5. Resolution of ambiguity

Frank Rudzicz
Toronto Rehabilitation Institute-UHN; and
Department of Computer Science, University of Toronto
Ambiguity resolution

Problem of chart parsing:

**Structural ambiguity:**

*Time flies like an arrow.*

*... paint the office in the building near the research center by the gym ...*

Our parsers, so far, find all possible parses.
Ambiguity resolution

Chart parsing is founded on idea of exploring large space of ambiguities.
  ◦ It can still be slow.
  ◦ It still does not really incorporate semantics.
  ◦ We have to streamline things.

Possible solution: stop at first parse.
  ◦ Problems?
Ambiguities and parsing

Questions:
- Are structural ambiguities really a problem?
- If so, what kinds of ambiguities?

Some real text:

In a general way such speculation is epistemologically relevant, as suggesting how organisms maturing and evolving in the physical environment we know might conceivably end up discoursing of abstract objects as we do.

— Quine

The sentence is unambiguous and we have found the parse. What is the problem?
In the usual way such people think, blithely ignorant as bleating sheep, politicians fulminating and bloviating on their oversized TVs, Americans ignore evidence credibly presented pointing out the results of their choices.
The 168-year-old Sunday tabloid will cease to exist after this week, Murdoch said today in an announcement to staff e-mailed to news organizations. ... Such has been the outcry over the phone hacking of everyday people during times of emotional turmoil that David Cameron’s government on Thursday postponed a decision on News Corp’s bid to purchase full control of BSkyB until September.
Combinatorial explosion of parses

Ordinary sentences can have hundreds of different parses due to combinatorial explosion (Church and Patil).

More than 300 parses for 2% of sentences in corpus.

E.g., 692 parses for:

For each plant give the ratio of 1973 to 1972 figures for each type of production cost and overhead cost.
Global and local ambiguity

**Global ambiguity:** A sentence has multiple interpretations.

I saw the man with the telescope.
Time flies.

- Count which interpretation(s) people prefer.

**Local ambiguity:** Resolved by later input.

- The horse raced...
- Mary expected the woman...
Derived from PoS ambiguity:

*Time flies.* Is ‘flies’ a noun or a verb?

Attachment of one phrase to another:

*examined* the fingerprint *with* the microscope

*the horse in the barn* *that* the vet examined

*learned* that Nadia *arrived* *on* Sunday

*He brought* the car back {undamaged|undismayed}.

Gap ambiguities:

*the boys that the police debated about fighting*

*Who did he tell you that to?*
Syntactic sources of ambiguity

Internal structure of a phrase:
- winter boot sale
- airport long term car park courtesy vehicle pickup point

Alternative analyses of constituent:
- The tourists objected to the guide that they couldn’t hear.
- I want the music box on the table.
What do people do all day? 1

Look at human behaviour:

- Expected / preferred interpretations.
- Clues for successfully pruning parses.

Some human strategies: ...
What do people do all day? 2

**Minimal attachment:**
Prefer the simplest structure.

Karen knew the schedule ...

1 \[ [S [NP [PN Karen]]] [VP [V knew [NP the schedule ...]]] \]
2 \[ [S [NP [PN Karen]]] [VP [V knew [S [NP the schedule ...]]]] \]

Fits 1.

Karen knew the schedule \{by heart | was wrong\}.

Requires 2; hence need to back up; longer processing time.

Recency (local/right association):
Associate new input with most recent part of the parse tree.

Karen met the mother of a singer who ...

1. $[\text{NP the mother} [\text{PP [P of]} [\text{NP a singer} [\text{S who} ...]]]

2. $[\text{NP the mother} [\text{PP [P of]} [\text{NP a singer}]] [\text{S who} ...]]

Notice that this might contradict minimal attachment. When?
Lexical preferences:

Words (especially verbs) may have defaults for their containing or nearby structures.

The tourists {objected | signalled} to the guide that they {couldn’t hear | didn’t like}.

1 Prefer: AGENT object to PATIENT
   (but AGENT object to PATIENT MESSAGE is also possible).

2 Prefer: AGENT signal to PATIENT MESSAGE
   (but AGENT signal to PATIENT is also possible).

Might contradict minimal attachment or recency.
PP attachment ambiguity

Prepositional phrase attachment.
- An example problem that is a focus of much work in disambiguation.
  - A common ambiguity.
  - A specific example of a very general type (modification ambiguity).
  - Representative of properties of many types of ambiguities.
Why is it hard?

Sometimes seems to require complex knowledge of the world:

Optical anisotropy of the copolyester melts can be determined by examination of the materials with the use of an optical microscope.

This is the first examination of the material with the impurity CVL in the region of deeply core shells.

The kinetic advantage arising upon using the NaH/Al mixture to prepare the doped hydride was well reproduced in our examination of the materials with variable dopant amounts and preparation conditions.

---

Many unambiguous cases.

The man with the telescope saw me.
The signals were analyzed with the oscilloscope.
Can often rule out structural possibilities:

- The preposition *of* almost never attaches to a transitive verb.
- Strong constraints on attaching PPs to pronouns and proper names.

He examined it with a microscope.
She examined John with a stethoscope.

But: *I saw {John | him} with a hat.* *(John | He) with a hat saw me.*

*Functioning as an AdjP, not restrictive*
Lexical preferences again

**Lexical preferences:** Words (especially verbs) may have defaults for their containing or nearby structures — *i.e.*, preferred disambiguation.

Examples for PP attachment:

- Preposition $p$ prefers to be attached to be a verb.
- Verb $v$ prefers PPs with preps $p_1$ or $p_2$ or nouns $n_1$ or $n_2$, but dislikes PPs with prep $p_3$ or noun $n_3$.
- A noun $n_1$ in the head of an NP in a PP prefers the PP to be attached to noun $n_2$ or $n_3$, or verb $v_1$ or $v_2$, if one of these is available.
Limitations of lexical preference

Preferences are only preferences:
  ◦ Might not be satisfiable.
  ◦ Might conflict.
  ◦ Might be overridden by coherence, plausibility.

A given attachment problem might have no applicable preferences.
How to use lexical preferences?

If a word $w$ had some preferences …

- How would we know what they are?
- How would we apply them in a parser?
Corpus-based attachment disambiguities

• Gather statistics for lexical usages from a corpus.
• Use statistics to train algorithm, set parameters.
• Apply algorithm to new cases.
Corpora

*Corpus* (*pl.* *corpora*): A large collection of text (or similar material).

- General or specialized content; e.g., news, blog, technical, ESL, errors, ...
- May be (manually or automatically) *annotated*; e.g., with parse, meaning, correction, ...
Some important corpora

Brown Corpus (1M words); British National Corpus (100M words).
  ◦ Tagged with part of speech of each word.


English Gigaword Corpus (~6B words).

Penn Treebank (1.6M sentences of WSJ).
  ◦ Each with complete human-created parse tree.

Canadian Hansard aligned French–English corpus.
Corpus statistics

You can count linguistic phenomena in corpora.
◦ E.g., count how many times a with-PP is noun-attached or verb-attached in Penn Treebank.

Use as data for statistically based methods.

Problems:
◦ Sparse data — even with large corpora.
◦ Required information is not explicit in corpus.
Use corpus statistics to *train* an algorithm — *i.e.*, set parameters.

- Typically output is *classification* of input.
- E.g., classify *(examine, the materials, with the microscope)* as a V-attachment or NP-attachment situation.
- Given input = *(V, NP, PP)*, should PP attach to V or to NP?
Types of training:

- **Supervised**: Learn from data with known answers: From set of pairs \{input, output\}, learn to classify new inputs.
- **Unsupervised**: Given inputs and possible classes only.
- In between:
  
  *Bootstrapping, minimally supervised.*
A three-way partition of corpus data

Training data.

Development (validation, verification) data.
  ◦ To test successive versions of algorithm under development, to guide adjustments to approach.

Test data.
  ◦ For testing of final version of algorithm. (No more tweaking allowed!)