Partial Parsing

- Instead of trying to generate a complete parse tree for a sentence, *partial parsers* generate only fragments representing local syntactic constituents.
- Partial parsers typically try to identify NPs, VPs, and PPs, and sometimes Proper Noun phrases.
- These local syntactic constituents can be identified (relatively) reliably using simple grammar rules and heuristics.
- Many partial parses use finite state machines to recognize a regular grammar.

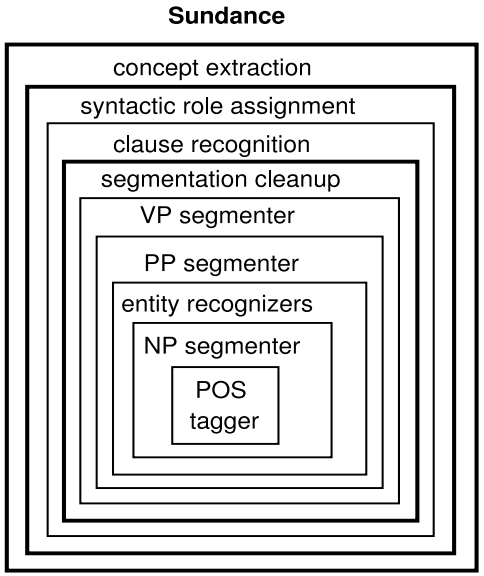
Benefits of Partial Parsing

- Deep syntactic structure may not be important for some NLP applications.
- Some ambiguity issues can be ignored if they are not critical for identifying the fragments.
- Some structural issues can be delayed and left for semantic analysis.
- Partial parsers are more robust with ungrammatical or ill-formed input.
- Partial parsers are usually much faster than full parsers.

The Sundance NLP System

- **Sundance** (Sentence UNDerstanding ANd Concept Extraction) was developed here at the University of Utah.
- Sundance is a conceptual sentence analyzer that supports:
  - obsessive modularity
  - variable depth processing
  - partial parsing
  - multi-pass heuristic constraint propagation
  - conceptual analysis for information extraction

The Sundance Architecture



Heuristic Constraint Propagation

The man shot the bear with a loaded gun.

Sample Sundance Output

```
> The man shot the bear with a loaded gun.
  NP SEGMENT (SUBJ):
    [The (root: the) (LEX)(ART)]
    [man (root: man) (LEX)(N SINGULAR)]

  VP SEGMENT (ACTIVE_VERB):
    [shot (root: shoot) (LEX)(V PAST)]

  NP SEGMENT (DOBJ):
    [the (root: the) (LEX)(ART)]
    [bear (root: bear) (LEX)(N SINGULAR)]

  PP SEGMENT (PREP):
    [with (root: with) (LEX)(PREP)]
    NP SEGMENT:
      [a (root: a) (LEX)(ART)]
      [loaded (root: load) (LEX)(ADJ)]
      [gun (root: gun) (LEX)(N SINGULAR)]
```


Example NP Disambiguation Heuristics

- If previous word is ART and current word can be ADJ/N
  ⇒ disambiguate as ADJ/N
- If previous word is NUMBER and current word can be plural N
  ⇒ disambiguate as plural N
- If previous word is PREP (not 'to') and current word can be N
  ⇒ disambiguate as N
- If previous word is POSS and current word can be N
  ⇒ disambiguate as N
- If previous word is base V and current word can be N
  ⇒ disambiguate as N
- If previous word is singular N and current word can be N or base verb
  ⇒ disambiguate as N

Example VP Disambiguation Heuristics

- If previous word is a form of 'to be' and current word can be a present participle V
  ⇒ disambiguate as V
- If previous word is 'to' and current word can be base V
  ⇒ disambiguate as V
- If next word is PRONOUN and current word can be V
  ⇒ disambiguate as V
- If next word is ART or (only) ADJ and current word can be present participle V
  ⇒ disambiguate as V
- If previous word is 'being' and current word can be past tense V
  ⇒ disambiguate as V

Clause Recognition

- Each sentence needs to be segmented into clauses. Each clause has one main verb, but it may also have infinitives.
- Relative pronouns usually signal the beginning of a new clause.

  Ex: *I bought the book that was #1 on the best-seller list.*

- A second verb may also signal the beginning of a new clause.

  Ex: *I bought a book and went home.*

- The subject of a clause often needs to be inferred from the previous clause.

Subject Inference

```
> Rover barked at the cat, which ran up a tree.
CLAUSE:
  NP SEGMENT (SUBJ):
    [Rover (?)(UNK)]

  VP SEGMENT (ACTIVE_VERB):
    [barked (INF-MOR)(V PAST)]

  PP SEGMENT (PREP):
    [at (root: at) (LEX)(PREP)]
    NP SEGMENT:
      [the (root: the) (LEX)(ART)]
      [cat (?)(UNK)]

  [>COMMA (LEX)(PUNC)]
```


Subject Inference

```
CLAUSE:
  NP SEGMENT (SUBJ):
    [the (root: the) (LEX)(ART)]
    [cat (?)(UNK)]

  VP SEGMENT (ACTIVE_VERB):
    [ran (root: run) (LEX)(V PAST)]

  PP SEGMENT (PREP):
    [up (root: up) (LEX)(PREP)]
    NP SEGMENT:
      [a (root: a) (LEX)(ART)]
      [tree (root: tree) (LEX)(N SINGULAR)]
```


Subject Inference (cont'd)

```
> the man went to the store and shot the clerk.
CLAUSE:
  NP SEGMENT (SUBJ):
    [the (root: the) (LEX)(ART)]
    [man (root: man) (LEX)(N SINGULAR)]

  VP SEGMENT (ACTIVE_VERB):
    [went (root: go) (LEX)(V PAST)]

  PP SEGMENT (PREP):
    [to (root: to) (LEX)(PREP)]
    NP SEGMENT:
      [the (root: the) (LEX)(ART)]
      [store (root: store) (LEX)(N SINGULAR)]

  [and (root: and) (LEX)(CONJ)]
```

Subject Inference (cont'd)

```
CLAUSE:
  NP SEGMENT (SUBJ):
    [the (root: the) (LEX)(ART)]
    [man (root: man) (LEX)(N SINGULAR)]

  VP SEGMENT (ACTIVE_VERB):
    [shot (root: shoot) (LEX)(V PAST)]

  NP SEGMENT (DOBJ):
    [the (root: the) (LEX)(ART)]
    [clerk (?)(UNK)]
```


Weaknesses of Partial Parsing

- Usually does not handle embedded relative clauses well.

  Ex: *I gave the boy that was sick some medicine.*

- Often has trouble recognizing reduced relative clauses.

  Ex: *The woman killed last night was an important diplomat.*

- PP attachment usually not attempted.

Information Extraction

Information extraction involves extracting predefined types of information from text.

Example: extracting companies, products, and locations associated with joint ventures.

> ***Japan Storage Battery Co.*** *announced it has teamed up with a leading French battery maker,* ***Saft S.A.****, to set up a joint venture in* ***Japan*** *to market* ***small batteries****.*
> ***Mitsui and Co.****, a major Japanese trading house, said Tuesday it will form a joint venture in* ***Canada next March*** *to produce* ***ethylene glycol (EG)*** *with* ***a subsidiary of Union Carbide Corp.*** *of the United States.*

Role Relationships

Information extraction (IE) is especially useful at identifying information that depends on *role relationships*.
For example:

- *perpetators and victims* of terrorism
- *buyers and buyees* in corporate acquisitions
- *monetary profits and losses* of business transactions

Dictionaries for Information Extraction

To build an IE system, you need the following resources:

- a part-of-speech dictionary or part-of-speech tagger
- a dictionary of domain-specific extraction patterns
- a dictionary with semantic features for domain-specific words

It also helps to have dictionaries of relevant proper names (e.g., people and location names) but they are not always available.

Extraction Patterns

Japan Storage Battery Co. announced it has teamed up with a leading French battery maker, Saft S.A,, to set up a joint venture in Japan to market small batteries.

<x> has teamed up	→	<x> is a partner company
has teamed up with <y>	→	<y> is a partner company
venture in <z>	→	<z> is location of venture
to market <p>	→	<p> is product of venture

Multi-Slot Extraction Patterns

Caseframe: (active_verb bombed)		
perpetrator	subject	TERRORIST
target	direct-obj	BUILDING
location	pp(in)	LOCATION

Caseframe: (active_verb detonated)		
perpetrator	subject	TERRORIST
instrument	subject	WEAPON
instrument	direct-obj	WEAPON

Arguments can be hard to predict...

Caseframe: (noun attack)		
target	pp(on)	BUILDING VEHICLE
victim	pp(on)	CIVILIAN GOVOFFICIAL
location	pp(on)	LOCATION
date	pp(on)	TIME
target	pp(against)	BUILDING VEHICLE
victim	pp(against)	CIVILIAN
target	pp(at)	BUILDING
location	pp(at)	LOCATION
location	pp(in)	LOCATION
date	pp(in)	TIME

Avoiding PP-attachment

- In a limited domain, PP attachment can often be done implicitly via case frame instantiation.
- The case frames can extract PPs that they are expecting, without having to do explicit PP attachment.

  For example:

  The boy was kidnapped by armed men with pistols.

Sample IE Output

```
> the man was killed by a lunatic with a machinegun.
---------------------------------------------------------
Caseframe KILLED_01(MURDER)
(KILLED; PASSIVE_VERB)

  SUBJ_EXTRACTION:      ``the man'' [CIVILIAN]: 1
---------------------------------------------------------
Caseframe KILLED_04(MURDER)
(KILLED; PASSIVE_VERB_PP)

  PREP_EXTRACTION:      ``a lunatic'' [UNKNOWN]: 1
---------------------------------------------------------
Caseframe KILLED_06(MURDER)
(KILLED; PASSIVE_VERB_PP)

  PREP_EXTRACTION:      ``a machinegun'' [WEAPON]: 1
---------------------------------------------------------
```


Weaknesses of Information Extraction

- Information is extracted piecemeal and often has to be glued back together.

  Ex: *The building was bombed when a grenade was thrown at it.*

  Ex: *The U.S. embassy was bombed. The bombing occurred yesterday afternoon.*

- Coreference resolution is crucial to understand how the pieces fit together.
- Determining when a text or a sentence is relevant can be difficult! Context is critical and metaphor is pervasive.

Strengths of Information Extraction

- Information extraction goes beyond keyword search but is practical for real-world applications.
- Information extraction is best at identifying information that depends on *role relationships*.
- Information extraction tasks are usually domain-specific and therefore more tractable than broad-coverage NLP tasks.
- Information extraction is especially appropriate for long-term interests so the necessary dictionaries can be built once and used repeatedly.
- Many researchers are developing methods to build IE dictionaries automatically or semi-automatically.