Valency in Other LRs: PropBank, VerbNet, FrameNet

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz
Proposition Bank
Annotations in PropBank

- based on Penn TreeBank
- transfer sentences to propositions
- goal is to annotate every tree systematically
  - so statistics in the corpus are meaningful
- generally more data-driven & bottom up
  - no level of abstraction beyond verb senses
  - annotate every verb you see
- Martha Palmer, Mitch Marcus, Joseph Rosenzweig, Paul Kingsbury, Hoa Dang, Karin Kipper, Scott Cotton, Laren Delfs, Christiane Fellbaum

http://verbs.colorado.edu/~mpalmer/projects/ace.html
When Powell met Zhu Rongji on Thursday they discussed the return of the spy plane.
Penn English Treebank

- 1.3 million words
- Wall Street Journal and other sources
- tagged with Part-of-Speech
- syntactically Parsed
- widely used in NLP community
- available from Linguistic Data Consortium

- the same data as in the (eng. part of) PCEDT
Lorillard Inc., the unit of New York-based Loews Corp. that *T*-2 makes Kent cigarettes, stopped using crocidolite in its Micronite cigarette filters in 1956.

[Lorillard, pobočka newyorské firmy Leows corp., která vyrábí cigarety Kent, přestala používat crocidolit ve filtrech svých cigaret Micronite v roce 1956.]
Penn English Treebank

original annotation → a-layer from PCEDT
Transfer sentences to propositions

- Penn TreeBank $\rightarrow$ PropBank
  - Add a semantic layer on Penn TreeBank
  - Define a set of semantic roles for each verb
  - Each verb’s roles are numbered

- Frameset: $\text{hate.01}$
  - ARG0: experiencer
  - ARG1: target

$\text{hate}$

\[\text{S} \rightarrow \text{NP} \rightarrow \text{NNS} \rightarrow \text{VB} \rightarrow \text{NN} \rightarrow \text{NP} \rightarrow \text{VP} \rightarrow \text{NN} \rightarrow \text{NN} \rightarrow \text{NN} \rightarrow \text{NN}\]

PDT: valency III Lopatková
Analysts have been expecting a GM-Jaguar pact that would give the U.S. car maker an eventual 30% stake in the British company.

[Analytici očekávali smlouvu mezi GM a Jaguárem, která by zaručovala výrobci aut případnou 30% investici v britské firmě.]
Analytici očekávali smlouvu mezi GM a Jaguárem, která by zaručovala výrobci aut případnou 30% investici v britské firmě.

[Expect(Analysts, GM-J pact) Give(GM-J pact, US car maker, 30% stake)]

PDT: valency III

Lopatková
Lorillard Inc., the unit of New York-based Loews Corp. that *T*-2 makes Kent cigarettes, stopped using crocidolite in its Micronite cigarette filters in 1956.
Lorillard Inc., the unit of New York-based Loews Corp. that *T*-2 makes Kent cigarettes, stopped using crocidolite in its Micronite cigarette filters in 1956.
PropBank: Predicate Annotation

- a general set of semantic roles for all types of predicates
- semantic roles for each verb and sense in the frame files
- the (core) arguments are labeled by numbers
- adjunct-like arguments – universal to all verbs
PropBank: Predicate Annotation

- **argument types** defined on a "per-verb" basis
- consistent across uses of a single verb (sense)
- but the same tags are used (Arg0, Arg1, Arg2, …)

  Arg0 ≈ proto-typical agent (Dowty)
  - Agent (85%), Experiencer (7%), Theme (2%), …

  Arg1 ≈ proto-typical patient
  - Theme (47%), Topic (23%), Patient (11%), …

  Arg2 = Recipient (22%), Extent (15%), Predicate (14%), …

  Arg3 = Asset (33%), Theme2 (14%), Recipient (13%), …

  Arg4 = Location (89%), Beneficiary (5%), …

  Arg5 = Location (94%), Destination (6%)
PropBank: Predicate Annotation

- **adjunct-like arguments** – universal to all verbs:
  - LOC (where at), TMP (when), DIR (where to), MNR (how), ADV (others), PRP (why)
  - REC (himself, themselves, each other), PRD (this argument refers to or modifies another)
  - NEG, MOD (modals and negation marked as ArgMs)
  - EXT (amount), PNC, DIS
Frame Examples

• Transfer sentences to propositions
  \[\text{Kristina} \text{ hit} \ \text{Scott} \rightarrow \text{hit(Kristina,Scott,with a baseball)}\]
Frame Examples

• Transfer sentences to propositions
  Kristina hit Scott → hit(Kristina, Scott, with a baseball)

• hit.01 “strike”
  Arg0: agent, hitter;
  Arg1: thing hit;
  Arg2: instrument, thing hit by or with
  \[A_0 \text{Kristina} \text{hit} \ A_1 \text{Scott} \ A_2 \text{with a baseball}] \text{yesterday.}
PropBank: Predicate Annotation

• look.02 “seeming”
  Arg0: seemer;
  Arg1: seemed like;
  Arg2: seemed to
  \[A0 \text{It} \text{looked}_{\text{REL}} [A2 \text{to her}] \text{like} [A1 \text{he deserved this}].\]

• deserve.01 “deserve”
  Arg0: deserving entity;
  Arg1: thing deserved;
  Arg2: in-exchange-for
  \[\text{It looked to her like} [A0 \text{he} \text{deserved}_{\text{REL}} [A1 \text{this}].\]

• expect.0? “anticipate”
  Arg0: expecter;
  Arg1: thing expected
  \[A0 \text{Portfolio managers} \text{expect}_{\text{REL}} [A1 \text{further declines in interest rates}].\]

• give.01 “transfer”
  Arg0: giver;
  Arg1: thing given;
  Arg2: entity given to
  \[A0 \text{The executives} \text{gave}_{\text{REL}} [A2 \text{the chefs}] [A1 \text{a standing ovation}].\]
PropBank: Argument Numbering

• task: not just undoing passives
  [A0 The earthquake] shook [A1 the building].

• how are arguments numbered?
  • Examination of example sentences
  • Determination of required / highly preferred elements
  • Sequential numbering, Arg0 is typical first argument
PropBank: Argument Numbering

- **task**: not just undoing passives
  
  \[A0 \text{The earthquake} \text{ shook} [A1 \text{ the building}].\]
  
  \[A1 \text{The walls} \text{ shook}; \quad [A1 \text{ the building}] \text{ rocked}.\]

- **how are arguments numbered?**
  
  - Examination of example sentences
  
  - Determination of required / highly preferred elements
  
  - Sequential numbering, Arg0 is typical first argument

**BUT:**  
- ergative/unaccusative verbs (shake example)

  rise.01 “go up”

  Arg1: Logical subject, patient, thing rising;

  Arg2: EXT, amount risen;

  Arg3* = start point

  Arg4 = end point

  Example: \[A1 \text{ Sales} \text{ rose}_\text{REL} [A2 4\%] [A4 \text{ to $3.28 billion}] [A3 \text{ from $3.16 billion}].\]

- arguments mapped for "synonymous" verbs

  (consistent between different verbs that share similar argument structures)
VerbNet
VerbNet – main characteristics

• the largest on-line verb lexicon for English
• Kipper-Schuler 2006
• main characteristics:
  • hierarchical
  • domain-independent
  • broad-coverage
• mappings to other lexical resources
  • PropBank … 90.86% coverage of PropBank tokens
  • WordNet (Miller, 1990; Fellbaum, 1998)
  • FrameNet (Baker et al., 1998)
  • Lexicalized Tree Adjoining Grammar (XTAG Research Group, 2001)
Organization of VerbNet

- organized into verb classes extending Levin (1993) classes
- syntactic and semantic coherence among members of a class
- verb class in VN is described by
  - thematic roles
  - selectional restrictions on the arguments
  - frames = syntactic description + semantic predicates
**VerbNet entry: Hit-18.1 class** (simplified)

### Class Hit-18.1

**Roles and Restrictions:**
- Agent[+int_control]
- Patient[+concrete]
- Instrument[+concrete]

**Members:**
- bang, bash, hit, kick, ...

### Frames:

<table>
<thead>
<tr>
<th>Name</th>
<th>Example</th>
<th>Syntax</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Transitive</td>
<td>Paula hit the ball</td>
<td>Agent V Patient</td>
<td>• cause(Agent, E)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• manner (during(E), directedmotion, Agent)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• !contact (during(E), Agent, Patient)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• manner(end(E), forceful, Agent)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• contact(end(E), Agent, Patient)</td>
</tr>
</tbody>
</table>
VerbNet – basic statistics

<table>
<thead>
<tr>
<th></th>
<th>VerbNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-level classes</td>
<td>274</td>
</tr>
<tr>
<td>Thematic roles</td>
<td>23</td>
</tr>
<tr>
<td>Semantic predicates</td>
<td>94</td>
</tr>
<tr>
<td>Syntactic restrictions (on sentential compl)</td>
<td>55</td>
</tr>
<tr>
<td>Number of verb senses</td>
<td>5257</td>
</tr>
<tr>
<td>Number of lemmas</td>
<td>3769</td>
</tr>
</tbody>
</table>
VerbNet – Thematic Roles

- Actor
- Agent
- Asset
- Attribute
- Beneficiary
- Cause
- Location, Destination, Source
- Destination
- Source
- Location
- Experiencer
- Extent
- Instrument
- Material and Product
- Material
- Product
- Patient
- Predicate
- Proposition
- Recipient
- Stimulus
- Theme
- Time
- Topic
VerbNet
Selectional Preferences

SelRestr
- concrete
  - time
  - state
  - abstract
  - scalar
- abstract
  - idea
  - sound
  - communication
- SelRestr
  - currency
  - location
  - organization
- concrete
  - solid
  - shape
  - rigid
  - pointed
  - elongated
- concrete
  - int-control
  - force
- concrete
  - natural
  - animal
  - plant
  - body-part
  - machine
  - garment
  - tool
  - non-rigid
  - comestible

PDT: valency III
VerbNet – Unified Verb Index

• complete list of the thematic roles, selectional and syntactic restrictions, predicates, and frame types:
  http://verbs.colorado.edu/verb-index/reference

• mappings:
  • VerbNet senses to FrameNet senses
    http://verbs.colorado.edu/verb-index/vn-fn.xml
  • VerbNet Role to FrameNet Frame Elements
    http://verbs.colorado.edu/verb-index/vn-fn-roles.xml
  • SemLink: PropBank, FrameNet, VerbNet
    http://verbs.colorado.edu/semlink/
FrameNet
FrameNet Background: Case Theory

• Fillmore: Case Theory (1968)
  • deep structure
  • surface-structure cases are derived
  • universal set of roles
    - $A(\text{gentive})$
    - $I(\text{instrumental})$
    - $D(\text{ative})$
    - $F(\text{actitive})$
    - $L(\text{ocative})$
    - $O(\text{bjective})$

• benefits
  • fewer tokens: fewer verb senses
  • verbs with the same semantics, but with different subject selection preferences
    - *like, please ...*\[\_O+D\]
  • only noun phrases of the same case may be conjoined
    *John and a hammer broke the window.  *The car broke the window with a fender

• drawbacks
  • limited set of cases
  • mapping rules syntax $\rightarrow$ semantics
From Case Theory to FrameNet

University of Berkeley
http://framenet.icsi.berkeley.edu/

• on-line lexical resource for English
• verbs, nouns, adjectives, prepositions
• the aim:
  to document the range of semantic and syntactic combinatory possibilities - valences - of each word in each of its senses
• FrameNet data:
  more than 13,000 lexical units,
  more than 1,200 hierarchically-related semantic frames exemplified in more than 200,000 annotated sentences
From Case Theory to FrameNet

• Frame semantics
  semantic frames and its frame elements
• Semantic frames are evoked by lexical units

• Semantic frame:
  a conceptual structure that describes a particular type of situation, object, or event
• Frame elements:
  components of semantic frames, semantic roles
• each word’s valence possibilities,
  the ways in which information about the frames is provided in the linguistic structures
FrameNet Data

FrameNet data is available online as browsable reports, a clickable visualization, and a searchable database. You can also download the data in XML format (OWL format is also available in the current data release).

Note that the data shown here is newer than the data in the data release, but has not been as carefully checked for completeness and consistency.

Full Text Annotation

In addition to our ongoing lexicographic work, FrameNet has begun to annotate some continuous texts, as a demonstration of how frame semantics can contribute to text understanding. This style of annotation typically involves marking frame elements of frames evoked by multiple predicators in each sentence or even in each clause.

Frame Index

Frame definitions, semantic roles/frame elements (FEs), and other frame information.

Lexical Unit Index

Word senses (Lexical Units) with annotation and related syntactic patterns report.
Communication

Definition:

A **Communicator** conveys a **Message** to an **Addressee**; the **Topic** and **Medium** of the communication also may be expressed. This frame includes no specification of the method of communication (speech, writing, gesture, etc.). The frames that inherit the general Communication frame can add elaboration to the **Medium** in a variety of ways (*in French, on the radio program, in a letter*) or to the **Manner** of communication (*babble, rant, shout, whisper*). There are also frames that either do not inherit all of the FEs of this frame or do not inherit them in a straightforward manner (such as Conversation, in which **Communicator** and **Addressee** alternate roles, and are often expressed by a single, plural NP).

Let’s hope it didn’t have time to **COMMUNICATE** any of its findings to its overlords.
Communication

FEs:

Core:

**Communicator [Com]**
Semantic Type: Sentient

The sentient entity that uses language in the written or spoken modality to convey a Message to another person.

He finds it hard to **COMMUNICATE** with people, not least his separated parents.

**Medium [Medium]**

The physical or abstract setting in which the Message is conveyed.

Opinions are usually **COMMUNICATED** **over the telephone** and are often given within 24 hours.

**Message [Msg]**
Semantic Type: Message

**Message** A proposition or set of propositions that the **Communicator** wants the **Addressee** to believe or take for granted.

How do you **COMMUNICATE** to them that you really like them?

**Topic [Top]**

Topic is the entity that the proposition or propositions conveyed relate to, that they are about.

Had someone **COMMUNICATED** to the capital **about the flagrant disregard of the religious law**?

Non-Core:

**Addressee [Add]**
Semantic Type: Sentient

The **Addressee** receives a **Message** from the **Communicator**.

The company must be able to **COMMUNICATE** to **potential customers** the way in which its product would satisfy their needs, and provide competitive value.

and others …

PDT: valency III
Communication

Frame-frame Relations:

Inherits From:
Is Inherited By: Communication_manner, Communication_noise, Communication_response, Gesture, Reassuring, Statement, Summarizing
Subframe of:
Has Subframes:
Precedes:
Is Preceded by:
Uses: Information, Topic
Is Used By: Attempt_suasion, Bail_setting, Candidness, Claim_ownership, Commitment, Communication_means, Contacting, Convey_importance, Deny_permission, Discussion, Encoding, Entering_of_plea, Expressing_publicly, Grant_permission, Hear, Justifying, Name_conferral, Notification_of_charges, Prevarication, Questioning, Reasoning, Reporting, Request, Ruling_legally, Sentencing, Speak_on_topic, Suasion, Successfully_communicate_message, Text_creation, Verdict, Volubility
Perspective on:
Is perspectiveized in: Mention
Is Causative of:
See Also:

Lexical Units

codeword.n, communicate.v, communication_{(act)}.n, communication_{(entity)}.n, contact.n, convey.v, indicate.v, password.n, signal.v, speech.n

lexical unit (LU) … a pairing of a word with a sense; roughly, a single word in a single meaning (Cruse 1986),
From Case Theory to **FrameNet**

example: convey … 3 lexical units
communicate.v

<table>
<thead>
<tr>
<th>Frame Elements</th>
<th>Core Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addressee</td>
<td>Core</td>
</tr>
<tr>
<td>Amount of information</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Communicator</td>
<td>Core</td>
</tr>
<tr>
<td>Depictive</td>
<td>Extra-Thematic</td>
</tr>
<tr>
<td>Duration</td>
<td>Extra-Thematic</td>
</tr>
<tr>
<td>Manner</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Means</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Medium</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Message</td>
<td>Core</td>
</tr>
<tr>
<td>Time</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Topic</td>
<td>Core</td>
</tr>
</tbody>
</table>

- T-W ith-(1)
  1. I have been **COMMUNICATING** with the Minister **since 1988** on that problem.
  2. Nerves that **COMMUNICATE** with the use of acetylcholine are said to be cholinergic and are found in the peripheral and central nervous systems.
  3. With the facility of computers, it is possible, as shown by other unions as well as forward-looking branches of our own, to **COMMUNICATE** with members on a personal basis.
  4. Let him **COMMUNICATE** with her by letter, if he must: let solicitors arrange money matters.
  5. By 1928 when the vote was granted to women over 21, the two sisters had ceased to **COMMUNICATE** with each other -- their ideas and lifestyles were now poles apart.
  6. EDI has become extremely important in international freight, enabling shippers, forwarders, carriers and HM Customs and Excise to **COMMUNICATE** with each other using a set of standard messages.

- T-NP information, message, idea-(1)
  1. Teletext uses a broadcast signal to **COMMUNICATE** information to specially adapted television sets, while Prestel uses a telephone link to allow access to information.
  2. Central to the overall strategy is the ability to **COMMUNICATE** information about individual patients and their care throughout the NHS.
Communication

relation of inheritance: semantic frames

Statement → Summarizing → Reassuring → Gesture → Communication_noise → Communication_manner → Communication_response → Response

10 children total
Communication

relation of inheritance: semantic elements

PDT: valency III
From Case Theory to FrameNet

• Principals: Charles J. Fillmore, Collin F. Baker
• Senior Linguist: Michael Ellsworth
• System Analyst: Jisup Hong
• FrameSQL Designer: Hiroaki Sato (Senshu University)
• Lexicographic Consultant: Sue Atkins
• Lexicographers, Annotators and Programmers:
  John Keesling, Albert Kong, Russell Lee-Goldman
• number of other participants and affiliates

(Fillmore, Baker et al., 1998; Ruppenhofer et al., 2010)
PropBank vs. FrameNet

FRAMENET ANNOTATION:

[Buyer Chuck] bought [Goods a car] [Seller from Jerry] [Payment for $1000].
[Seller Jerry] sold [Goods a car] [Buyer to Chuck] [Payment for $1000].

PROPBNANK ANNOTATION:

[Arg0 Chuck] bought [Arg1 a car] [Arg2 from Jerry] [Arg3 for $1000].
[Arg0 Jerry] sold [Arg1 a car] [Arg2 to Chuck] [Arg3 for $1000].
References

- Svozilová et al. (1997) *Slovesa pro praxi*. Praha, Academia
- Svozilová et al. (2005) *Slovník slovesných, substantivních a adjektivních vazeb a spojení*
Valency in Other Lexical Resources: Czech

- **VALLEX**
- most frequent Czech verbs
- valency frames ~ meaning
  - aspectual counterparts
  - diatheses
  - reflexivity and reciprocity
  - semantic classes

http://ufal.mff.cuni.cz/vallex/
Valency in Other Lexical Resources: Czech

• **VALLEX 3.0**
  • the lexicon structure
  • lexeme as an abstract unit consisting of
    • a 'formal' part … a list of word forms represented by (one or more) lemmas
    • a 'functional' part … a list of lexical units

• a single lexeme for aspectual counterparts

- lemma: *write*
- paradigm: {*write, writes, writing, written, wrote*}

- gloss: *to make letters/numbers on a surface*
- syntax: *sb writes st (with st)*

- gloss: *to send a message via letter*
- syntax: *sb writes st to sb (about st)*

...
Valency in Other Lexical Resources: Czech

- **VALLEX 3.0**
  - 2,722 lexemes, 6,711 lexical units
  - 4,586 verbs, 10,821 meanings
  - coverage 98% of verbal occurrences in the Czech National Corpus (SYN2000)
Valency in Other Lexical Resources: Czech

- **VerbaLex** … (Hlaváčková, Horák, 2006)
Valency in Other Lexical Resources

- Czech Syntactic Lexicon (Skoumalová, 2001)
- Slovesa pro praxi (Svozilová et al., 1997)
- Slovník slovesných, substantivních a adjektivních vazeb a spojení (Svozilová et al., 2005)

- Pattern Dictionary of English Verbs (PDEV) (Hanks, Pustejovský)
  - based on Corpus Pattern Analysis (CPE)
  - prototypes and patterns
  - exploitation
  - *semantic types* and *semantic roles*

Cognitive concepts such as Human, Institution, Animal, Event, etc.

Intrinsic property of nouns normally found in that argument slot, e.g. judge, criminal, punishment for to sentence
References

• Šindlerová, J., Toman, J., Cinková, S., Semecký, J. (2007) EngVallex 1.0. UFAL MFF UK
• Lopatková et al. (2008) Valenční slovník českých sloves. Praha, Karolinum
• Svozilová et al. (1997) Slovesa pro praxi. Praha, Academia
• Svozilová et al. (2005) Slovník slovesných, substantivních a adjektivních vazeb a spojení