified as a part of something or a substance of something or a member of some group. Holonymy can also be specified in the same manner, each cross reference representing the semantic relation opposite to the corresponding meronymy relation. Many cross references are reflexive, that is, if a set contains a cross reference to another set, the other set should contain a corresponding reflexive cross reference back to the original set. The Compiler can automatically generate the relations for missing reflexive cross references.

**3.2.2.2. COMPILER SYSTEM**

The Compiler System will primarily compile the lexical resource files into a database format and sends it to storage system to facilitate machine retrieval of the information in word net. The Compiler will have several options that control its operation on a set of input files. The Compiler System will compile the lexical resource files, keep the integrity of these files, verify the syntax of the files, resolve cross references, then generate the word net database which will be stored in the storage system for retrieval.

**3.2.2.3. STORAGE SYSTEM**

Storage system will work as an intermediary between Compiler System and Retrieval System. The cooked database of the compiler system will be stored in the Storage System for retrieval.

**3.2.2.4. RETRIEVAL SYSTEM**

For the information scientist the primary focus in word net construction is perhaps 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. A user friendly interface helps the user to get the needed information with great ease. The retrieval system faces the following issues:

1. Transfer of classificatory and network scheme into commands for electronic type setting
2. Transfer of classificatory and network scheme into information retrieval system
3. On-line retrieval: combining the classificatory and network scheme with code searching.

It is important to recognize the difference between a printed dictionary and a lexical database while considering the role of interface. Word net's interface software creates its responses to a user's requests on the fly. Unlike an on-line version of a printed dictionary, where information is stored in a fixed format and displayed on demand, word net's information is stored in a format that would be meaningless to an ordinary reader. 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.

**4. CONCLUSION**

The theme of lexical semantics, computational lexicography, and computational semantics are altering rapidly. The availability of machine-readable resources and newly developed tools for analyzing and manipulating lexical entries make it possible to build a massive word net for a language. In present state of affairs it is quite feasible to build one such tool for Tamil. Building of a word net for Tamil is an immediate requirement in the context of information technology equipped with internet in which the web sites in Tamil are getting added up day by day. For the development of a search engine for Tamil, word net is the intermediate solution. The hierarchical representation of vocabulary of Tamil in terms of lexical relations such as hyponymy-hyperonymy, meronnymy-holonymy, etc. can be used in the construction of information retrieval systems. The theory of lexical semantics throws a flood of light in this line of thinking. It is not possible to show the lexical relations existing between words in a paper in a comprehensive manner. The computer offers us to do so effectively. So, it is high time that linguists and computer scientist should come forward to build a comprehensive word net for Tamil.

\*\*\* \*\*\* \*\*\*

Note: Preparation of Tamil WordNet is an ongoing project in Tamil University in Collaboration with AUKC Research Centre at MIT Campus, Chennai in which the author of the present paper is involved.

**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.

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

Beckwith, R., Miller, G.A. and Tengi, R. 1993. Design and Implementation of the WordNet Lexical Database and Searching Software. (Down loaded from internet).

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. 1993. 'English Verbs as a Semantic Net.' (Down loaded from internet)

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

Fellbaum, C., Gross, D. and Miller, K. 1993. 'Adjectives in WordNet.' (Down loaded form internet).

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

Gross, D. and Katherine, J.M.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. 1993. 'Nouns in WordNet: A Lexical Inheritence System.' (Down loaded from internet)

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.

Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D, and Miller, K. 1993. 'Introduction to WordNet: An On-line Lexical Database.' (Down loaded from internet)

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.

Rajendran. S, 1976. Verbs of Cooking in Tamil. *Wording papers in Linguistics,* Number 1, May, 1976, 1:8-16.

-------, 1978. *Syntax and Semantics of Tamil Verbs.* Ph.D. Thesis. Poona: University of Poona.

-------, 1981. Semantic structure of verbs, 13th All India University Tamil University Tamil Teacher's Association Conference *aayvukkoovai,* vol 2: 305-310.

-------,1982. Semantics of Spatial Expressions in Tamil. 14th All India University Tamil University Tamil Teacher's Association Conference *aayvukkoovai.* vol. 2: 262-67.

----- 1982. Verbs of 'seeing' in Tamil. Poona: *Bulletin of the Deccan College Research Institue,* 41:151-159.

----- 1983. Semantic structure of directional verbs of movement. *Working papers in linguistics,* Deccan College, Poona, 1983, vol.7, 19-37.

----- 1983. Temporal Expressions in Tamil. *Bulletin of the Deccan College Research Institute,* vo1. 42, 138-147.

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

----- 1995. Componential Analysis of 'eating' in Tamil. *PILC Journal of Dravidic Studies,* 5:2.175-181

-----1995. 'Towards a Compilation of a Thesaurus for Modern Tamil.' In: *South Asian Language Review.* 5.1:62-99.

-----, 1996. 'The feasibility of Preparing a Thesaurus using Corpus'. Workshop on Indian Language Corpus and its applications (29,29 October, 1996), Central Institute of Indian Languages, Mysore.

-----,1997. 'Preparation of an Electronic Thesaurus for Tamil. Paper read in Symposium on Natural Language Processing conducted by CALT, Central University of Hyderabad from 21st to 26th March, 1997.

-----, 1997. 'Intricacies involved in the preparation of lexical net for Tamil'. Paper read in DLA conference, Telugu University, Hyderabad.

-----,1998. 'Prerequisite for the preparation of an electronic thesaurus for a text processor'. Paper read in Workshop for Preparation of Thesaurus for Indian languages, Tirupati.

-----, 1999. 'poruTpula vakaipaaTum coRkaLanjciyamum'. *Pulamai,* vol. 25, No.2, pp 47-66.

-----, 2000. 'peyaraTaiyaakkam'. cittiraputtiran, ec., iraaparT cattiya joocap, ta, and paarvati, maa (eds.). ndanjcil. tanjcaavur: ndiyuu vican veLiyiiTTaka,pp 66-97.

-----, 2001. *taRkaalat tamizc coRkaLanjciyam [Thesaurus for Modern Tamil].* Thanjavur: Tamil University.

----, 2001. Preliminaries to the preparation of a Word Net for Tamil. Paper read in National seminar on Computational Linguistics and Dravidian Languages held in Centre of Advanced Studies in Linguistics, Annamalai University from February, 22-24, 2001.

Rajendran, S, S Arulmozi, B Kumara Shanmugam, S Baskaran and S Thiagarajan. 2002. Tamil WordNet. Paper read in First International Conference on Global WordNet held in CIIL, Mysore from January 22-25, 2001.

\*\*\* \*\*\* \*\*\*

[HOME PAGE](http://www.languageinindia.com/index.html)| [Headlines in Indian Vernacula](http://www.languageinindia.com/march2002/geethakumari.html)