VerbNet: extensions and mappings to other lexical resources

Karin Kipper Schuler
kipper@linc.cis.upenn.edu

June 26th, 2006

Overview

Real world applications need resources with rich syntactic and semantic representations.

- Many existing broad-coverage resources provide only a shallow semantic representation
- Rich representations are needed
- Verbs are key elements in providing this

Overview

Natural language applications are currently limited to specific domains with hand-crafted lexicons.

- not available to the whole community
- expensive and time-consuming to build

Many available broad-coverage resources either focus on syntax or on semantics and do not provide a clear association between the two.

Semantic representation must be tied to the syntactic information:

- Differences between syntactic frames can help:

  Eng: John **left** the soccer field. (exited)
  Port: John saiu do campo.

  Eng: John **left** the ball on the field. (left)

  Port: John deixou a bola no campo.

- But syntax alone is not sufficient:

  Eng: John **left** the soccer field. (exited)
  Port: John saiu do campo.

  Eng: John **left** a fortune. (gave away)

  Port: John deixou uma fortuna.

Overview

Predicate argument relations are of interest for NLP, providing generalizations over data:

- *Ronaldo scored a goal for the Brazilian team*
- *A goal was scored by Ronaldo for the Brazilian team*
- *Ronaldo wanted to score a goal for the Brazilian team*

Outline

- Overview
- VerbNet
- Extensions of VerbNet
- Mappings to other Resources

VerbNet class entries

Kipper, Dang and Palmer, 2000

- verb classes based on Levin's classification
- classes defined by syntactic properties
- capture generalizations about verb behavior
- for each verb class
  - thematic roles
  - syntactic frames
  - selectional restrictions for the arguments in each frame
  - each frame includes semantic predicates with a time function

Thematic roles

- small set of roles (Agent, Theme, Location,..)
- roles used across classes
- provide as much information as possible for each class
- roles have semantic restrictions

Syntactic Frames

Describe possible surface realizations for verbs in a class

- constructions such as transitive, intransitive, resultative, and a large set of Levin's alternations
- Examples:

  1. *Agent V Patient*
     (John hit the ball)
  2. *Agent V at Patient*
     (John hit at the window)
  3. *Agent V Patient[+plural] together*
     (John hit the sticks together)

Semantic Predicates

Semantics of a syntactic frame captured through a conjunction of semantic predicates

- each semantic predicate includes a time function showing at what stage in the event the predicate holds
  start(E), during(E), end(E), result(E)
- similar to Moens and Steedman's event decomposition
- semantic predicates can be:
  General (e.g.,*motion* and *cause*), Specific (e.g.,*suffocate*), or Variable (Prep)

Hit class

Class	hit-18.1		
Parent	—		
Members	bang (1,3), bash(1), batter(1,2,3), beat(2,5), ..., hit(2,4,7,10), kick(3), ...		
Themroles	Agent Patient Instrument		
Selrestr	Agent[+int_control] Patient[+concrete] Instrument[+concrete]		
Frames	Name	Syntax	Semantic Predicates
	Transitive	Agent V Patient "Paula hit the ball"	cause(Agent, E) ∧ manner(during(E),directedmotion,Agent) ∧ !contact(during(E), Agent, Patient) ∧ manner(end(E),forceful, Agent) ∧ contact(end(E), Agent, Patient)
	Transitive with Instrument	Agent V Patient Prep(with) Instrument "Paula hit the ball with a stick"	cause(Agent, E) ∧ manner(during(E),directedmotion,Agent) ∧ !contact(during(E),Instrument,Patient) ∧ manner(end(E),forceful, Agent) ∧ contact(end(E), Instrument,Patient)

Hierarchical organization

Refinement of Levin classes

- verb classes are hierarchically organized
  - the original set of Levin classes has been further subdivided into additional subclasses which are more syntactic and semantically coherent
  - members have common semantic predicates, thematic roles, syntactic frames
  - a particular verb or subclass inherit from parent and may add more information

Current status of VerbNet

- 237 top-level classes, 194 additional subclasses
  - 5,000 verb senses (3,800 lemmas)
- characterized by:
  - 23 thematic roles types
    - 36 semantic restrictions on thematic roles
  - 131 syntactic frames (357 thematic role variants)
    - 55 syntactic restrictions
- 94 semantic predicates

Parameterized Action Representation (PAR)

Badler et al. (1999)

Interface to agents in an animation system.

Needs a semantically precise representation.

- Representation of actions
  - instructions to a virtual human
  - used in a simulated 3D environment
- Represented as
  - parameterized structures
  - hierarchical organization

PARs and VerbNet

PARs for animating agents require precise semantics associated with syntax provided by VerbNet.

- participants of an action are the arguments of a verb
- selectional restrictions on the arguments
- event structure (during, end, result)
- semantic components expressed by predicates

Outline

- Overview
- VerbNet
- Extensions of VerbNet
- Mappings to other Resources

Description of Korhonen and Briscoe's classes

(Korhonen and Briscoe, 2004)

Classes created using a semi-automatic approach to extend Levin's classification:

- 106 new diathesis alternations identified (many for sentential complements)
- 57 new classes identified (2-45 members each), with frames related by diathesis alternations

Integrating VerbNet and K&B's new classes

(Kipper, Korhonen, Ryant and Palmer, 2006)

Two major tasks were involved in this integration:

1. assigning VerbNet-style detailed syntactic-semantic descriptions to the new classes
   - because of the different sets of subcategorization frames uncovered in K&B, new roles, new syntactic descriptions and restrictions, and new semantic predicates needed to be added to VN
2. incorporating the new classes into the VerbNet database

Integrating VerbNet and K&B's new classes

Assigning VerbNet-style syntactic-semantic descriptions to the new classes required the addition of:

- thematic roles (+2)
- syntactic frames to account for new alternations (+76)
- syntactic restrictions (+52)
  (to account for object control, subject control, and different types of complements)
- semantic predicates (+30)
- increased number of classes from 191 to 237
- 320 new verb senses and 200 new lemmas added

Integrating VerbNet and K&B's new classes

We used 55 of the initial 57 classes in the integration.
These classes fell in three categories:

- entirely new classes (35)
  Classes did not overlap with existing VerbNet classes
  (e.g., URGE, FORBID)
- included as subclasses of existing classes (7)
  New class semantically or syntactically similar to existing class
  (e.g., CONVERT AND SHIFT added as subclasses of *Turn-26.6*)
- reorganization of the original classes (13)
  Existing classes focused mainly on NP and PP,
  many verbs classify better by sentential complements
  (e.g., WANT and *Want-32.1*)

Notes on K&B integration

New classes have already been uncovered (Korhonen and Ryant, 2005) and added to VerbNet (Euralex 2006).

Total number of classes after both integrations is 274

Addressing coverage:

- investigated the coverage of the 274 classes over PropBank
- *without* new classes VerbNet matches **78.45%** of the verb tokens in the annotated PropBank data (88,584 occurrences)
- *including* new classes VerbNet matches **90.86%** of the verb tokens in PropBank

Extending VerbNet's members – LCS

Dorr (2001)

Addition of members from the LCS database

- inspected 1,266 verbs present in the LCS database and not in VerbNet
- 429 (426 lemmas) were initially integrated into our lexicon
- verbs had been acquired automatically, data noisy

Automatic acquisition of verbs – Clusters

Kingsbury and Kipper (2003); Kingsbury (2004)

- used PropBank subcategorization frames (e.g., *Arg0.V.Arg1*)
- 121 clusters from the EM algorithm (0 to 45 elements each)
- 1,278 verbs which occurred at least 10 times in the PropBank annotation were used as data
- 484 verbs were already in VerbNet class
  (824 potential candidates for inclusion in VerbNet classes)

Automatic acquisition of verbs – Clusters

Results:

- 5.6% of the candidates were included in VerbNet
- large clusters were not predictive of any classes
- small clusters did not offer many candidates
- 12.6% if using only "good clusters"
- need better way to filter the clusters
- impoverished features
- senses predicted in VerbNet and PropBank are different

Extending VerbNet with WordNet

(Loper, Kipper and Palmer)

- use WordNet as a source of candidates for inclusion in VerbNet
- use syntactic contexts of these verbs in Propbank
- candidates are filtered based on the grammatical patterns and the relationship between those patterns and known members of VerbNet classes
- 707 lemmas suggested, 849 senses
- 208 lemmas, 255 senses integrated into the suggested classes
- experiment done on version 1.5 of VerbNet

Extending VerbNet with WordNet

Experiment redone using version 2.2 of VerbNet:

- 9,302 senses (4,992 lemmas) suggested
- inspected only candidates with similar context as VerbNet member
- 179 (out of 413) added to VerbNet (43.34%)
- lack of semantic features limited the experiment

Summary of Extensions

Source	Lemmas added	Senses added
K&B 359/410	200	320
LCS 1134/1266	426	429
Cluster 824	47	47
WordNet 707/849	208	255

Outline

- Overview
- VerbNet
- Extensions of VerbNet
- Mappings to other Resources

Linking resources

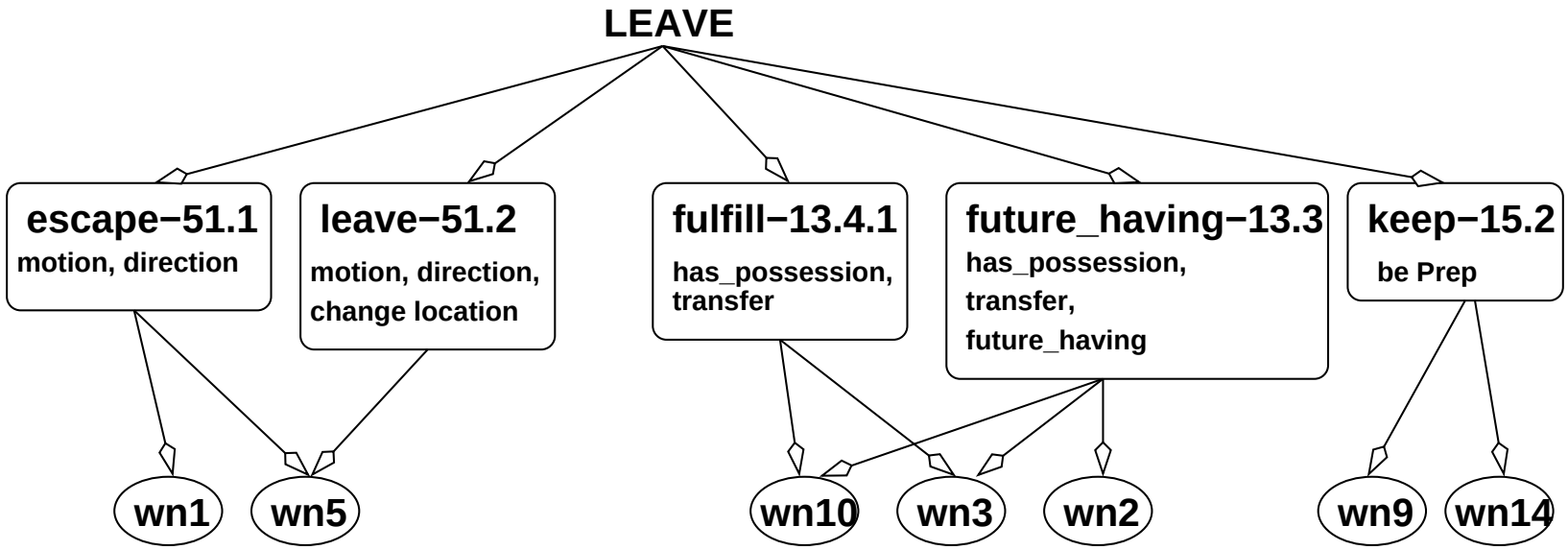
Many applications would benefit by merging the results of different lexical resources and annotation projects:

- compatibility between resources
- inherent theoretical differences
- different levels of representation

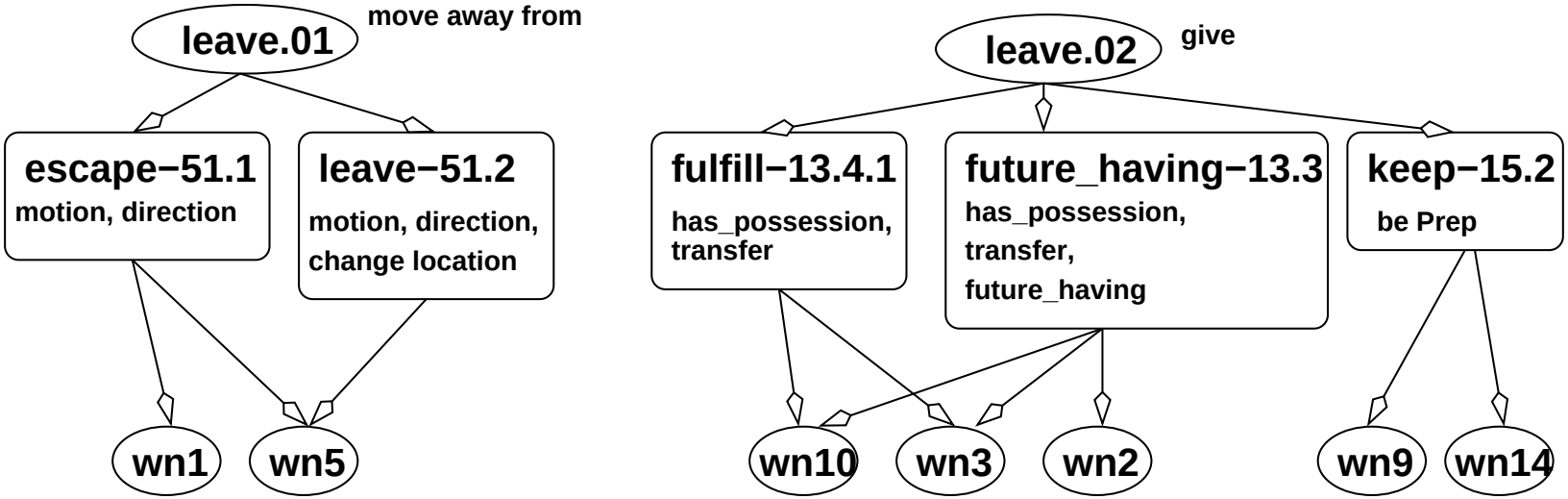
Semlink: develop computationally explicit connections between FrameNet, PropBank, and VerbNet.

Mappings between VerbNet and WordNet

Each verb in VerbNet is mapped to its corresponding synset(s) in WordNet, if available.



PropBank/VerbNet/WordNet



Mappings between VerbNet and FrameNet

Two steps:

1. mappings between VerbNet verb senses to FrameNet;

VN class	VN member	FN frame
9.1	arrange	(diff. sense)
9.1	immerse	Placing
9.1	lodge	Placing
9.1	mount	Placing
9.1	sling	

2. mappings from VerbNet thematic roles to the FrameNet frame elements

VNclass 9.1	FNframe "Placing"
VN role	FN frame element
Agent	Agent
Agent	Cause
Destination	Goal
Theme	Theme

Mappings between VerbNet and FrameNet

Dolbey, Kipper, and Palmer (in progress)

- 4756 VN verb senses
- 3294 FN senses (2333 lemmas)
  - 2170 have corresponding entry in FN
  - 796 different sense in FN
  - 1790 VN lemma does not exist in FN
- 673 mappings
- 263 unique FrameNet frames assigned to VerbNet

Assigning Xtag trees to VerbNet

Ryant and Kipper (2004)

- VerbNet only describes declarative frames
- Xtag provides detailed account of syntactic transformations
- mapping VerbNet syntactic frames to Xtag trees extends VerbNet syntactic coverage while providing semantics for the Xtag trees
- 104 VerbNet syntactic frames (out of 131) map to 19 Xtag tree families

Annotating PropBank with VerbNet

Loper and Palmer (in progress)

The annotation consists of:

- each PB frameset is annotated with VN class
- each PB corpus instance is annotated with VN role labels (instead of argument numbers Arg0, Arg1,..)

Uses:

- train classifiers to automatically map PB-labeled instances to VN-labeled instances
- train semantic role labelers that use VerbNet role labels instead of PropBank argument numbers

Annotating PropBank with VerbNet

Mapping was done in two ways:

- lexical mapping:
  - define the set of possible mappings between the two lexicons
  - semi-automatic while creating PB frame files and later revised
- instance classifier:
  - chooses the best mapping for each instance in the corpus
  - uses two heuristic classifiers:
    - SenseLearner WSD engine: finds WN class of verb instance, selects VN class based on mappings between WN and VN
    - examines syntactic context of verb instance and selects class with syntactic frame that most closely matches the instance's context

Annotating PropBank with VerbNet

The mapping does not currently cover all instances from the PropBank:

- 19% of the PB instances use verbs that are not present in VN
- 6% of the PB instances use verb senses that are not currently covered by VN
- 23% of the verbs are included contain mappings between verb instance and VN class but the individual arguments cannot currently be mapped

Mappings

These resources are:

- Complementary
- Redundancy is harmless, may even be useful
- PropBank provides great training data
- VerbNet provides clear links between syntax and semantics
- FrameNet provides rich semantics
- Together they give us the most comprehensive coverage

Conclusion

To achieve the detailed level of representation required for natural language applications we need resources capable of providing a rich semantic representation tied to syntax.

VerbNet:

- broad-coverage, general purpose natural language resource
- focuses on both syntax and semantics and provides a clear association between the two, necessary for characterizing verbs

Conclusion

The extensions provided by Korhonen and Briscoe, Korhonen and Ryant, the LCS database, and WordNet greatly increased VerbNet's coverage.

The efforts in mapping between resources provide the community with several complementary layers of syntactic-semantic representation.