

# WordNet: What is to be Done?

Patrick Hanks,  
Berlin-Brandenburg Academy of Sciences and Brandeis University

## Good Things about WordNet

Excellent coverage: look up any word and you will find out something about it.  
A rich array of structures showing semantic relations among words.  
Definitions, capturing the gist of each sense.  
Technical terms linked to domain categories.  
Freely available to all on the Internet.  
User-friendly – very easy to use.  
Explicit discussion in accompanying publications of the psycholinguistic theories on which WordNet is founded.

It would be possible to expatiate on these themes, but the purpose of this document is to consider possible improvements, not to sing a paean in praise of WordNet. It is therefore necessary to identify problem areas, but this should not be done in a negative spirit. WordNet is a tremendous achievement. That's why it is worth considering how to make it even better.

## Bad Things about WordNet

Most of the bad things arise from problems in the execution of the idea rather than the idea itself. The devil, as they say, is in the details. From a lexicographical point of view, the main problems are:

Failure to link meaning to use. (It is based on pre-corpus work on the lexicon)  
No indication of relative frequency or importance of senses.  
Poorly motivated sense distinctions.  
Some very weird words – e.g. *minify*, *desquamate*.  
Phraseology is sometimes inconsistent and confusing – esp. in phrasal verbs:

*deal with* is not the same as *deal*.

Non-synonymous synonyms, e.g. *idea* = *content* (?)

Many good modern synonym links are missing.

Hierarchical ontology overdone.

Unnatural made-up examples.

Examples do not apply to all members of the synsets they are attached to.

## Proposals for Improving WordNet

### 1. Make Senses Mutually Exclusive

WordNet has entries for which there is considerable overlap between synsets, e.g.:

2. venture, **hazard**, adventure, stake, jeopardize  
gamble, chance, risk, **hazard**, take chances, adventure, run a risk, take a chance

**anger**, cholera, ire -- (a strong emotion; a feeling that is oriented toward some real or supposed grievance)

**anger**, angriness -- (the state of being angry)

wrath, **anger**, ire, ira -- (belligerence aroused by a real or supposed wrong (personified as one of the deadly sins))

stigmatize, stigmatise, **brand**, denounce, mark -- (to accuse or condemn or openly or formally or brand as disgraceful; "He denounced the government action"; "She was stigmatized by society because she had a child out of wedlock")

post, **brand** -- (mark or expose as infamous; "She was branded a loose woman")

Synsets such as these could be integrated with consequent improvements in clarity and usability.

This would be a considerable task, involving a full-scale re-organization of the sense distinctions of WordNet. If undertaken at all, it would need to be undertaken on the basis of empirical evidence from corpora. This could be done either independently or in association with one or more of the major corpus-based British learner's dictionaries of the past ten years (OALD, LDOCE, Cobuild, CIDE, Macmillan). See 2 below.

### 2. Show Selectional Preferences

WordNet does not at present show selectional preferences. Researchers from many different backgrounds have suggested that it should.

This would be a nontrivial task, as it would be difficult (impossible?) to attach meaningful and contrastive selectional preferences to the existing definition structure of WordNet. The task would therefore need to be done in association with the restructuring mentioned in 1 above.

Selectional preferences could be expressed in the form of contrastive prototypical syntagmatic patterns derived from the CPA project, e.g.:

[[Person]] abandon [[Process]]: *abandon a project*  
[[Person 1]] abandon [[Person 2 | Animate(=Pet)]]: *abandon one's wife and children*  
[[Person]] abandon [[Artefact]]: *they abandoned the car and ran off; an abandoned fridge*  
[[Person(=Military)]] abandon [[Location]]: *after Napoleon abandoned Moscow*  
[[Person]] abandon [SELF] {to [[Activity | Emotion]]}: *abandon oneself to grief*

Each of these five senses of *abandon* belongs in a different synset. They do not map well onto WordNet's existing synset groupings.

*Abandon* is not an exceptional case. Indeed, the sense distinctions for most verbs in WordNet are not well motivated in terms of selectional preferences.

### 3. Delete Weird and Nonexistent Words

*minify* as a synonym for *decrease*, *lessen*. [BNC frequency: 0]

*desquamate* as a synonym for *peel off*. [BNC frequency: 0]

*ira* as a synonym for *anger*. [BNC frequency: 0]

This would be easy to do in association with almost any of the other editorial tasks mentioned in this document. It is, of course, necessary to beware of the failure-to-find fallacy. (The fact that a word is not found or is unfamiliar to a particular person does not mean that it does not exist.) However, the existence of very large corpora now provides the needed evidence to support the trained lexicographer's intuitive judgement that these words are rare, weird, or nonexistent.

### 4. Delete Bad Links

*firewall* as a synonym of *drive* (sense 1)

"he slammed the throttle to the firewall" – WN example.

If authentic (doubtful), this is at best an unusual **exploitation** of the normal sense of *firewall*.

*content* as a synonym for *idea*.

*fan* as a synonym for *winnow*.

Synset: {*abandon*, forsake, desolate, desert}

Is *desolate* a synonym of *abandon*, verb?

*Desolate* is mainly an adjective. Is it also a verb? – YES.

*She was desolated by the sudden loss of her husband.* [BNC freq of verb: 7 in 100 million]

Is it a synonym for *abandon*?

\*The mother desolated her children -- NO.

Synset: {guess, venture, pretend, *hazard*}: *pretend*??

“He *pretended* that she was still there” **does not mean** “he *guessed* that she was still there.”

Re the synset: {cover, *treat*, handle, plow, deal, address}:

*plow*?? *deal*?? “This book plows all aspects of lexicography” **does not mean** “this book covers all aspects of lexicography”.

Clarify the confusion between *love* (verb) and *make love to*. A systematic distinction is made English between the set of words denoting the emotion and the set of words denoting the physical act. This distinction, like many others, is blurred by WordNet.

Cleaning up WordNet’s links would require considerable lexicographic art and judgement. This is not a task for a Postdoc, and would require considerable resources.

## 5. Add Links for Good Matches

(Many good synonyms are missing from WordNet synsets)

E.g. *exasperate* as a synonym or troponym of *annoy*

*discontinue* and *terminate* as synonyms of *abandon* (in the sense “give up (a project)”)

Missing synonyms and hyponyms could be obtained quite easily by consulting recent publications in mainstream English lexicography.

## 6. Add Frequencies or Relative Importance

Any standard English learners’ dictionary would help.

But NOT a dictionary compiled on historical principles such as Merriam Webster or OED.

## 7. Correct the Phraseology

*Abandon oneself to something* is not the same as *abandon*.

*deal* is not the same as *deal with*: “This book deals all aspects of lexicography” is syntactically ill-formed.

*plow*?? Maybe “This book plows *through* all aspects of lexicography”, but not “this book plows all aspects of lex

Synset: {discard, *fling*, toss, toss out, toss away, chuck out, cast aside, *dispose*, throw out, cast out, throw away, cast away, put away}

Here should be *fling out*, *dispose of* -- other particles and prepositions *are* mentioned in this synset.

## 7. Flatten the Procrustean Hierarchies

Especially among abstract nouns...

“An *activity* is a kind of *human activity*” (??)

Is *idea* a hyponym of *concept*? Or is *concept* a hyponym of *idea*? — Surely NEITHER! They are synonyms of each other. The notion of semantic hierarchies is greatly overdone in WordNet.

Another example:

slake, abate, slack -- (make less active or intense)

=> decrease, lessen, minify -- (make smaller; "He decreased his staff")

Is *decrease* a hyponym of *slake*, *abate*, and *slack[en]*? Surely not. They are synonyms.

## 8. Simplify the Remaining Hierarchies

Hierarchies such as “*canary=>bird => vertebrate => animate thing => physical object => entity*” are realistic and verifiable.

But terms such as *chordate* and *craniate* belong in a specialist zoological domain-specific WordNet (if anywhere). They are not part of general English. They merely confuse the hierarchy. Even the good hierarchies need to be simplified.

## 9. Replace Examples

Replace examples with corpus-derived examples selected to show typical usage.

Examples should be chosen to exemplify the typical usage of each word in the relevant sense; the examples should be attached to words, not to synsets.

## Halliday’s Suggestion

“A thesaurus of English based on formal criteria, giving collocationally defined lexical sets with citations to indicate the defining environments, would be a valuable complement to Roget’s brilliant work of intuitive semantic classification.” —M. A. K. Halliday, 1966

This suggestion has never been implemented. Indeed, up until about 1995, it *could not* have been implemented – there was simply not enough data in machine-readable form to provide a reliable basis for such a project.

## Conclusion

Some of the above suggestions could be implemented cheaply and efficiently within the existing WordNet framework, given the availability of corpus-trained lexicographers. Such lexicographers are plentiful in England, thanks to the activity of publishers of learners’ dictionaries. There are none in America.

It is an open question whether it would be better to implement some or all of these proposals as a revision of the existing WordNet or as an independent, parallel project.