

# WordNet Has No ‘Recycle Bin’

B. A. Sharada and P. M. Girish

Central Institute of Indian Languages  
Manasagangotri, Mysore-570 006. India  
Email: sharada@ciil.stpmv.soft.net, drsharada@sancharnet.in,  
pm\_girish@rediffmail.com

**Abstract.** This paper is conceived and prepared to provide an overview of the compound words in the WordNet, the miracle lexicon of the new millennium. Indeed meanings are not expressed by single words only such as noun, verb, etc., but also languages do have many ways to express content and the concept. Compound words are one among them. Wide range of words and expressions are included in the WordNet. They express a clear view on the existence of concepts in language and culture. After a keen verification, it is found that, some very frequent compound words are not included in the WordNet available online. This paper lists out some such frequent compound words in English. As far as WordNet is concerned – this study is more an application oriented than architecture. Algorithms followed in the development of Subject Heading list are suggested.

## 1 Introduction

A compound word is a combination of two or more words used to express a single concept. In English, words, particularly adjectives and nouns are combined to form compound words in a variety of ways. Two words will be joined together by a hyphen “fire-fly” and then joined as one word “firefly” [5]. Meys W J states that “Functionally, compounding is clearly a linguistic economy-mechanism allowing one to express in a concise way something which would otherwise have to be rendered by means of an – often much more elaborate – phrase” [12]. Many studies have been undertaken in evolving the theories of combining two or more words by Aronoff [1], Chomsky [4], Bauer [2], Bresnan [3], Marantz [11], Williams [16], Lieber [9], Roeper and Siegel [15], etc.

The combinations may be among two nouns, an adjective and a noun, a noun and a verb, etc. such as:

N N	postman	N – Noun
N A	color-blind	V – Verb
A N	high school	A – Adjective
A A	super-fine	P – Preposition
P N	under wears	
V N	pound-rice	
V A	diehard	
N V	spoon feed	
A V	deep-fry	
P V	incoming	
V V	drop-kick	
P P	within	

In the above combinations, VA, VN and PA are predicted to be a rare possibility [10]. But few exceptions could be found for VN in the context of Indian English. For example: boiled rice.

## 2 Need for the Study

The need for the present study arose while doing linguistic analysis of some texts relating to language learning and information retrieval applications. The powerful online database “WordNet” was checked and it was found that many terms that we call compound words could not be located in the WordNet. Hence these terms were separately listed so that they could be included in the WordNet and make it more comprehensive.

## 3 Source

The words are collected from the articles that appeared in periodicals, newspapers and other mass media published in India. In order to make the study more wide some more words were collected and checked for which, intuitive knowledge was one of the criteria for the data collection.

## 4 Compound Words and Analysis

The treatment of compound words in WordNet was very insignificant in its earlier version 1.6. Some of the compound words got entered as one word in the later version. That is, the orthographic representation of a compound word will be entered as one term without giving any space or hyphen. The word “compound” itself has become a part of such words like – compound fraction, compound fracture, compound interest, compound word, compound eye, etc. When the search word “compound word” was entered in the WordNet for different senses, ‘Sorry, no matches found.’ was displayed.

There are three forms of compound words [5]:

- a. Closed form: words joined together such as – keyword, textbook, lineup, newspaper, etc.
- b. Hyphenated form: Words joined with a hyphen such as – World-wide, Indo-Aryan, Mother-in-law, brief-case, etc.
- c. Open form: neither of the above such as – Compound word, Preview theater, Match box, etc.

All the three forms mentioned above are present in WordNet.

For closed form example – wildfire, mailman, manhood, etc. That is, compound words are without hyphen and space (pre-nominal entered as prenominal). In such a case, words listed in the present study also could be treated in the similar way.

With regard to hyphenation, WordNet has stated, “the hyphenation presents special difficulties when searching WordNet” [13]. But in the recent version [7] some of the compound words have hyphen in the middle [Cross-country].

In case of Open form, space is considered as a delimiter in WordNet [13]. Example: sky blue, white collar, etc.

In pursuance of the dictionary of compound words in the search-engine, Dictionary: compound [7] was located (updated up to July 23<sup>rd</sup> 2003). This Hyper dictionary has English dictionary, Dream dictionary, domain specific dictionaries such as Computer and Medical and a thesaurus. The definition of the word along with its grammatical category in brackets is provided with link to each and every word used in the definition. Synonyms and 'See Also' entries are followed. Here also each word has a hyperlink.

The compound words are listed in **Appendix 1**, which were tested in the WordNet, as they are. Among the 180 compound words, 50 words that are in italics were located in the WordNet. Some of the usages of the compound words that were not found in the WordNet can be seen in the following ways:

1. In some cases, though the affixes such as 'co-', 'sub', 'super', 'pre', 'hood', etc., have semantic value, they cannot function as independent words, in their affix-meanings [12].
2. The semantic elements of compound words are different from what the words actually represent as primary meaning. A specific meaning is obtained only when they are used together. In this case, both semantics and pragmatics have wide role to play in dissecting the meaning.

For example:

Operation flood	- Use and production of milk products in a large quantity.
Collective unconscious	- Is a Freudian terminology to express a sort of socio-mental attitude.
Fall guy	- A person who is punished for the wrong doing of another person.
Recycle bin	- To treat a computer file that has already been used so that it can be used again. It is a component in all the computers.
Lion hearted	- A person having hard nature

3. Some compound words have the thematic or connotative meaning which is completely different from the primary meaning. In the initial stages, it will have limitations in its frequency of usages.

Limitation may be among – age, gender, profession and other social variables such as religion, education, etc.

For example, the thematic or connotative meaning for:

Chief minister	- A person who takes a decision in a family
Central Government	- Parents
Tree cover	- Fresh look
Snake gourd	- Very thin person

4. It is a known fact that language is culture bound. So various culture specific words can also be seen.

For example:

Auto rickshaw - Auto rickshaw is a three-wheeled vehicle and an economic variety of transportation.

Mid day meal - Mid day meal refers to the meal that is offered in the school for children free of cost to promote education in economically backward community. Though there is a word in English as 'Lunch' it is not used in this context for differentiating.

Regional language - Regional language is the language that is in currency in a particular state or a part in the union and in totality of a region.

Panchayat raj - Village administration

Like wise, terms like Snow-clearing may be in currency in the place where snow fall is a routine matter.

Let us look at the compound words such as: gang shooting, breast-feeding and food poisoning. Although shoot, feed and poison are typically used as transitive verbs, the meanings are not compatible with interpretation such as "to shoot gangs", "to feed breast" and "to poison food". Rather, gang shooting is a shooting incident somehow related to gang activities, breast-feeding is a way to feed babies, and food poisoning is a case of illness caused by unsanitary food [14]. The head word usually at the right side of the compound word gives clue to the description of the meaning. Though the latter two are added in the hyper Dictionary few terms like 'Gang shooting' are not found in the WordNet. The conventional meanings of some of the other terms are mentioned in **Appendix 2**.

## 5 Conclusion

As Lieber, Rochelle states that, a major goal of current linguistic research is to construct a theory of the lexicon which allows us to characterize the notion of possible word in a simple manner with a minimum of theoretical machinery. Such a theory would ideally predict the possibility of certain sorts of inflected or derived forms, compound, and reduplicated words, while ruling out others [10]. WordNet has mentioned in its third objective that "meanings are not just expressed by nouns and verbs or single words. Language uses a variety of ways to express content ..." [13]. In addition to this, WordNet has improved much within a span of two years. In 2002, WordNet hardly included compound words. It may be recalled here that in the GWN 2002 conference held at the Central Institute of Indian Languages, Mysore, India, it was discussed in the concluding session to include compound words in its lexicon. But now a hyper dictionary is available on the net and that is a tremendous development in WordNet.

For some words in general category and domain specific compound words, it is suggested that the algorithm followed in constructing List of Subject Headings (SH) could be followed. SH is a part of Indexing Language and is sharp and equal to summarized text. In SH the importance is given only to the concepts and not to the structure words. If the SH contains two words it will be the combination of an Adjective and a Noun. This order will be inverted to give importance to the Noun. For Example: 'pumping machinery' will be rendered as 'machinery, pumping'. Controlled vocabulary is used in forming the concepts [8]. A

controlled vocabulary contains a unique term for each meaning. Also this may not hold well in all compound words.

This study shows that there is a great potential for WordNet to deal with compound words appearing not only in different grammatical categories but also from all disciplines including interdisciplinary and multidisciplinary research.

## References

1. Aronoff, M.: Word Formation in Generative Grammar. Linguistic Inquiry Monograph I. MIT Press, Cambridge, Mass (1976).
2. Bauer, Laurie.: English Word-formation. Cambridge University Press, Cambridge (1984).
3. Bresnan, J.: Passivization: Part II of a theory of Lexical Rules and Representations. MIT Press, Cambridge, Mass (1980).
4. Chomsky, N.: Lectures on Government and Binding. Dordrecht, Foris (1981).
5. Compound Words. <http://webster.comnet.edu/grammar/compounds.htm> (2003).
6. Fairbairn, G. J and Winch. C.: Reading, Writing and Reasoning: A guide for all students. SRHE and Open University Press, Buckingham (1991).
7. Hyperdictionary. <http://www.hyperdictionary.com/dictionary/compound> (2003).
8. LCSH.: Library of Congress Subject Headings. 17<sup>th</sup> Edition. Library of Congress, Washington D. C. (1993).
9. Lieber, Rochelle: On the organization of the lexicon. Doctoral dissertation. MIT Press, Cambridge, Mass (1980).
10. Lieber, Rochelle: Argument linking and compounds. *Linguistic Inquiry*, 14/2 (1983) pp. 251–285.
11. Marantz, A.: On the nature of grammatical relations, Doctoral dissertation, MIT Press, Cambridge, Mass (1981).
12. Meys, W.J.: Compound Adjectives in English and the Ideal Speaker-listener. North – Holland Amsterdam (1975).
13. MORPHY(7WN): <http://www.cogsci.princeton.edu/~wn/man/morphy.7WN.html>.
14. Oshita, Hiroyuki.: A view from Suffixation and A-structure alteration. In: Booig, G., Marle, J. V. (eds): Yearbook of morphology 1994. Kluwer Academic Publishers, Dordrecht (1995) 179–205.
15. Roeper, T. and Siegel, M.E.A.: A Lexical Transformation for Verbal Compounds. *Linguistic Inquiry*, 9 (1978) 199–260.
16. Williams, E.: Argument Structure and Morphology. University of Massachusetts at Amherst (1980).

## Appendix 1

**Table 1.** List of Compound Words (Words found in wordNet are in Italics)

Abundant promise	<i>Bill Collector</i>	Body spray
<i>Alma mater</i>	Bitter gourd	Boiled rice
Assistant master	Black lash	Branch-brown
Auto rickshaw	Black leg	<i>Breathtaking</i>
Bad shot	Black Master	<i>Broad sheet</i>
Benefit show	Black money	Cable Network
Big bull	<i>Blue collar</i>	<i>Cell phone</i>

Central government	Hand-made	Over-ground
<i>Chain Smoker</i>	<i>Handwriting</i>	Overwhelm
Cheque leaf	Hanging cot	Own house
Chief Minister	Heart-breaking	<i>Painstaking</i>
Closed chapter	Help line	Pan fried
Co brother	Hercules task	Panchayat raj
Co-editor	Hidden agenda	Play act
Co sister	Hidden cost	Post-modify
Collective-unconscious	<i>Hollywood</i>	Pound rice
<i>Color-blind</i>	<i>Home page</i>	<i>Power delivery</i>
Community-hall	Hot drinks	Pre press
<i>Compound eye</i>	House-top	Press-button
<i>Compound fraction</i>	<i>Ice cream</i>	Pressroom
<i>Compound fracture</i>	<i>Ill ommened</i>	Proof-read
<i>Compound interest</i>	Inner politics	Provident fund
Compound word	Jackpot	Recycle bin
Contact programme	<i>Kingmaker</i>	Re-do
Cross border	<i>Knock-out</i>	Red army
Cross-country race	Land mark	Red carpet welcome
Door-leveler	Leech gathers	Red street
Draw-sheet	Left-branching	<i>Red tape</i>
<i>Dry-clean</i>	Letter-writing	Regional language
<i>Dry ginger</i>	<i>Long sight</i>	Right-branching
Dying patient	<i>Lower-house</i>	Rented house
E-Magazine	Magic world	Rough note
Earmarked	<i>Mailman</i>	Rough tough
Eco feminism	Mail shot (Advertisement post)	Scorching sun
Eco-linguistics	<i>Manhood</i>	<i>Seafood</i>
Evergreen-hits	<i>Many-sided</i>	Search-box
Ever last	Mega-hit	<i>Search-engine</i>
Fall gay	Meta-analytical	<i>Search term</i>
Fan mail	Mid day	See off
<i>Fat cat</i>	Mid day meal	<i>Self-respect</i>
<i>Fire-Fighter</i>	Mid noon	<i>Short-circuit</i>
Forest cover	Mixed Language	Short sight
Fresh-smell	Morpho-thematic	Sign-post
Gang shooting	New hand	<i>Silver screen</i>
Giant killer	North Indian	Sister concern
Girl-crazy	Off shot	<i>Sky blue</i>
Glass palace	One act play	Slow dry
Golden opportunity	Overwhelm	Snake gourd
Green Rebellion	Open book	Snow-clearing
Green-crazy	Operation flood	Soft drinks
Green signal	Out look	South Indian
Group music		Spider man

<i>Spoon-feed</i>	Teacher aspirant	Visiting time
<i>Stand-by</i>	Tell-tale	<i>Water-resistant</i>
Stress-pattern	<i>Test drive</i>	<i>White collar</i>
Sub-section	Total starvation	White money
Sub urban	Tree cover	White rebellion
Super-hit	<i>Tree-diagram</i>	<i>Wildfire</i>
<i>Superimpose</i>	Twelfth hour	Word formation
<i>Tailor-made</i>	Upper-house	Yellow card
<i>Talk show</i>	<i>Vacuum cleaner</i>	

## Appendix 2

**Table 2.** Meanings

Abundant promise	Excellent, great in number or quantity
Assistant master	Designation of teacher in a public school
Big-bull	A person who is important and highly influential
Black-log	A person who continues to work when his/her fellow workers are on strike: cheater: one who betrays his friend
Black-money	Money earned by illegal means
Body-spray	A spray used for body freshness: A pleasing personality
Boiled rice	A variety of rice where the paddy grains are boiled before making rice out of it
Branch down	Ruin: quite arrogant
Central government	Government of a country having a number of states and its governments: parent
Chief minister	A chief among state ministers: diplomat: a person who takes a decision in the family
Closed chapter	Broken friendship or relationship to a person or an establishment
Co-brother	Cousin brother
Co-sister	Cousin sister
Community-hall	Hall for a group of people of the same race.
Contact-program	A program for helping teaching program in person to students getting education through correspondence course
Cooked story	Gossip
Cross border	Frontier: keeping rivalry: annoyed relationship
Door-leveler	Give exposure to somebody/someone
Draw sheet	Lucky enough
Dying patient	A person who deserves sympathy
Fan mail	Letters from fans to the persons they admire.
Forest cover	Large area of land thickly covered with trees
Fresh smell	Innovative venture
Giant killer	Person who defeats another one stronger than him(Sports): Win over an unusually large person
Glass palace	Illusion

Golden opportunity	Most favorable situation
Green rebellion	Agricultural progress
Green signal	Sanction
Group music	Group song: Unique demand: uniform decision
Hand – made	Not so professional
Hanging cot	Alter position
Heart breaking	Shocking
Hercules task	Most difficult work
Hidden agenda	Mysterious political plans
Hidden cost	Black market price
House top	A parliament section: super
Inner politics	Under current play of an issue
Leech gather	Traditional doctor
Left over	Food remaining at the end of a meal
Left branching	Marxian terminology to denote progressive development in accordance with their theoretical applications
Magic world	Unreal world
Mail shot	An advertisement post
Mid day	Afternoon
Mid noon	Peak at the noon
Mixed language	Mixing two or more different language
New hand	New cover
North Indian	An Indian cultural as well as geographical sphere
One act play	No twist and turn
Open book	Plain and clean: clean image of a person
Over – ground	Unreal
Own house	Permanent place for living
Pound rice	A variety of rice where the paddy is soaked and pound to get a flat variety of rice
Press room	News room
Proof read	To read and correct a piece of written or printed work before publication.: be cleared before actions
Recycle bin	To treat a computer file that has already been used so that it can be used again
Re-do	To do again differently: to place a thing as it is again
Red army	Conspiracy wing
Red carpet welcome	Warm welcome: receive somebody with an open heart
Red street	Anti social place where prostitutes live together and treat customers
Right – branching	Proper development
Rented house	Not permanent place for living
Rough note	Not a fare copy: not justified
Scorching sun	Doing hard work: great difficulty
See off	Farewell
Short sight	A person who does not have future plans
Sister concern	Branch of an institution



Snake guard	A kind of vegetable: very thin person
Snow clearing	Route get cleared
South India	A cultural and geographic sphere in India
Stress pattern	Accent
Sub-urban	Partially urban
Tell- tale	Gossip on cinema actresses:
Total starvation	Too much of suffering
Tree cover	Greenish plants: fresh look
Twelfth hour	Last moment
Upper house	A division in an assembly/parliament
Visiting time	See a person in a proper time
White money	Authorized currency
White rebellion	A good progress in milk products
Word formation	New creation of the word
Yellow card	Punishment: convict.