Spoken Language Understanding
Spoken Language Understanding

- Spoken language understanding is to map natural language speech to frame structure encoding of its meanings.

- What’s difference between NLU and SLU?
  - **Robustness**: noise and ungrammatical spoken language
  - **Domain-dependent**: further deep-level semantics (e.g. Person vs. Cast)
  - **Dialog**: dialog history dependent and utt. by utt. analysis

- Traditional approaches; natural language to SQL conversion

A typical ATIS system (from [Wang et al., 2005])
Semantic Representation

- **Semantic frame** (frame and slot/value structure) [Gildea and Jurafsky, 2002]
  - An intermediate semantic representation to serve as the interface between user and dialog system
  - Each frame contains several typed components called *slots*. The type of a slot specifies what kind of fillers it is expecting.

  “Show me flights from Seattle to Boston”

  XML format (left) and hierarchical representation (right) [Wang et al., 2005]
Semantic Representation

• Two common components in semantic frame
  ► **Dialog acts (DA)**; the meaning of an utterance. At the discourse level, and it is approximately the equivalent of *intent* or *subject slot* in the practice.
  ► **Named entities (NE)**; the identifier of entity such as person, location, organization, or time. In SLU, it represents domain-specific meaning of a word (or word group).

• Example (ATIS and EPG domain, simplified representation)

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Dialog Act</th>
<th>From Location City Name</th>
<th>To Location City Name</th>
<th>Month Name</th>
<th>Day Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show me flights from Denver to New York on Nov. 18th</td>
<td>Flight</td>
<td>Denver</td>
<td>New York</td>
<td>Nov.</td>
<td>18th</td>
</tr>
<tr>
<td>I want to watch LOST</td>
<td>Search Program</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DIALOG_ACT = Program
PROGRAM = LOST
## Semantic Frame Extraction

• **Semantic Frame Extraction** (~ *Information Extraction Approach*)
  1) Dialog act / Main action Identification ~ *Classification*
  2) Frame-Slot Object Extraction ~ *Named Entity Recognition*
  3) Object-Attribute Attachment ~ *Relation Extraction*
  
  ▶ 1) + 2) + 3) ~ *Unification*

---

**Examples of semantic frame structure**

<table>
<thead>
<tr>
<th>롯데월드에 어떻게 가나요?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain: Navigation</td>
</tr>
<tr>
<td>Dialog Act: WH-question</td>
</tr>
<tr>
<td>Main Action: Search</td>
</tr>
<tr>
<td>Object.Location.Destination=롯데월드</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>난 롯데월드가 너무 좋아.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain: Chat</td>
</tr>
<tr>
<td>Dialog Act: Statement</td>
</tr>
<tr>
<td>Main Action: Like</td>
</tr>
<tr>
<td>Object.Location=롯데월드</td>
</tr>
</tbody>
</table>

---

*Overall architecture for semantic analyzer*
Knowledge-based Systems

• Knowledge-based systems:
  ► Developers write a syntactic/semantic grammar
  ► A robust parser analyzes the input text with the grammar
  ► Without a large amount of training data

• Previous works
  ► MIT: TINA (natural language understanding) [Seneff, 1992]
  ► CMU: PHEONIX [Pellom et al., 1999]
  ► SRI: GEMINI [Dowding et al., 1993]

• Disadvantages
  1) Grammar development is an error-prone process
  2) It takes multiple rounds to fine-tune a grammar
  3) Combined linguistic and engineering expertise is required to construct a grammar with good coverage and optimized performance
  4) Such a grammar is difficult and expensive to maintain
Statistical Systems

- Statistical SLU approaches:
  - System can *automatically learn* from example sentences with their corresponding semantics
  - The annotation are much easier to create and do not require specialized knowledge

- Previous works
  - Microsoft: HMM/CFG composite model [Wang et al., 2005]
  - AT&T: CHRONUS (Finite-state transducers) [Levin and Pieraccini, 1995]
  - Cambridge Univ: Hidden vector state model [He and Young, 2005]
  - Postech: Semantic frame extraction using statistical classifiers [Eun et al., 2004; Eun et al., 2005; Jeong and Lee, 2006]

- Disadvantages
  1) Data-sparse problem; system requires a large amount of corpus
  2) Lack of domain knowledge
Machine Learning for SLU

• Relational Learning (RL) or Structured Prediction (SP) [Dietterich, 2002; Sutton and McCallum, 2004]
  ▶ Structured or relational patterns are important because they can be exploited to improve the prediction accuracy of our classifier
  ▶ Argmax search (e.g. Sum-Max, Belief propagation, Viterbi etc)

\[ y^* = \arg \max_y P(y|x) \]

• Basically, RL for language processing is to use a left-to-right structure (a.k.a linear-chain or sequence structure)
• Algorithms: **CRFs**, Max-Margin Markov Net (**M3N**), SVM for Independent and Structured Output (**SVM-ISO**), Structured Perceptron, etc.
Machine Learning for SLU

- Maximum Entropy (a.k.a logistic regression)
  - Conditional and discriminative manner
  - Unstructured! (no dependency in $y$)
  - *Dialog act* classification problem
    $$p(z|x) = \frac{1}{Z_z(x)} \exp \left( \sum_k \nu_k h_k(z, x) \right)$$

- Conditional Random Fields [Lafferty et al. 2001]
  - Structured versions of MaxEnt (argmax search in inference)
  - Undirected graphical models
  - Popular in language and text processing
  - Linear-chain structure for practical implementation
  - *Named entity* recognition problem
    $$p(y|x) = \frac{1}{Z_y(x)} \exp \left( \sum_{t=1}^{T} \sum_k \lambda_k f_k(y_{t-1}, y_t) + \mu_k g_k(y_t, x) \right)$$
Semantic NER as Sequence Labeling

- Relational Learning for language processing
  - Left-to-right $n$-th order Markov model (linear-chain or sequence)
  - E.g. Part-of-speech tagging, Noun Phrase chunking, Information Extraction, Speech Recognition, etc.
  - Very large size of feature space (e.g. state-of-the-art NP chunking > 1M)
  - Open problem is how to reduce the training cost (even 1st order Markov)
- Transformation to BIO representation [Ramshaw and Marcus, 1995]
  - Begin of entity, Inside of entity, and Outside

<table>
<thead>
<tr>
<th>Show</th>
<th>me</th>
<th>flight</th>
<th>from</th>
<th>Denver</th>
<th>to</th>
<th>New</th>
<th>York</th>
<th>on</th>
<th>Nov.</th>
<th>18th</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>F.CITY-B</td>
<td>O</td>
<td>T.CITY-B</td>
<td>T.CITY-I</td>
<td>O</td>
<td>MONTH-B</td>
<td>DAY-B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F.CITY</td>
<td></td>
<td>T.CITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Long-distance Dependency Problem

- Most practical NLP models employ the local feature set
  - Local context feature (sliding window)
  - E.g.) for “dec.”; current = dec., prev-2 = on, prev-1 = chicago, next+1 = 10th, next+2 = 1999, POS-tag = NN, chunk = PP
  - However, there is exactly same feature set for two states “dec.” (even different labels)
- Non-local feature or high-order dependency should be considered
Using Non-local Information

• Syntactic parser-based approach
  ▶ Parse tree path, governing category, or head word
  ▶ Adv: global structure of language
  ▶ Disadv: informal language – *speech, email*

• Data-driven approach
  ▶ Identical word, lexical co-occurrence, or *triggering*
  ▶ Adv: easy to extract, fit to ungrammatical form
  ▶ Disadv: depends on data set
Hidden Vector State Model

- CFG-like Model [He and Young, 2005]
  - Extends ‘flat-concept’ HMM model
  - Represents hierarchical structure (right-branching) using hidden state vectors
  - Each state expanded to encode stack of a push down automaton

I want to return to Chicago on Dec.

DUMMY RETURN TOLOC CITY MONTH

DUMMY RETURN TOLOC CITY ON MONTH

PUSH(DUMMY) POP(DUMMY) PUSH(RETURN) PUSH(TOLOC) PUSH(CITY) POP(CITY) POP(TOLOC) PUSH(ON) PUSH(MONTH)
Trigger Feature Selection

- Using Concurrence Information [Jeong and Lee, 2006]
  - Finding non-local and long-distance dependency
  - Based on a feature selection method for exponential models

**Definition 1 (Trigger Pair)** Let two elements $a \in A$ and $b \in B$, $A$ is a set of history features and $B$ is a set of target features in training examples. If a feature element $a$ is significantly correlated with another element $b$, then $a \rightarrow b$ is considered as a trigger, with $a$ being the trigger element and $b$ the trigger ed element.

I want to return to chicago on dec.
Joint Prediction for SLU
Motivation: Independent System

- DA and NE tasks are totally independent.
  - System produces the DA (z) and NE (y) given input words (x), and passes them to DM.
  - MaxEnt for DA classification, CRFs for NE recognition
- Preliminary result in previous section
Motivation: Cascaded System

• Current state-of-the-art system design [Gupta et al. 2006, AT&T system]
  ► Training the NE module, and use its prediction as a feature for the DA module (or vice versa)
  ► Significant drawback; cannot take advantage of information from DA task (or vice versa)
  ► Cascaded system can improve only single task performance rather than both.

![Diagram of Spoken Language Understanding System](attachment:Spoken_Language_Understanding_System_Diagram.png)
Motivation: Joint System

- Joint Prediction of DA and NE [Jeong and Lee, 2006]
  - DA and NE are mutually dependent
  - An integration of DA and NE model $\rightarrow$ encoding inter-dependency
  - GOAL is to improve both performances of DA and NE task.

![Diagram of joint system]

Automatic Speech Recognition $\xrightarrow{x}$ Joint Inference $\xrightarrow{x,y,z}$ Dialog Management

Joint Inference

- Sequence Labeling (Named Entity / Frame Slot)
- Classification (Dialog Act / Intent)

Joint Model (e.g. TriCRFs)
Triangular-chain CRFs

- Modeling the inter-dependency \((y \leftrightarrow z)\)

\[
p(y, z | x; \Theta) = \frac{1}{Z(x)} \exp(\phi(x, y, z))
\]

- Factorizing the potential

\[
\phi(x, y, z) = \sum_t \sum_k \{ \lambda_k f_k(y_{t-1}, y_t, z) + \mu_k g_k(y_t, x) \} + \sum_k \nu_k h_k(z, x)
\]

- \(f_k\) can be a function of triangulated cliques.

However, we assume that NE state transition is independent from DA, i.e., DA operates as an observation feature to identify NE labels.
CRFs Family

Graphical illustrations of linear-chain CRF (left), factorial CRF [Sutton et al. 2004] (middle) and triangular-chain CRF (right)
Reducing the Human Effort
Reducing the Effort of Human Annotation

- The goal is to reduce the labeling effort for spoken language understanding
  - Preparation of human-labeled utterances is labor intensive and time consuming
- Supervised learning
  - Requires the large amount of labeled data
  - Given the data \((x_1, y_1), \ldots, (x_n, y_n) \equiv \{x_i, y_i\}_{1}^{n}\)
  - We find a function \(f : X \rightarrow Y\)
  - \(f\) can be any classifiers (e.g. MaxEnt, SVM, Boosting, Decision tree, etc.)
Reducing the Effort of Human Annotation

- **Active learning**
  - *learning algorithm to query the label of any instance (included into labeled data set)*
  - Artificial membership queries (Cohn et al. 1994)
  - Text categorization (Lewis and Carlett, 1994)
  - Support vector machine (Schohn and Cohn, 2000; Tong and Koller, 2001)
  - Natural language parsing and information extraction (Thompson et al., 1999; Tang et al., 2002), Word segmentation (Sassano, 2002)
  - Spoken language understanding (Tur et al., 2002)

- **Semi-supervised learning**
  - Co-training (Blum and Mitchell, 1998)
  - Co-EM (Nigam and Ghani, 2000), Co-EM with ECOC (Ghani, 2002)
  - Natural language call routing (Iyer et al. 2002)

- **Combining two techniques**
  - Text categorization (McCallum and Nigam, 1998)
  - Speech recognition (Riccardi and Hakkani-Tur, 2003)
Active Learning

- Certainty-based method
  - Predict the candidate raw data
  - Estimate confidence (e.g. Prob) and use it

\[ S_{active} = \{ s_i : P(y_i|x_i) < threshold_{active} \} \]
Semi-supervised Learning

- Augmenting the machine-labeled data
  \[ S_{semi} = \{ s_i : P(y_i|x_i) > threshold_{semi} \} \]

- Augmenting the classification model
  - Like as adaptation (model interpolation), classifier dependent method
Combining Two Methods

- Use whole raw data, and divide them into two sets $S_{raw} = S_{active} + S_{semi}$

$$S_{active} = \{ s_i : P(y_i|x_i) < \text{threshold} \}$$

$$S_{semi} = \{ s_i : P(y_i|x_i) > \text{threshold} \}$$
To General Understanding: Semantic Role Labeling
Semantic Role Labeling (SRL)

- More general natural language understanding tasks
- The process of assigning a WHO did WHAT to WHOM, WHEN, WHERE, WHY, HOW etc. structure to plain text
- **Natural Language Understanding**: domain-specific / hand-crafted → domain-independent / machine learning

**Examples**

- [Judge She] **blames** [Evaluee the Government] [Reason for failing to do enough to help]. (JUDGEMENT)
- [Message “I’ll knock on your door at quarter to six”] [Speaker Susan] **said**. (STATEMENT)

**Applications**

- information extraction, question-answering, spoken dialogue system, machine translation, text summarization, parsing etc.
Semantic Role Labeling (SRL)

- SRL = Domain-independent shallow semantic parsing
- There is not always a direct mapping between syntax and semantics.
  - Verb-specific Role (FrameNet)
    - **CONVERSATION** including
      - verb: argue, banter, debate, converse, gossip
      - noun: dispute, discussion
  - Thematic Role (FrameNet & PropBank)
    - tend to be mainly verb
# Two Projects

<table>
<thead>
<tr>
<th></th>
<th>FrameNet</th>
<th>PropBank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target Corpus</strong></td>
<td>British National Corpus</td>
<td>Wall Street Journal</td>
</tr>
<tr>
<td><strong>Labels</strong></td>
<td>10 general abstract thematic roles + the thousands of potential verb-specific roles</td>
<td>Predicate independent labels ARG0~5, ARG-Ms etc</td>
</tr>
<tr>
<td><strong>Training set</strong></td>
<td>50,000 sentence / 100,000 frames</td>
<td>85,000 sentences / 250,000 arguments</td>
</tr>
<tr>
<td><strong>Test set</strong></td>
<td>N/A</td>
<td>5,000 sentences / 12,000 arguments</td>
</tr>
<tr>
<td><strong>Release</strong></td>
<td>Release 1.2, June 14, 2005</td>
<td>Feb, 2004 / March, 2005</td>
</tr>
<tr>
<td><strong>Evaluation Task</strong></td>
<td>Senseval task</td>
<td>CoNLL shared task</td>
</tr>
<tr>
<td><strong>Distributor</strong></td>
<td>UC. Berkeley</td>
<td>Univ. Penn / ACE project</td>
</tr>
<tr>
<td><strong>Fee</strong></td>
<td>Free for Academy</td>
<td>Free for CoNLL</td>
</tr>
</tbody>
</table>
CoNLL 04/05 Shared Tasks

• Open and Competitive Task

• Dataset
  - Released on March 4th, 2005
  - WSJ sections 02-21 as training set
  - WSJ section 24 as development set
  - WSJ section 23 + fresh data as test set
  - POS, Chunk, Collins/Charniak parse tree, NE tag
  - Special labels
    - R-*: a reference to some others (e.g. that)
    - C-*: a continuation phrase

• Example
  - \([A0\ He\ ]\ [AM-MOD\ would\ ]\ [AM-NEG\ n't\ ]\ [V\ accept\ ]\ [A1\ anything\ of\ value\ ]\ from\ [A2\ those\ he\ was\ writing\ about\ ].\)
Sample domains and frames from the FrameNet lexicon
## FrameNet - Abstract Roles

Table 17
Abstract semantic roles, with representative examples from the FrameNet corpus.

<table>
<thead>
<tr>
<th>Role</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGENT</td>
<td><em>Henry pushed the door open and went in.</em></td>
</tr>
<tr>
<td>CAUSE</td>
<td><em>Jeez, that amazes me as well as riles me.</em></td>
</tr>
<tr>
<td>DEGREE</td>
<td><em>I rather deplore the recent manifestation of Pop; it doesn’t seem to me to have the intellectual force of the art of the Sixties.</em></td>
</tr>
<tr>
<td>EXPERIENCER</td>
<td><em>It may even have been that John anticipating his imminent doom ratified some such arrangement perhaps in the ceremony at the Jordan.</em></td>
</tr>
<tr>
<td>FORCE</td>
<td><em>If this is the case can it be substantiated by evidence from the history of developed societies?</em></td>
</tr>
<tr>
<td>GOAL</td>
<td><em>Distant across the river the towers of the castle rose against the sky straddling the only land approach into Shrewsbury.</em></td>
</tr>
<tr>
<td>INSTRUMENT</td>
<td><em>In the children with colonic contractions fasting motility did not differentiate children with and without constipation.</em></td>
</tr>
<tr>
<td>LOCATION</td>
<td><em>These fleshy appendages are used to detect and taste food amongst the weed and debris on the bottom of a river.</em></td>
</tr>
<tr>
<td>MANNER</td>
<td><em>His brow arched delicately.</em></td>
</tr>
<tr>
<td>NULL</td>
<td><em>Yet while she had no intention of surrendering her home, it would be foolish to let the atmosphere between them become too acrimonious.</em></td>
</tr>
<tr>
<td>PATH</td>
<td><em>The dung-collector ambled slowly over, one eye on Sir John.</em></td>
</tr>
<tr>
<td>PATIENT</td>
<td><em>As soon as a character lays a hand on this item, the skeletal Cleric grips it more tightly.</em></td>
</tr>
<tr>
<td>PERCEPT</td>
<td><em>What is apparent is that this manual is aimed at the non-specialist technician, possibly an embalmer who has good knowledge of some medical procedures.</em></td>
</tr>
<tr>
<td>PROPOSITION</td>
<td><em>It says that rotation of partners does not demonstrate independence.</em></td>
</tr>
<tr>
<td>RESULT</td>
<td><em>All the arrangements for stay-behind agents in north-west Europe collapsed, but Dansey was able to charm most of the governments in exile in London into recruiting spies.</em></td>
</tr>
<tr>
<td>SOURCE</td>
<td><em>He heard the sound of liquid slurping in a metal container as Farrell approached him from behind.</em></td>
</tr>
<tr>
<td>STATE</td>
<td><em>Rex spied out Sam Maggott hollering at all and sundry and making good use of his over-sized red gingham handkerchief.</em></td>
</tr>
<tr>
<td>TOPIC</td>
<td><em>He said, “We would urge people to be aware and be alert with fireworks because your fun might be someone else’s tragedy.”</em></td>
</tr>
</tbody>
</table>
PropBank - Roles

Table I. Argument labels associated with the predicate *operate* (sense: work) in the PropBank corpus.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG0</td>
<td>Agent, operator</td>
</tr>
<tr>
<td>ARG1</td>
<td>Thing operated</td>
</tr>
<tr>
<td>ARG2</td>
<td>Explicit patient (thing operated on)</td>
</tr>
<tr>
<td>ARG3</td>
<td>Explicit argument</td>
</tr>
<tr>
<td>ARG4</td>
<td>Explicit instrument</td>
</tr>
</tbody>
</table>

Table II. List of adjunctive arguments in PropBank – ARGMS

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARGM-LOC</td>
<td>Locative</td>
<td>the museum, in Westborough, Mass.</td>
</tr>
<tr>
<td>ARGM-TMP</td>
<td>Temporal</td>
<td>now, by next summer</td>
</tr>
<tr>
<td>ARGM-MNR</td>
<td>Manner</td>
<td>heavily, clearly, at a rapid rate</td>
</tr>
<tr>
<td>ARGM-DIR</td>
<td>Direction</td>
<td>to market, to Bangkok</td>
</tr>
<tr>
<td>ARGM-CAU</td>
<td>Cause</td>
<td>In response to the ruling</td>
</tr>
<tr>
<td>ARGM-DIS</td>
<td>Discourse</td>
<td>for example, in part, Similarly</td>
</tr>
<tr>
<td>ARGM-EXT</td>
<td>Extent</td>
<td>at $38.375, 50 points</td>
</tr>
<tr>
<td>ARGM-PRP</td>
<td>Purpose</td>
<td>to pay for the plant</td>
</tr>
<tr>
<td>ARGM-NEG</td>
<td>Negation</td>
<td>not, n’t</td>
</tr>
<tr>
<td>ARGM-MOD</td>
<td>Modal</td>
<td>can, might, should, will</td>
</tr>
<tr>
<td>ARGM-REC</td>
<td>Reciproceals</td>
<td>each other</td>
</tr>
<tr>
<td>ARGM-PRD</td>
<td>Secondary Predication</td>
<td>to become a teacher</td>
</tr>
<tr>
<td>ARGM</td>
<td>Bare ArgM</td>
<td>with a police escort</td>
</tr>
<tr>
<td>ARGM-ADV</td>
<td>Adverbials</td>
<td>(none of the above)</td>
</tr>
</tbody>
</table>
PropBank - Example

[ARG0 It] [predicate operates] [ARG1 stores] [ARGM-LOC mostly in Iowa and Nebraska].

Figure 1. Syntax tree for a sentence illustrating the PropBank tags

Figure 2. A sample sentence from the PropBank corpus
Features in SRL

• The relationship between surface manifestations (syntax) and semantic roles = Linking theory (Levin and Hovav 1996)
  ▶ the syntactic realization of arguments of a predicate is predicable from semantics

• Syntactic Structure
  ▶ Using Shallow and Full parsing

• Lexical information
  ▶ Word Statistics / Semantics

• Basic Features provided by CoNLL Task
  ▶ POS, Chunk, Parse tree, NE tag

• Bottleneck = Parse Tree Error
Using Grammatical Function

- Phrase Type (= Chunk)
  - [Speaker We] talked [Topic about the proposal] [Medium over the phone].
- Governing Category: S (subject) and VP (object)
  - “if there is an underlying AGENT, it becomes the syntactic subject.”
- Position: before or after the predicate
  - to overcome parse error
- Voice: active of passive
- Head Word
  - from parse tree output
- Tree Path
**Parse Tree Path**

- Syntactic relation between target word and constituent
- Syntactic path through the parse tree from the parse constituent to the predicate being classified

He → ate :
VB↑VP↑S↓NP

The lawyer → went :
The lawyers ARG0
VBD went to work
predicate NULL ARG4

*FrameNet (Gildea)  *PropBank (Pradhan)