Lexical relations between verbs

Verbs can be

- **synonyms**, e.g., *pass away–die*
- **hyponyms**, e.g., *walk–move*
- **meronyms**, e.g., *wash – soak, scrub, wring out, (dry)*. [rare]
- **opposites**
  - **indirect converses** such as *bequeath–inherit; give–receive*
  - **reversives** such as *enter–leave, mount–dismount*

Overall, lexical relationships between verbs are weak and unsystematic in comparison to those in operation between nouns.
**Hyponymy vs Entailment**

X is a verbal **hyponym** of Y if the following test frame succeeds:

**Test frame “To X is necessarily to Y”**

To murder someone is necessarily to kill them.
To strangle someone is necessarily to kill them.

**Entailment:** a causal relationship between propositions, which includes and goes beyond hyponymy.

He killed the bee – The bee died.
He snored – He was sleeping.

**Troponymy:** subtype of hyponymy; manner of an action.
(Cruse (1979) calls this property **verbal taxonomy**.)

**Test frame: “To X is a way of Y-ing”**

To strangle/?murder somebody is a way of killing.
To crawl/?travel is a way of moving.

Thus, strangle is a troponym of kill. murder is not a troponym of kill, but of commit a crime.

Wordnet distinguishes four types of lexical relations between verbs: hyponymy, troponymy, entailment, meronymy. Few instances in comparison to nouns.
NLP methods for finding verb similarities

Verbs with similar semantics tend to . . .

- have similar subcategorisation behaviour → cluster verbs by their subcategorisation patterns; e.g., Schulte (2006); Lin and Korhonen (2009)
- undergo the same diathesis alternations. → manually classify verbs (Levin 1993)
- have similar selectional restrictions → determine the difference between two verbs' selectional restrictions; e.g., Resnik (1995)
- have similar thematic roles, i.e., participants in the actions they denote → perform semantic role labelling, e.g., Gildea and Jurafsky (2002)

Linguistic Selection

A selector imposes semantic constraints on its selectees.

**Head–complement construction**

*I have been waiting for hours.*  
*I have been waiting for the bus.*  
Selector: verb, Selectee: arguments

**Head–modifier constructions**

graceful degradation  
Selector: modifier, Selectee: head

**Verb–subject constructions**

*The water froze within seconds.*  
Selector: verb, Selectee: subject (most likely)
**Selectional restrictions:** Violation of selector’s presuppositions results in paradox or incongruity.

- This cannot be resolved by replacement with synonym
- But it can be resolved by replacement with near hypernym (in the case of paradox).

- ? my male aunt – paradox; resolvable (relation).
- ? the cat barked – paradox; resolvable (animal).
- ? a lustful affix – incongruity; unresolvable (except by thing).

---

**Collocational restrictions:** Violation of selector’s presuppositions results in inappropriateness; resolvable by replacement with synonym.

- ? The aspidistra kicked the bucket – resolvable (died).

<table>
<thead>
<tr>
<th>unblemished</th>
<th>spotless</th>
<th>flawless</th>
<th>immaculate</th>
<th>impeccable</th>
</tr>
</thead>
<tbody>
<tr>
<td>performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>argument</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>complexion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>behaviour</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kitchen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>record</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reputation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>taste</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>order</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>credentials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Phenomenology: Aspects of similarity in verbs
Selectional Restrictions and Subcategorisation Frames
Frame Semantics
Semantic Role Labelling

Resnik 1995
Automatic verb clustering

... vs collocational restrictions

Collocational restrictions are highly **unpredictable**.

<table>
<thead>
<tr>
<th></th>
<th>unblemished</th>
<th>spotless</th>
<th>flawless</th>
<th>immaculate</th>
<th>impeccable</th>
</tr>
</thead>
<tbody>
<tr>
<td>performance</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>argument</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>complexion</td>
<td>?</td>
<td>?</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>behaviour</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>kitchen</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
</tr>
<tr>
<td>record</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>?</td>
<td>X</td>
</tr>
<tr>
<td>reputation</td>
<td>?</td>
<td>X</td>
<td>-</td>
<td>?</td>
<td>-</td>
</tr>
<tr>
<td>taste</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>?</td>
<td>X</td>
</tr>
<tr>
<td>order</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>credentials</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
</tbody>
</table>

**Quantifying selectional preferences: Resnik 1995**

- **Selectional preference strength** $S_R(v)$ of verb $v$: the degree of selectiveness of a predicate about the semantic class of its arguments; expressed in bits of information.

- Semantic classes $c$ are WordNet synsets

- $S_R(v)$ is based on difference in distribution between
  - $P(c)$ – likelihood of direct object of falling into semantic class $c$
  - $P(c|v)$ – likelihood of direct object of falling into semantic class $c$ if associated with verb $v$

- Use KL divergence to determine $S_R(v) = D(P(c|v)||P(c))$:

$$S_R(v) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}$$
Resnik (1995), ctd

- Selectional association between a verb and a class (synset) is the **relative** contribution to the overall selectionality of the verb

\[ A_R(v, c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)} \]

Example result:

<table>
<thead>
<tr>
<th>Verb</th>
<th>Dir. Obj. (preferred)</th>
<th>Assoc</th>
<th>Dir Obj. (dispreferred)</th>
<th>Assoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>read</td>
<td>WRITING</td>
<td>6.80</td>
<td>ACTIVITY</td>
<td>-0.20</td>
</tr>
<tr>
<td>write</td>
<td>WRITING</td>
<td>7.26</td>
<td>COMMERCE</td>
<td>0</td>
</tr>
<tr>
<td>see</td>
<td>ENTITY</td>
<td>5.79</td>
<td>METHOD</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

- The Resnik algorithm can be used to perform WSD.

**Diathesis alternation; Levin (1993)**

**Definition**

Systematic variations in the expression of arguments, sometimes accompanied by changes in meaning (Levin, 1993)

Famous example:

**Dative alternation**

- Doris gives flowers to the headmistress.
- Doris gives the headmistress flowers.

This pattern is meaning-preserving and covers several semantic classes:

- verbs of “future having”: advance, allocate, offer, owe, lend
- verbs of “sending”: forward, hand, mail
- verbs of “throwing”: kick, pass, throw
The sailors gullied the whales.

“gully” is an archaic whaling term. What does it mean?

We observe a strong correlation between syntactic behaviour and semantic class.

Whales gully easily.

We observe a strong correlation between syntactic behaviour and semantic class.
The sailors gullied the whales.

“gully” is an archaic whaling term. What does it mean?

Whales gully easily.

Has your hypothesis changed?

We observe a strong correlation between syntactic behaviour and semantic class.

Other diathesis alternations

John loaded the truck with hay.
John loaded hay on the truck.

Semantic difference?
Other verbs following this pattern? (spray? fill? pour? dump? cover? (this is called the locative alternation.)

John cuts the bread.
The bread cuts nicely. (middle)
John cut Mary’s arm/Mary on the arm (bodypart possessor ascension)
John cut at the bread (conative)

Other verbs following this pattern?
**Alternations and Semantic Dimensions**

- Bodypart Possessor Ascension Alternation is sensitive to **contact** — separating out *break* as a non-contact verb (pure change of state)
- Conative Alternation is sensitive to **both motion and contact** — separating out *touch* as a verb of contact (non-change of state)
- Middle Alternation is sensitive to **change of state** — identifying hit as non-change-of-state (contact by motion verb), whereas cut is a verb of “cause of change of state by moving sth int contact with entity that changes state”
Levin’s (1993) Verb Classification

- Based on 79 diathesis alternations
- Covers 3200 verbs in 48 main classes (191 subdivided ones)
- \textit{break} class contains: \textit{break, chip, crack, crash, crush, fracture, rip, shatter, smash, snap, splinter, split} and \textit{tear}.
- Diathesis alternations are difficult to detect automatically
- But: we can use the fact that similar alternations result in similar SCF (subcategorisation frames).

Verb clustering with subcategorisation frames and selectional restrictions

- Use spectral clustering algorithm and many features
- Evaluation:
  - Standard Test set 1 (TS1): 15 course- and fine-grained Levin classes, 10-15 verbs per class; 205 verbs
  - Test set 2 (TS2): 17 fine-grained Levin classes with 12 members each, resulting in 204 verbs
- Use all occurrences of verb (up to 10,000) from corpora
- Far superior results to previous literature (unsupervised); 0.58 F-measure (previously 0.31) on T1; 0.80 F-measure on T2 (previously best unsupervised 0.51)
Sun and Korhonen’s Features

- Collocation (CO): 4 words immediately preceding and following lemmatised verb. Remove stop words, keep 600 most frequent words.
- Prepositional preference (PP): type and frequency of prepositions in direct object relation
- Lexical Preference (LP): type and frequency of nouns and prepositions in subject, object, indirect object relation
- Subcategorisation frames and relative frequencies with verbs
- Selectional Preferences
- Tense of verb
- Many combinations of these

Frame Semantics

Due to Fillmore (1976);
- a frame describes a prototypical situation;
- it is evoked by a frame evoking element (FEE);
- it can have several frame elements (semantic roles).
Due to Fillmore (1976);
a frame describes a prototypical situation;
it is evoked by a frame evoking element (FEE);
it can have several frame elements (semantic roles).

Mathilde fried the catfish in a heavy iron skillet.
Due to Fillmore (1976);
- a frame describes a prototypical situation;
- it is evoked by a frame evoking element (FEE);
- it can have several frame elements (semantic roles).

\[
\begin{align*}
\text{Mathilde} & \quad \text{fried} \quad \text{the catfish} & \quad \text{in a heavy iron skillet.}
\end{align*}
\]

Properties of Frame Semantics
- Provides a shallow semantic analysis (no modality, scope);
- generalizes well across languages;
- can benefit various NLP tasks (IR, QA).
Types of thematic roles

- Verb-specific frames and domain-specific roles:
  - kiss — Kisser – Kissee
  - From-Airport, To-Airport, Departure-Time
- Only two roles: Proto-Agent, Proto-Patient
- Mid-level: AGENT, EXPERIENCER, INSTRUMENT, OBJECT, SOURCE, GOAL, LOCATION, TIME, and PATH (Fillmore, 1971).
- Granularity in FrameNet is situated between mid-level and verb-specific.

FrameNet Corpus

FrameNet is a corpus with frame semantics markup:

- uses a tagset of 76 semantic roles (frame elements) from 12 general semantic domains (body, cognition, communication);
- consists of a sample of sentences from the BNC annotated with frame elements;
- 49,013 sentences and 99,232 frame elements in total;
- this includes 927 verbs, 339 nouns, 175 adjectives.

The sentences in the corpus were not chosen from the BNC at random; rather representative usages were selected.
Some FrameNet Examples for RISK

- She **risked** her life on the ascent of K2.
- You would not really want to **risk** annoying her.

**Agent/Protagonist, BadOutcome, Asset, Action**

She risked **one of her elaborate disguises** when she went out that day
- ...because she had been hidden in that hotel room for long enough.
- ...because she suspected they already had a photo of her in it.

---

**RISK in FrameNet**

<table>
<thead>
<tr>
<th>FrameNet</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DARING</strong></td>
<td>an Agent performs an Action which can harm the Agent and which is considered imprudent.</td>
</tr>
<tr>
<td><strong>RUN_RISK</strong></td>
<td>Protagonist is exposed to a dangerous situation, which may result in a BadOutcome or the loss of an Asset. There is no implication of intentionality on behalf of the Protagonist.</td>
</tr>
<tr>
<td><strong>RISKY_SITUATION</strong></td>
<td>a Situation is likely to result in a (non-mentioned) harmful event befalling an Asset</td>
</tr>
<tr>
<td><strong>BEING_AT_RISK</strong></td>
<td>An Asset is exposed to or otherwise liable to be affected by a HarmfulEvent, which may occur as DangerousEntity.</td>
</tr>
</tbody>
</table>
Gildea and Jurafsky (2002):

1. Parse the training corpus using Collin’s parser;
2. Match frame elements to constituents;
3. Extract features from the parse tree;
4. Train probabilistic model on the features.

The start and end word of each parsed constituent is found and matched against a frame element with the same start and end. No match is possible in 13% of the cases (parsing errors).
Assume the sentences are parsed, then the following features can be extracted for role labeling:

- **Phrase Type**: syntactic type of the phrase expressing the semantic role (e.g., NP, VP, S);
- **Governing Category**: syntactic type of the phrase governing the semantic role (NP, VP); distinguishes subject-NPs from object-NPs;
- **Parse Tree Path**: path through the parse tree from the target word to the phrase expressing the grammatical role;
- **Position**: whether the constituent occurs before or after the predicate; useful for incorrect parses;
- **Voice**: active or passive; use heuristics to identify passives;
- **Head Word**: the lexical head of the constituent.

Path from target *ate* to frame element *He*: VB↑VP↑S↓NP

“If there is an underlying AGENT, it becomes the syntactic subject (Fillmore, 1968)”
Paths and Grammatical Roles

<table>
<thead>
<tr>
<th>Freq.</th>
<th>Path</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.2%</td>
<td>VB↑VP↓PP</td>
<td>PP argument or adjunct</td>
</tr>
<tr>
<td>11.8</td>
<td>VB↑VP↑S↓NP</td>
<td>subject</td>
</tr>
<tr>
<td>10.1</td>
<td>VB↑VP↓NP</td>
<td>object</td>
</tr>
<tr>
<td>7.9</td>
<td>VB↑VP↑VP↑S↓NP</td>
<td>subject of embedded VP</td>
</tr>
<tr>
<td>4.1</td>
<td>VB↑VP↓ADVP</td>
<td>adverbial adjunct</td>
</tr>
<tr>
<td>3.0</td>
<td>NN↑NP↑NP↓PP</td>
<td>prepos. complement of noun</td>
</tr>
<tr>
<td>1.7</td>
<td>VB↑VP↓PRT</td>
<td>adverbial particle</td>
</tr>
<tr>
<td>1.6</td>
<td>VB↑VP↑VP↑S↓NP</td>
<td>subject of embedded VP</td>
</tr>
<tr>
<td>14.2</td>
<td>no matching parse constituent</td>
<td>other</td>
</tr>
<tr>
<td>31.4</td>
<td>other</td>
<td></td>
</tr>
</tbody>
</table>

Divide the FrameNet corpus into:

- 10% test set;
- 10% development set;
- 80% training set;

Relatively small training set: average number of sentences per target word is 34, number of sentences per frame is 732.

Simone Teufel L113 Word Meaning and Discourse Understanding 30
Build a classifier by combining conditional distributions of the features. Compute the distribution from the training data, e.g.:

\[ P(r|pt, t) = \frac{\#(r, pt, t)}{\#(pt, t)} \]  

where:
- \( r \) = semantic role
- \( pt \) = phrase type
- \( gov \) = governing category
- \( pos \) = position
- \( voice \) = voice
- \( h \) = head word
- \( t \) = target word (predicate)

Measure the performance of a distribution using the following metrics:

- **Coverage**: percentage of the test data for which the conditioning event has been seen in the training data.
- **Accuracy**: percentage of covered test data for which the correct role is predicted.
- **Performance**: product of coverage and accuracy.

Baseline: always choose most probable role for each target word (40.9%)
Evaluation

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Coverage</th>
<th>Accuracy</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(r</td>
<td>t)$</td>
<td>100</td>
<td>40.9</td>
</tr>
<tr>
<td>$P(r</td>
<td>pt, t)$</td>
<td>92.5</td>
<td>60.1</td>
</tr>
<tr>
<td>$P(r</td>
<td>pt, gov, t)$</td>
<td>92.0</td>
<td>66.6</td>
</tr>
<tr>
<td>$P(r</td>
<td>pt, pos, voice)$</td>
<td>98.8</td>
<td>57.1</td>
</tr>
<tr>
<td>$P(r</td>
<td>pt, pos, voice, t)$</td>
<td>90.8</td>
<td>70.1</td>
</tr>
<tr>
<td>$P(r</td>
<td>h)$</td>
<td>80.3</td>
<td>73.6</td>
</tr>
<tr>
<td>$P(r</td>
<td>h, t)$</td>
<td>56.0</td>
<td>86.6</td>
</tr>
<tr>
<td>$P(r</td>
<td>h, pt, t)$</td>
<td>50.1</td>
<td>87.4</td>
</tr>
</tbody>
</table>

Final system performance 80.4, using head word, phrase type, target word, path and voice.

But there are 3 features modelling grammatical function – which is best (pos, path, gov)?

Voice is beneficial only if at least one of these 3 is used.

If we don’t have voice, position is best (79.9%).

Position + voice instead of either path or governing category is equivalent;

Head words are very accurate indicators of a constituent’s semantic role; $P(r|h, t)$ can only be evaluated on 56.0% of the date, but was 86.7% correct.
So far, we have assumed that the boundaries of the frame elements are known; this is not the case in a realistic setting. The system is no longer given frame element boundaries but has to find them.

fei is a binary feature: is given constituent a frame element or not. The same features can be used to identify frame elements.

Compute the following conditional distribution:

\[ P(\text{fe}|p, h, t) = \lambda_1 P(\text{fe}|p) + \lambda_2 P(\text{fe}|p, t) + \lambda_3 P(\text{fe}|h, t) \]

The features used are Parse Tree Path, Head Word, and Target Word, as previously introduced.

<table>
<thead>
<tr>
<th>Type of Overlap</th>
<th>% identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exactly matching boundaries</td>
<td>66</td>
</tr>
<tr>
<td>Identified constituent entirely with FE</td>
<td>8</td>
</tr>
<tr>
<td>FE entirely within identified constituent</td>
<td>7</td>
</tr>
<tr>
<td>Partial overlap</td>
<td>0</td>
</tr>
<tr>
<td>No match to true frame element</td>
<td>13</td>
</tr>
</tbody>
</table>
Head words are good predictors of semantic role, but data is sparse. This can be overcome using:

- **Clustering**: find words that are similar to head words that do not occur in the training data; increases performance to 85%.
- **WordNet**: if a word is not in the training data, use its hypernym in WordNet; percolate co-occurrence counts up the WordNet hierarchy (problem: multiple hierarchies and multiple word senses); increases accuracy to 84.3%.
- **Bootstrapping**: label unannotated data with the automatic system, use the resulting data as training data; increases accuracy to 83.2%.

**Summary**

- Semantic role labeling means identifying the constituents (frame elements) that participate in a prototypical situation (frame) and labeling them with their roles.
- This provides a shallow semantic analysis that can benefit various NLP tasks.
- FrameNet is a corpus/dictionary marked up with semantic roles.
- A simple probabilistic model combining lexical and syntactic features performs well on the task.
- The model interpolates distributions or performs backoff.
- Similar features can be used for identifying frame elements.
- In both models, lexical statistics are sparse, which can be addressed with clustering, WordNet, or bootstrapping.
Lin and Korhonen (2009). EMNLP


Jurafsky and Martin, chapters 19.4, 20.4.2 (selectional restrictions) and 20.9 (frames)

Thematic Roles reviewed, Chapter 10 of ?

Allan, Frames, Fields and Semantic components – chapter 8 of book “Natural Language Semantics”.

Cruse, 2.2 (arguments) 14.4.4 (thematic roles)