Putting Semantics into WordNet's "Morphosemantic" Links

Christiane Fellbaum
Anne Osherson

Princeton University
Princeton University
Princeton, NJ 08540 (USA)
{fellbaum,aosher}@princeton.edu

Peter E. Clark

Boeing Company
Seattle, WA (USA)
peter.e.clark@boeing.com

Abstract
To add to WordNet's contents, and specifically to aid automatic reasoning with WordNet, we classify and label the current relations among derivationally and semantically related noun-verb pairs. Manual inspection of thousands of pairs shows that the form-meaning mappings of affixes are not in a one-to-one relation and far less regularity than expected. We determine a set of semantic relations found across a number of morphologically defined noun-verb pair classes.

1. Introduction

Natural Language Processing applications such as Information Retrieval and Machine Translation rely critically on lexical resources. Such resources often do not include words that are morphologically derived from base forms on the assumption that morphology is regular and that affixes carry unambiguous information both about the part of speech and the meaning of the derived word. (Many traditional paper dictionaries include derivations, but list them as run-ons without any information on their meaning.) Dorr and Habash (2003), recognizing the importance of morphology-based lexical nests for NLP, created "CatVar," a large-scale database of categorial variations of English lexemes. CatVar relates lexemes belonging to different syntactic categories (part of speech) and sharing a stem, such as *hunger* (n.), *hunger* (v.) and *hungry* (adj.). CatVar is a valuable resource containing some 100,000 unique English word forms; however, no information is given on the words' meanings.

A complementary resource to CatVar is WordNet (Miller 1995, Fellbaum 1998), which focuses on semantics and expresses the meanings of some 155,000 English words in terms of semantic relations such as synonymy, antonymy, hyponymy, and meronymy. Most of WordNet's relations are paradigmatic, i.e., they link words belonging to the same syntactic category.

Miller and Fellbaum (2003) describe the addition of "morphosemantic links" to WordNet, which connect words (synset members) that are similar in meaning and where one word is derived from the other by means of a morphological affix. For example, the verb *direct* (defined in WordNet as "guide the actors in plays and films") is linked to the noun *director* (glossed as "someone who supervises the actors and directs the action in the production of a show"). Another link was created

for the verb-noun pair *direct/director*, meaning "be in charge of" and "someone who controls resources and expenditures," respectively. Most of these links connect words from different classes (noun-verb, noun-adjective, verb-adjective), though there are also noun-noun pairs like *gang-gangster*.

English derivationally morphology is highly regular and productive, and the addition of a given affix to a base form produces a new word whose meaning differs from that of the base word in a predictable way. For example, adding the affix *-en* to many adjectives yields a verb that denotes a change event, where an entity acquires the property denoted by the adjective:

(1) *red-redden*
dark-darken
sad-sadden
fat-fatten
etc.

English has many such affixes and associated meaning-change rules (Marchand 1969).

When the morphosemantic links were added to WordNet, their semantic nature was not made explicit, as it was assumed --- following conventional wisdom --- that the meanings of the affixes are highly regular and that there is a one-to-one mapping between the affix forms and their meanings.

2. Labeling morphosemantic links

Systems for robust textual inference make extensive use of WordNet as an informal source of knowledge (e.g., MacCartney et al. 2006). We are currently working to transform WordNet into a Knowledge Base that better supports such reasoning and inferencing (Clark et al. 2007). WordNet's morphosemantic arcs could be

valuable if the semantics of the links were spelled out. For example, while humans can easily infer that the (b,c) statements are entailed by the (a) statements, automatic systems are having trouble with this task:

(2) a. The Zoopraxiscope was invented by Mulbridge.
   b. The inventor of the Zoopraxiscope is Mulbridge.
   c. The Zoopraxiscope is an invention by Mulbridge.

(3) a. Shareholders criticized Dodge, which produces ProHeart devices.
   b. ProHeart is a product.
   c. Dodge is the producer of ProHeart.

Currently, WordNet is able to link *invented* with *inventor* and *invention* in (2) as well as *produce, product,* and *producer* in (3). But it does not tell us that *inventor* and *producer* are the Agents of the events denoted by the verbs *invent* and *produce*, respectively, or that ProHeart and the Zooprxiscope are products. We considered examples of statements and possible inferences and noticed that in many cases, spelling out the relation between the nouns and verbs in the two sentences would facilitate the evaluation of the entailments.

We began our efforts to add explicit meaning to WordNet's morphosemantic links with those noun-verb pairs where the nouns are derived from the verbs by *-er* and *-or* suffixation (*invent-inventor, produce-producer, build-builder*). We assumed that, with rare exceptions, the nouns denote the Agents of the event referred to by the verb, as expressed by the phrases in (4):

(4) an inventor invents
   a producer produces
   a builder builds
   etc.

We will refer to pattern in (4) as the *Agentive* pattern.
We automatically extracted some 4,000 morphosemantic noun-verb pairs related via *-er* affixation from WordNet. All were manually inspected. False hits were discarded and those pairs that did not conform to the Agentive pattern were placed into a separate file.
We repeated this process with noun-verb pairs where the deverbal nouns ended in *–al* (*reverse-reversal*), *-ment* (*amaze-amazement*) and *–ion* (*infuse-infusion*). In addition, we extracted meaningfully related pairs where the verb was derived from a noun via *–ize* affixation (*alphabet-alphabetize*) or from an adjective by means of the addition of *–ify* (*beauty-beautify*).

3. How regular are morphosemantics?
For the *-er* derivations, expected to find few "exceptions" to the Agentive pattern. But surprisingly, only two thirds of the pairs could be classified as in (4). For the remaining pairs, different semantic relations were formulated and the pairs were labeled accordingly (words in parentheses identify the intended senses of polysemous nouns).

(5) Instrument
   rule- ruler (measuring stick)
   shred- shredder
   aspirate- aspirator

Instruments are distinct from Agents: anInstrument does not act alone but implies an Agent who controls it, usually with intention. Both can co-occur as arguments of the verb: *John ruled straight lines with his ruler*. Wojcik (1976) distinguished "enabling" from "facilitating" instruments: the former but not the latter may appear in subject position (*the shredder shredded the paper* vs. **the ruler ruled the lines*).

(6) Inanimate Agent/Cause
   block-blocker (drug)
   whiten-whitener
   sense-sensor
   soften-softener

Inanimate Agents or Causers are often substances and, unlike Instruments, can act without the direct control of a human Agent*: the softener softened the clothes* vs. **the ruler ruled the lines.*

(7) Body part
   adduct-adductor

Like Inanimate Agents/Causers, Body Parts take the place of Agents in events*.* The *adductor* muscle *adducts*, etc. Unlike with Agents, volition and intention are not implied.

(8) Purpose/Function
   line-liner (coating/layer)
   read- reader (book)
   train- trainer (shoes)

The verbs express the intended purpose of function of the nouns: *trainers* are for *training*, *readers* are for *reading*, etc.

(9) Vehicle
   commute-commuter (train)
   cruise-cruiser (boat)

The event denoted by the verb takes place in the vehicle that the noun refers to.

(10) Location
   plant-planter(pot)
   sleep-sleeper (sofabed)
   hang-hanger
   lock- locker

The noun denotes the Location of the event or state expressed by the verb: one *plants* something in a *planter*, *sleeps* in a *sleeper*, *hangs* something on a *hange*r, *locks* something in a *locker*, etc.

(11) Undergoer/Patient

break-breaker (wave)
broil-broiler (chicken)
steam-steamer (clam)
loan- loaner

The noun denotes the Undergoer of the event: the wave (*breaker*) *breaks*, the chicken (*broiler*) *broils*, a *loaner* is an item that has is *loaned*, etc.
The presence of an Undergoer blocks an Agent that instigates the event (**The breaker was broken by the storm*).

(12) Event
dine-dinner
pelt-pelter

A *dining* event is a *dinner*, *pelting* rain is a *pelter*, etc.

(13) Result or Cause
groan-groaner (bad joke)
err-error

The event produces the entity denoted by the noun (*erring* results in an *error*), or, conversely, the noun causes the event (*groaner-groan*). We do not distinguish between eventive and product readings of the Result, though WordNet often draws this regular distinction among polysemous nouns.

Table 1 shows the number of pairs for each semantic class among all the *–er*-related pairs that were extracted from WordNet 3.0.

Agent	2,584
Instrument	482
Inanimate agent/Cause	302
Event	224
Result	97
Undergoer	62
Body part	49
Purpose	57
Vehicle	36
Location	36

Table 1: Distribution of *–er* verb-noun pair relations

4. Relations

The same relations turn up in the classification of other morphosemantically linked noun-verb pairs in WordNet that we have inspected.. The *-er/-or* pairs exhibit a particularly wide spectrum of relations; some other classes we have looked at so far include pairs that can be classified with a subset of categories only and the semantics of the affix seems to be somewhat more regular. Some affix-based classes suggest categories not found in the *–er* class. We have not yet settled on an inventory of categories that covers all morphological classes, but it seems that the number will be between fifteen and twenty.

Of course, the categories could be distinguished more finely (e.g., Result and Cause could be separated) or collapsed (Body Part and Inanimate Cause), and our classification is somewhat subjective, though we tried to motivate it syntactically wherever possible. For example, Purpose is different from Instrument, in that it does not allow a PP headed by *with*, a characteristic of Instruments. Similarly, Vehicles are expressed in a PP headed by a spatial proposition (i*n/on,* etc.) rather than *with*, though Vehicles could be considered a subclass of Locations. Instruments presuppose an Agent who acts with intent and volition and Instrument and Agent must therefore be distinguished. Locations are where the event denoted by the verb takes place and where the Agent or Patient is located.

The labels we assigned refer to well-known well known semantic categories and have been studied or applied in different contexts. The Cases proposed by Fillmore (1968) and the FrameElements of FrameNet (Ruppenhofer et al. 2006) also refer to Agents, Undergoers/Patients, Instruments, etc.

5. Polysemy

We saw that the *–er* suffix is polysemous; although the Agentive pattern is the default reading for deverbal nouns, there are several other patterns. For example, the verb *broil* is paired with three distinct senses of the noun *broiler* (Agent, Location, and Undergoer). Similarly, we have two pairs *plant-planter* (Agent and Location), just as in the case of *dine-diner*. Some highly polysemous noun-verb pairs enter into an even larger number of relations involving different senses of both the nouns and the verbs (*run-runner, roll-roller*).

For virtually all cases we examined, the default agentive reading of the noun is always possible, though it is not always lexicalized (and does not have an entry in WordNet or other lexicons). Speakers easily generate and process ad-hoc nouns like *planter* (gardener), but only in its (non-default) location reading ("pot") is the noun part of the lexicon, as its meaning cannot be guessed from its structure.

Examining other morphological patterns, we found that polysemy of affixes is widespread. Thus, nouns derived from verbs by *–ion* suffixation exhibit regular polysemy between event and result readings (*the exam lasted two hours/the exam was lying on his desk*, Pustejovsky 1995).

We also fnd one-to-many mappings for semantic patterns and affixes: a semantic category can be expressed by means of several distinct affixes, though there seems to be a default semantics associated with a given affix. Thus, while many –er nouns denote Events, event nouns are regularly derived from verbs via *–al* suffixation (*bomb-bombardment*, *punish-punishment*, etc.)

Patterns are partly predictable from the thematic structure of the verb. Thus, nouns derived from unergative verbs (intransitives whose subject is an Agent) are Agents, and the pattern is productive:

(14) *runner, dancer, singer, speaker, sleeper,*

Nouns derived from unaccusative verbs (intransitives whose subject is a Patient/Undergoer) are Patients:

(15) *breaker* (wave), *streamer* (banner)

This pattern is far from productive:

(16) *faller, ?arriver, ?leaver,

Many verbs have both transitive (causative) and intransitive readings (cf. Levin 1993):

(17)a. The cook roasted the chicken
     b. The chicken was roasting

For many such verbs, there are two corresponding readings of the derived nouns: both the *host* in (17a) and the chicken in the (17b) can be referred to as a *roaster*. Other examples of Agent and Patient nouns derived from the transitive and intransitive readings of verbs are *(best)seller, (fast) developer, broiler*. But the pattern is not productive, as nouns like *cracker, stopper*, and *freezer* show.

6. Related Work

Clark and Clark (1979) examine the large number of English noun-verb pairs related by zero-affix morphology, i.e., homographic pairs of semantically related verbs and nouns (*roof, lunch, Xerox*, etc.) Clark and Clark note that this pattern of deriving verbs from nouns in English is productive and speakers readily interpret the meaning of novel verbs based on their knowledge of the nouns, even though the relations among verbs and nouns do not permit a unified semantic description. Clark and Clark distinguish a large number of semantic noun classes that have spawned derived verbs, including Agent, Location, Instrument, Body Part, Meals, Elements, and Proper Names. Clark and Clark conclude that the meanings of the verbs depend on the time, place and circumstances of their use and are somewhat conventionalized.

In the context of the EuroWordNet project (Vossen 1998), Peters (n. d.) manually established noun-verb and adjective-verb pairs that were both morphologically and semantically related. (EuroWordNet was based on WordNet version 1.5, which lacked the morphosemantic relations added in version 2.0). Of the relations that Peters considered, the following match the ones we identified: Agent, Instrument, Location, Patient, Cause. But Peters's methodology differed from ours. While we proceeded from the previously classified morphosemantic links and assumed a default semantic relation for pairs with a given affix, Peters selected pairs of word forms that were both morphologically related and where at least one member had only a single sense in WordNet. These were then manually disambiguated and semantically classified, regardless of regular morphosemantic patterns.

Pala and Hlaváčková (2007) automatically enrich Czech WordNet with "derivational nests," groups of morphologically and semantically related words. Czech morphology is very rich and relatively regular, allowing Pala and Hlaváčková to construct a tool that generates new word forms derived from stems by adding affixes associated with specific semantics. The noun-verb relations that the Czech team identified overlap with those reported on here for English. (Czech WordNet, and the morphological-semantic links make by Pala and Hlaváčková , are linked to the Princeton WordNet.)

7. Conclusions and Future Work

Manual inspection of morphologically related verb-noun pair classes shows that, contrary to what is commonly assumed, there is no one-to-one mapping of affixes and meaning. Rather, affixes can be highly polysemous, and a given affix can be associated with several meanings, though there seems to a default reading for each affix (such as the Agentive pattern for *–er* deverbal nouns).When a default reading is not applicable, speakers can compute the reading from context, as suggested by Clark and Clark. Homographic derivations, like *sleeper* ("sleeping person" vs. "sleeping car"), like all cases of polysemy, require separate entries for each sense both in speakers' mental lexicon as well as in computational lexicons.

Conversely, a given semantic pattern can be associated with more than one affix. For example, both *–al* and *–ment* can denote deverbal events, as in *disbursal, disbursement* and *committal, commitment*, making the two derived nouns synonyms in each case.

Our work so far, and the independent efforts by Peters and Pala and Hlaváčková suggest that the meanings of affixes can be classified into a finite, relatively small number of semantic categories. It is important to note that the inventory of relations we presented here is somewhat arbitrary; one could certainly propose a more fine-grained or a more coarse-grained one. We expect to encode additional relations as we consider other types of morphosemantic pairs, though we anticipate a fairly small number of relations, most likely a subset of those discussed by Clark and Clark (1979). We started to explore the encoding of these relations crosslinguistically, focusing on several Bantu languages (Bosch, Fellbaum and Pala, in prep.) Encoding the semantics of the relations in WordNet will make it a more useful tool for automated reasoning and inferencing.

References
Bosch, S., Fellbaum, C., and Pala, K. (in prep.) Derivational Morphology in Bantu, Czech, and English. To appear in the Proceedings of ALASA/HLT track.

Clark, E. and Clark, H. (1979). *When nouns surface as verbs*. Language *55,* 767-811.

Clark, P., Harrison, P., Thompson, J., Murray, W., Hobbs, J., and Fellbaum, C. (2007). On the Role of Lexical and World Knowledge in RTE3. *ACL-PASCAL Workshop on Textual Entailment and Paraphrases,* June 2007, Prague, CZ.

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

Fillmore, C**.** (1968). The Case for Case. In: Bach, E., and R. Harms (Eds.) *Universals in linguistic theory*. NY: Holt.

Habash, N. and B. Dorr (2003). A Categorial Variation Database for English. *Proceedings of the North American Association for Computational Linguistics*, Edmonton, Canada, pp. 96-102, 2003.

Levin, B. (1993). *English Verb Classes and Alternations.* Chicago, IL: University of Chicago Press.

Marchand, H. (1969). The categories and types of present-day English word formation. Munich: Beck.

MacCartney, B., Grenager, T., de Marneffe, M.-C., Cer, D., and Manning, C.D. (2006). Learning to recognize features of valid textual entailments. Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006), pp. 41-48.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*. 38.11:39-41

Miller, G. A. and Fellbaum, C. (2003). Morphosemantic links in WordNet. *Traitement automatique de langue,* 44.2:69-80.

Pala, K. and Hlaváčková, D. (2007). Derivational Relations in Czech WordNet. In: Proceedings fo the Workshop on Balto-Slavonic, ACL, Prague, 75-81.

Peters, W**.** (n.d.) The English wordnet, EWN Deliverable D032D033. University of Sheffield, England.

Pustejovsky, J. (1995). The Generative Lexicon. Cambridge, MA: MIT Press.

Ruppenhofer, Baker, C. and Fillmore, C.J. (2002): The FrameNet Database and Software Tools. In Braasch, A. and C. Povlsen (eds.), *Proceedings of the Tenth Euralex International Congress.* Copenhagen, Denmark. Vol. I: 371-375.

Vossen, P. (1998, Ed.). *EuroWordNet.* Dordrecht: Kluwer.

Woijcik, R. (1976). Where do Instrumental NPs come From? In: M. Shibatani (Ed.), *Syntax and Semantics* 6: 165-180.