

WordNet and FrameNet as Complementary Resources for Annotation

Collin F. Baker

International Computer Science Institute
1947 Center St., Berkeley, California 94704
collinb@icsi.berkeley.edu

Christiane Fellbaum

Princeton University
Princeton, NJ 08540-5233
fellbaum@princeton.edu

Abstract

WordNet and FrameNet are widely used lexical resources, but they are very different from each other and are often used in completely different ways in NLP. In a case study in which a short passage is annotated in both frameworks, we show how the synsets and definitions of WordNet and the syntagmatic information from FrameNet can complement each other, forming a more complete representation of the lexical semantic of a text than either could alone. Close comparisons between them also suggest ways in which they can be brought into alignment.

1 Background and motivation

FrameNet and WordNet are two lexical databases that are widely used for NLP, often in conjunction. Because of their complementary designs they are obvious candidates for alignment, and an exploratory research project within the larger context of the semantic annotation of the the American national Corpus is currently underway. We give specific illustrative examples of annotations against both resources, highlighting their different contributions towards a rich semantic analysis.

WordNet (WN):¹ (Fellbaum, 1998), is a large electronic lexical database of English. Originally conceived as a full-scale model of human semantic organization, it was quickly embraced by the Natural Language Processing (NLP) community, a development that guided its subsequent growth and design. WordNet has become the lexical database of choice for NLP and has been incorporated into other language tools, including VerbNet (Kipper et al., 2000) and OntoNotes (Hovy et al., 2006). Numerous on-line dictionaries, including Google's "define" function, rely significantly on WordNet.

WordNet's coverage is sometimes criticized as being too fine-grained for automatic processing, though its inventory is not larger than that of a standard collegiate dictionary. But the present limitation of automatic WSD cannot be entirely blamed on existing systems; for example, Fellbaum and Grabowski (1997)

have shown that humans, too, have difficulties identifying context-appropriate dictionary senses. One answer is clearly that meanings do not exist outside contexts. Furthermore, although WN does contain "sentence frames" such as "Somebody —s something" for a transitive verb with a human agent, it provides little syntagmatic information, except for what can be gleaned from the example sentences. WordNet's great strength is its extensive coverage, with more than 117,000 synonym sets (**synsets**), each with a definition and relations to other synsets covering almost all the general vocabulary of English.

FrameNet (FN):² (Fontenelle, 2003) is a lexical resource organized not around words *per se*, but **semantic frames** (Fillmore, 1976): characterizations of events, relations, and states which are the conceptual basis for understanding groups of word senses, called **lexical units (LUs)**. Frames are distinguished by the set of roles involved, known as **frame elements (FEs)**. Much of the information in the FrameNet lexicon is derived by annotating corpus sentences; for each LU, groups of sentences are extracted from a corpus, sentences which collectively exemplify all of the lexically relevant syntactic patterns in which the LU occurs. A few examples of each pattern are annotated; annotators not only mark the target word which **evokes** the frame in the mind of the hearer, but also mark those phrases which are syntactically related to the target word and express its frame elements. FrameNet is much smaller than WordNet, covering roughly 11,000 LUs, but contains very rich syntagmatic information about the combinatorial possibilities of each LU.

Given these two lexical resources with different strengths, it seems clear that combining WN and FN annotation will produce a more complete semantic representation of the meaning of a text than either could alone. What follows is intended as an example of how they can usefully be combined.

2 Case Study: Aegean History

The text chosen for this study is a paragraph from the American National Corpus³ (Ide et al., 2002), from the Berlitz travel guide to Greece, discussing the history of

¹<http://wordnet.princeton.edu>

²<http://framenet.icsi.berkeley.edu>

³<http://www.americannationalcorpus.org>

Greece, specifically the Aegean islands after the fall of Byzantium to the Crusaders. Although brief, its three sentences provide ample material to demonstrate some of the subtlety of both WN and FN annotation:

(1) *While Byzantine land was being divided, there was no one in control of the seas, so pirates raided towns on many of the islands.* (2) *To counter this, the populations moved from their homes on the coast and built settlements inland, out of sight of the raiding parties.* (3) *This created a pattern seen today throughout the Aegean of a small port (skala) which serves an inland settlement or chora, making it easier to protect the island from attack.*

Below, we present three tables containing the annotation of both the WordNet synsets for each open class (content) word in the text⁴ and the FrameNet frames and the fillers of the frame elements in each sentence. We also provide brief notes on some interesting features of the semantics of each sentence.

2.1 Discussion of Sentence 1, shown in Table 1 on page 4 :

(2) Information about what the land was separated into is not given in the sentence nor clear from the context, so the PARTS FE has been annotated as “indefinite null instantiated” (INI). Clearly this is an intentional action, but because the verb is passive, the agent can be (and is) omitted, so the AGENT FE is marked as “constructionally null instantiated” (CNI).⁵

(4) In addition to FEs and their phrase types and grammatical functions, FrameNet annotates a limited set of syntactic facts: here, *in* is annotated as at “support preposition”, allowing *control* to function as an adjectival, and *was* as a copula, allowing *no one* to fill the External syntactic position of *in control*.

(5) Since FN is based on semantic frames, annotation of nouns is largely limited to those which express events (e.g. *destruction*), relations (*brother*), or states (*height*). For the most part, nouns denoting artifacts and natural kinds evoke relatively uninteresting frames, and hence relatively few of them have been included in FN. However, there are three such instances in this sentence, *seas*, *islands* (9), and *towns* (12); In all three cases, the frame-evoking noun also **denotes** the filler of the FE LOCALE.

(6) At the top level of organization, *so* evokes the Causation frame. Actually, it is misleading to simply annotate *control of the seas* in the frames *Be_in_control* and *Natural_features*; here, we regard *seas* as metonymic for “ship traffic on the seas”, but neither the FN annotation nor the WN definition indicates this.

(7) The noun *pirates* evokes the very rich frame of

⁴Note that for reasons of space, many WN examples have been omitted.

⁵In fact, the previous sentence describes the sack of Constantinople by the Crusaders, so they can be inferred to be the dividers of the lands, as well.

Piracy, and also denotes the filler of the FE PERPETRATOR, but that is the only FE filled in in that frame. Instead, *pirates* actually fills the ASSAILANT FE of the Attack frame, (8); the main idea is about the raids, not the piratical acts on the seas that the same people have a habit of committing. Note that the WN definition takes the view that raiding coastal towns is a typical part of piracy.

(10) *Political_locales* roughly corresponds to “Geopolitical entity” in named entity recognition.

Despite the relatively fine level of detail of the annotations, there are still many important semantic features of the sentence not represented in FrameNet or WordNet. For example, there is no treatment of negation *cum* quantification, no representation of the fact that *there was no one in control* should mean that *Be_in_control* is not happening.

2.2 Discussion of Sentence 2, shown in Table 2 on page 5:

The two highest level predicates in this sentence are *moved* (2) and *built* (6), in the frames Motion and Building respectively; since they are conjoined, the phrase *to counter this* fills the FE PURPOSE in both frames.⁶ In (2) the GOAL FE of the Motion is marked as definite null instantiation (DNI), because, although it is not expressed in the VP headed by *moved*, it is recoverable from context (i.e. the second VP).

(4) Note that FN puts this sense of *home* in the Buildings frame⁷, but WN has a less specific definition. (6) *Coast* is a *Relational_natural_feature* because it is defined in relation to another natural feature; a coast has to be the coast **of** some land mass, although here the land mass is DNI. (9) *Inland* both evokes a *Locative_relation* and denotes the GROUND FE. (10) FN and WN agree on a sense of *sight* denoting the range of vision. (11) WN’s example sentence for *raid* is precisely about pirates.

2.3 Discussion of Sentence 3 shown in Table 3 on page 5:

(2) The concept of “pattern” is very slippery—the arrangement of port and inland settlement is both spatial and temporal in terms of building practices over centuries. (3) This sense of *see* can refer to the area in which something is seen, the time, or the conditions under which it can be seen; these are subsumed by the FE STATE. (4) *Today* expresses a *Temporal_collocation* and denotes the LANDMARK. (Repetitions of the words *settlement* and *island* have been omitted.) The interrelation among (7), (10), (11) and (12) is rather complex: the arrangement in which the port serves the settlement has the making easier as a result. The arrangement is also the CAUSE FE of *making*. *Easier* in the *Difficulty* frame requires an EX-

⁶This is a **peripheral** FE, common to all frames which inherit from the *Intentionally_act* frame.

⁷Not to be confused with the Building frame, in (7).

PERIENCER FE which is not specified here (thus INI) and an ACTIVITY FE, *to protect*. The FE PROTECTION (which can be a person, a thing, or an activity) is marked CNI, because it is the external argument of the infinitive.

3 Towards an alignment of WordNet and FrameNet

We hope these examples have shown that finding related WN and FN senses can contribute to text understanding. Fellbaum and Baker (2008) discuss the respective strengths and weaknesses of WN and FN as well as their complementary advantages that could be fruitfully exploited aligning the two resources. Work of this type is actually underway; researchers are semi-automatically annotating selected lemmas in the American National Corpus with both FN frames and WN senses. The lemmas are chosen so as to reflect the part of speech distribution in text and to represent a spectrum of frequency and polysemy. A preliminary group of instances are manually tagged by trained annotators, and then the teams working on WN and FN annotation discuss and resolve discrepancies among the taggers before the remaining tokens are annotated.

Three cases sum up the annotation and alignment process:

(1) In the very unlikely case that a synset and a frame contain exactly the same set of lexemes, their correspondence is simply recorded.

(2) In the more common case in which all the words in a synset are a subset of those in the frame, or all the words in a frame are a subset of those in the synset, this fact is also recorded.

(3) In case two synsets are subsets of the LUs of one frame, we will record this and note that it as a possible candidate for collapsing the synsets, respectively.

FN and WN are two comprehensive but complementary lexical resources. Both WN's paradigmatic and FN's syntagmatic approach to lexical semantics are needed for a rich representation of word meaning in context. We have demonstrated how text can be annotated against both resources to provide the foundation for deep language understanding and, as an important by-product, help to align the word senses of these widely-used resources. Of course, these examples were manually annotated, but automatic systems for word-sense disambiguation (largely based on WordNet) and FrameNet role labeling (Johansson and Nugues, 2007; Coppola et al., 2008) are improving rapidly. The project just described is intended to provide more gold-standard annotation (both WN and FN) to help train automatic systems for both WN and FN annotation, which are clearly related tasks e.g. (Pradhan et al., 2007; Erk, 2005).

Acknowledgment

We gratefully acknowledge support from the National Science Foundation (#IIS-0705199) for the work reported here.

References

- Bonaventura Coppola, Alessandro Moschitti, Sara Tonelli, and Giuseppe Riccardi. 2008. Automatic framenet-based annotation of conversational speech. In *Proceedings of IEEE-SLT 2008*, pages 73–76, Goa, India, December.
- Katrin Erk. 2005. Frame assignment as word sense disambiguation. In *Proceedings of IWCS 6*, Tilburg.
- Christiane Fellbaum and Collin F. Baker. 2008. Can WordNet and FrameNet be made “interoperable”? In Jonathan Webster, Nancy Ide, and Alex Chengyu Fang, editors, *Proceedings of The First International Conference on Global Interoperability for Language Resources*, pages 67–74, Hong Kong. City University.
- Christiane Fellbaum and J. Grabowski. 1997. Analysis of a hand-tagging task. In *Proceedings of the ACL/Siglex workshop*. Association for Computational Linguistics.
- Christiane Fellbaum, editor. 1998. *WordNet. An electronic lexical database*. MIT Press, Cambridge/Mass.
- Charles J. Fillmore. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280:20–32.
- Thierry Fontenelle, editor. 2003. *International Journal of Lexicography—Special Issue on FrameNet*, volume 16. Oxford University Press.
- Eduard H. Hovy, Mitch Marcus, Martha Palmer, Sameer Pradhan, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: The 90% solution. In *Proceedings of HLT-NAACL 2006*, New York.
- Nancy Ide, Randi Reppen, and Keith Suderman. 2002. The American National Corpus: More than the web can provide. In *Proceedings of the Third Language Resources and Evaluation Conference (LREC)*, pages 839–44, Las Palmas, Canary Islands, Spain.
- Richard Johansson and Pierre Nugues. 2007. LTH: Semantic structure extraction using nonprojective dependency trees. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 227–230, Prague, Czech Republic, June. Association for Computational Linguistics.
- Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *Seventeenth National Conference on Artificial Intelligence*, Austin, TX. AAAI-2000.

Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. Semeval-2007 task-17: English lexical sample, srl and all words. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 87–92, Prague, Czech Republic, June. Association for Computational Linguistics.

1. Frame: Political_locales: [CONTAINER_POSSESSOR Byzantine] [LOCALE LAND] WN: (adj) <i>Byzantine</i> (of or relating to or characteristic of the Byzantine Empire or the ancient city of Byzantium) (n) domain, demesne, <i>land</i> (territory over which rule or control is exercised) “his domain extended into Europe”; “he made it the law of the land”
2. Frame: Separating: [WHOLE Byzantine land] was being DIVIDED [AGENT CNI] [PARTS INI] WN: (v) <i>divide</i> , split, split up, separate, dissever, carve up (separate into parts or portions) “divide the cake into three equal parts”; “The British carved up the Ottoman Empire after World War I”
3. Frame: Existence: [TIME While Byzantine land was being divided], THERE WAS [ENTITY no one in control of the seas]
4. Frame: Be_in_control: there [was COPULA] [CONTROLLING_ENTITY no one] [in SUPPORT] CONTROL [DEPENDENT_ENTITY of the seas] WN: (n) <i>control</i> (power to direct or determine) “under control”
5. Frame: Natural_features: [LOCALE SEAS] WN: (n) <i>sea</i> (a division of an ocean or a large body of salt water partially enclosed by land)
6. Frame: Causation: [CAUSE While Byzantine land was being divided, there was no one in control of the seas], SO [EFFECT pirates raided towns on many of the islands]
7. Frame: Piracy: [PERPETRATOR PIRATES] WN: (n) <i>pirate</i> , buccaneer, sea robber, sea rover (someone who robs at sea or plunders the land from the sea without having a commission from any sovereign nation)
8. Frame: Attack: [ASSAILANT pirates] RAIDED [VICTIM towns on many of the islands] WN: (v) foray into, <i>raid</i> (enter someone else’s territory and take spoils) “The pirates raided the coastal villages regularly”
9. Frame: Political_locales: [LOCALE TOWNS] [RELATIVE_LOCATION on many of the islands]. WN: (n) town (an urban area with a fixed boundary that is smaller than a city)
10. Frame: Locative_relation: [FIGURE towns] ON [GROUND many of the islands]
11. Frame: Quantity: [QUANTITY MANY] [INDIVIDUALS of the islands]
12. Frame: Natural_features: [LOCALE ISLANDS] WN: (n) <i>island</i> (a land mass (smaller than a continent) that is surrounded by water)

Table 1: FN/WN Annotation of sentence 1

1. Frame: Thwarting: To COUNTER [ACTION this], [PREVENTING_CAUSE the populations moved ... raiding parties] WN:(v) anticipate, foresee, forestall, <i>counter</i> (act in advance of; deal with ahead of time)
2. Frame: Aggregate: [AGGREGATE POPULATIONS] WN: (n) <i>population</i> (the people who inhabit a territory or state) “the population seemed to be well fed and clothed”
3. Frame: Motion: [PURPOSE To counter this], [THEME the populations] MOVED [SOURCE from their homes on the coast] [GOAL DNI] WN: (v) <i>move</i> (change residence, affiliation, or place of employment)
4. Frame: Buildings: [BUILDING HOMES] [PLACE on the coast] WN: (n) <i>home</i> , place (where you live at a particular time) “deliver the package to my home”
5. Frame: Locative_relation: [FIGURE their homes] ON [GROUND the coast]
6. Frame: Relational_natural_features: [FOCAL_FEATURE COAST] [RELATIVE_LOCATION DNI] WN: (n) seashore, <i>coast</i> , seacoast, sea-coast (the shore of a sea or ocean)
7. Frame: Building: [PURPOSE To counter this], [AGENT the populations] ... BUILT [CREATED_ENTITY settlements] [PLACE inland], [PLACE out of sight of the raiding parties]. WN: (v) construct, <i>build</i> , make (make by combining materials and parts)
8. Frame: Locale_by_use: [LOCALE SETTLEMENTS] WN: (n) village, small town, <i>settlement</i> (a community of people smaller than a town)
9. Frame: Locative_relation: built [FIGURE settlements] [GROUND INLAND] WN: (adv) inland (towards or into the interior of a region) “the town is five miles inland”
10. Frame: Range: ...out of [DISTANCE SIGHT] [PARTICIPANT of the raiding parties] WN: (n) <i>sight</i> , ken (the range of vision) “out of sight of land”
11. Frame: Attack: RAIDING [ASSAILANT parties] WN: (v) foray into, <i>raid</i> (enter someone else’s territory and take spoils) “The pirates raided the coastal villages regularly”
12. Frame: Aggregate: [AGGREGATEPROPERTY raiding] [AGGREGATE PARTIES] WN: (n) <i>party</i> , company (a band of people associated temporarily in some activity) “they organized a party to search for food”

Table 2: FN/WN Annotation of sentence 2

1. Frame: Creating: [CAUSE This] CREATED [CREATED_ENTITY a pattern seen today ... from attack]. WN: (v) create (bring into existence) “He created a new movement in painting”
2. Frame: Pattern: PATTERN [DESCRIPTOR seen today throughout the Aegean] [ENTITIES of a small port (skala) which serves an inland settlement or chora] WN: (n) practice, pattern (a customary way of operation or behavior) “they changed their dietary pattern”
3. Frame: Perception_experience: [PHENOMENON a pattern] SEEN [TIME today] [STATE throughout the Aegean] [PHENOMENON of a small port ... from attack]. [PERCEIVER_PASSIVE CNI] WN: (v) witness, find, see (perceive or be contemporaneous with) “You’ll see a lot of cheating in this school”
4. Frame: Temporal_collocation: [TRAJECTOR_EVENT a pattern seen] [LANDMARK_EVENT TODAY] [TRAJECTOR_EVENT throughout the Aegean... attack] WN: (n) today (the present time or age) “the world of today” (n) Aegean, Aegean Sea (an arm of the Mediterranean between Greece and Turkey...)
5. Frame: Dimension: [DIMENSION SMALL] [OBJECT port] WN: (adj) small, little (limited or below average in number or quantity or magnitude or extent)
6. Frame: Locale_by_use: [DESCRIPTOR small] [LOCALE PORT] WN: (n) port (a place (seaport or airport) where people and merchandise can enter or leave a country)
7. Frame: Assistance: [HELPER a small port (skala)] [HELPER which] SERVES [BENEFITED_PARTY an inland settlement or chora], [RESULT making it easier to protect the island from attack] WN: (v) service, serve (be used by; as of a utility) “The sewage plant served the neighboring communities”
8. Frame: Locative_relation: [GROUND INLAND] [FIGURE settlement]
10. Frame: causation: [CAUSE a small port (skala) which serves an inland settlement or chora], MAKING it [EFFECT easier to protect the island from attack.] [AFFECTED DNI] WN: chora: not in WordNet (v) <i>make</i> , get (give certain properties to something) “This invention will make you a millionaire”
11. Frame: Difficulty: EASIER [ACTIVITY to protect the island from attack]. [EXPERIENCER INI] WN: (adj) easy (posing no difficulty; requiring little effort) “an easy job”; “an easy victory”
12. Frame: Protecting: [PROTECTION CNI] PROTECT [ASSET the island] [DANGER from attack] WN: (v) protect (shield from danger, injury, destruction, or damage) “Weatherbeater protects your roof from the rain”
14. Frame: Attack: from ATTACK . [ASSAILANT DNI] WN: (n) attack, onslaught, onset, onrush ((military) an offensive against an enemy (using weapons)) “the attack began at dawn”

Table 3: FN/WN Annotation of sentence 3