Computational Linguistics

CSC 2501 / 485
Fall 2019

5A

5a. Resolution of ambiguity

Gerald Penn
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 …

- Parser finds all possible parses.

Ambiguity resolution

- Chart parsing is founded on idea of exploring large space of ambiguities.
  - Slow? Not that slow, given a typical median sentence length of < 40.
  - There are simply too many parses on average – it's proven too hard to write grammars for all and only the right readings.
  - Too much work for semantics.
  - Have to narrow down this potential.
- Possible solution: stop at first parse.
  - Problems?

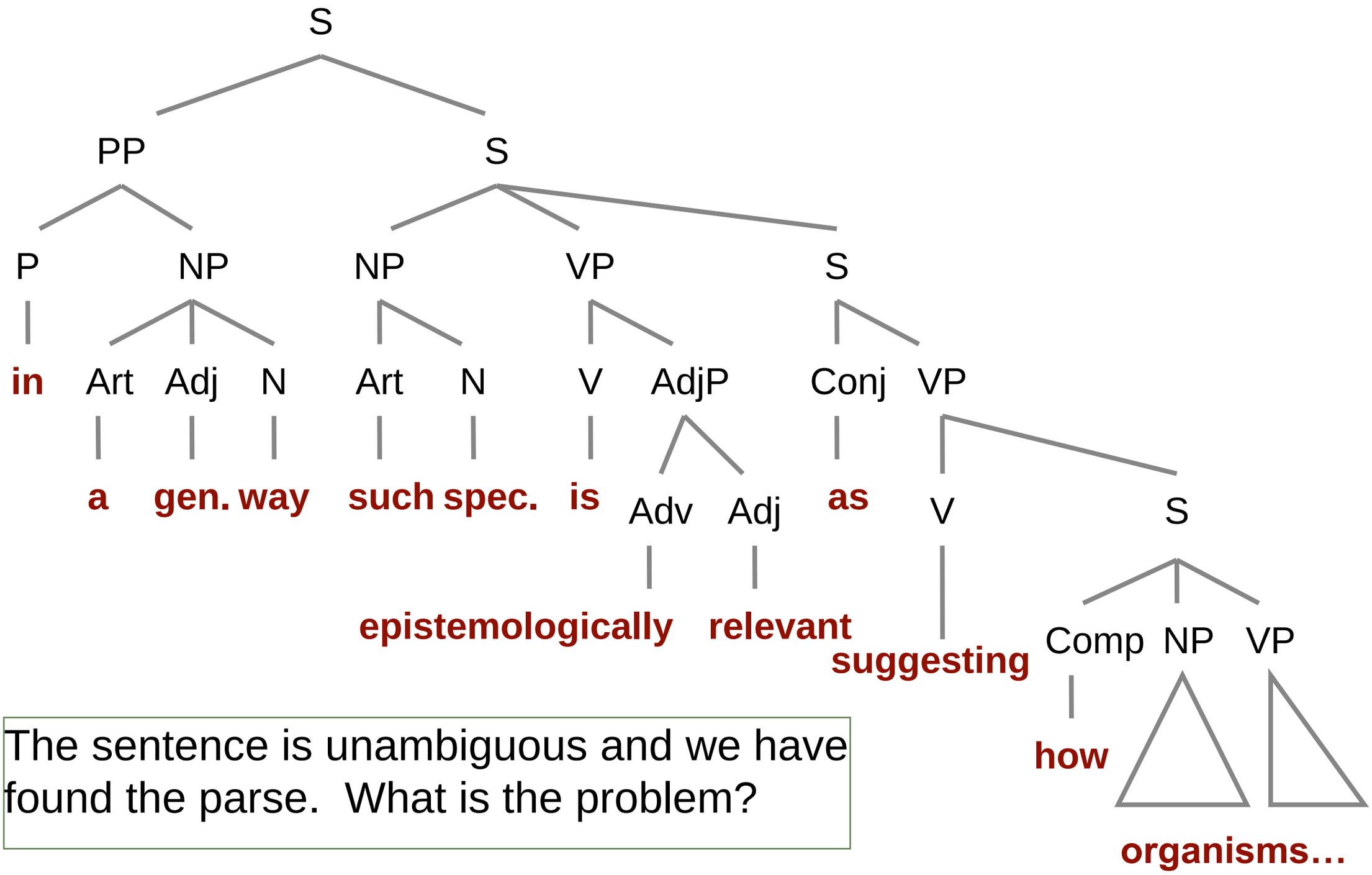
Ambiguities and parsing

- 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

W.V. Quine. "Speaking of objects." *Proceedings and Addresses of the American Philosophical Association*, Vol. 31 (1957–1958), pp. 5–22. Quoted in: Steven Abney, "Statistical methods and linguistics." In: Judith Klavans and Philip Resnik (eds.), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language.* The MIT Press, Cambridge, MA. 1996.

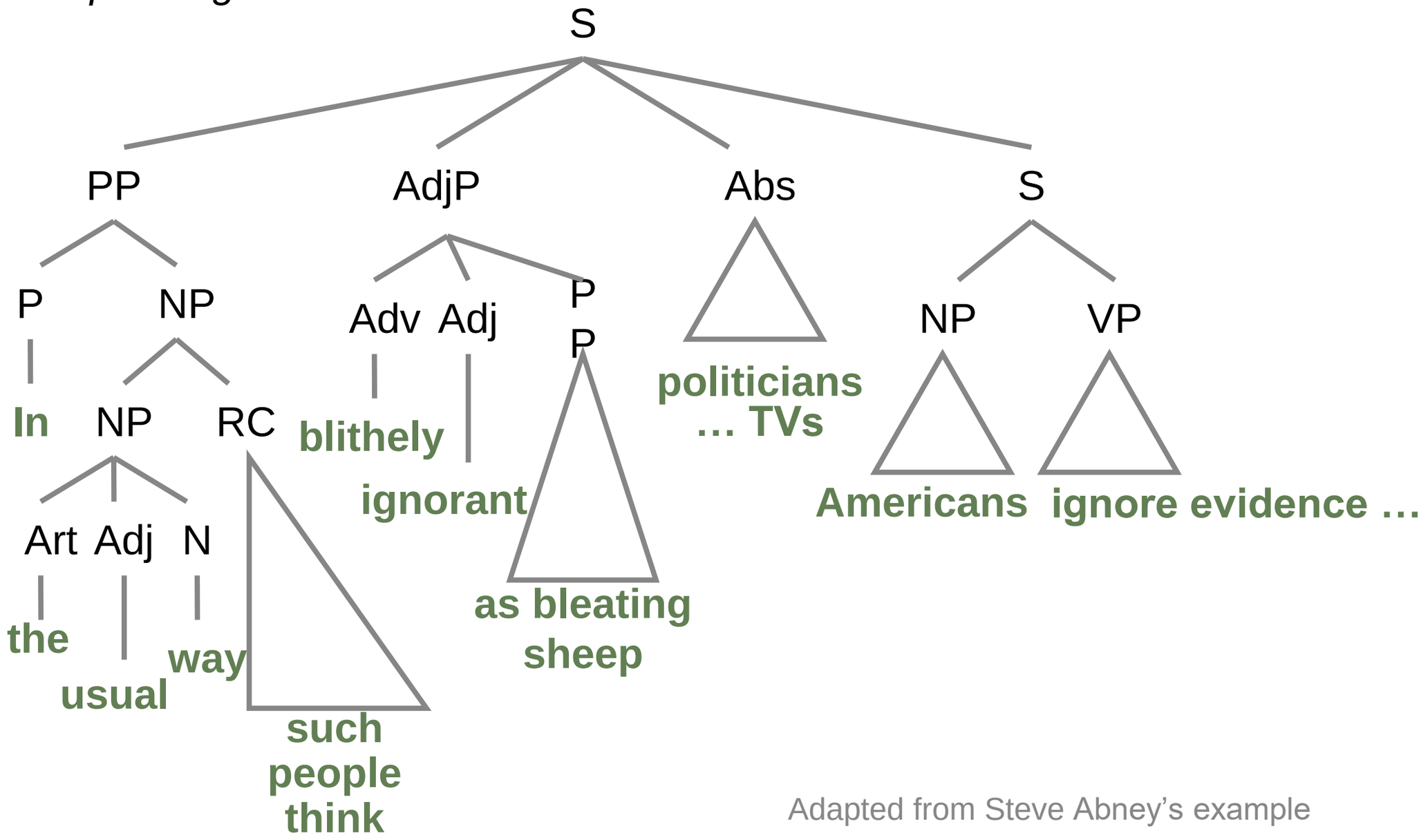
Ambiguities and parsing: Example



The sentence is unambiguous and we have found the parse. What is the problem?

Another example

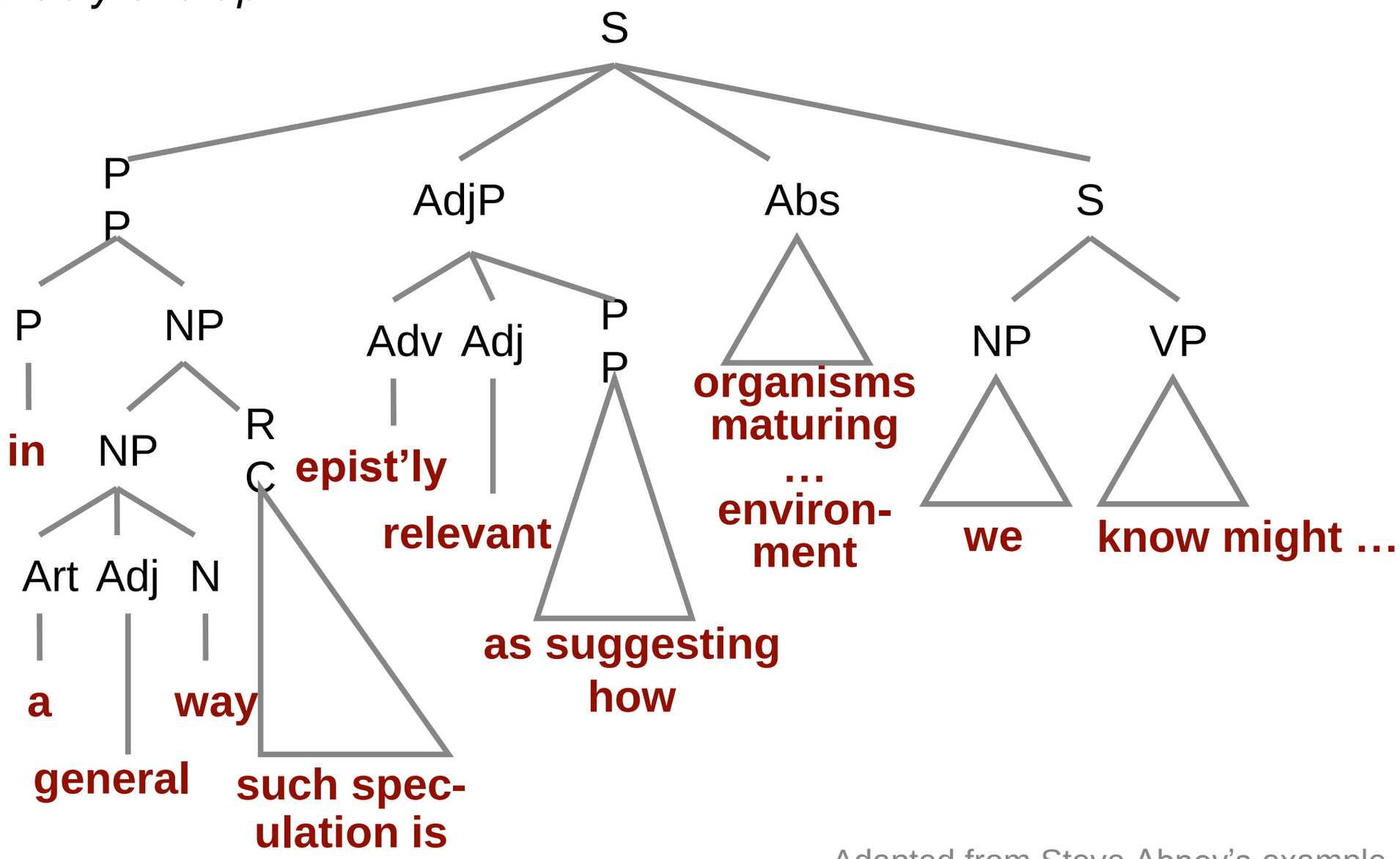
In the usual way such people think, blithely ignorant as bleating sheep, politicians fulminating and bloviating on their oversized TVs, Americans ignore evidence credibly presented pointing out the results of their choices.



Adapted from Steve Abney's example

Another example

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 …



Adapted from Steve Abney's example

Combinatorial explosions of parses

- Ordinary sentences can have hundreds of different parses due to combinatorial explosion (Church and Patil).
- That combinatorial explosion arises to a great extent from the fact that syntactic categories do not – arguably *cannot* – incorporate all of the real-world knowledge that we bring to bear on this problem.

Combinatorial explosions of parses

- 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.

Find the structural ambiguities

The 168-year-old Sunday tabloid will cease to exist after this week, Murdoch said today in an announce-ment 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.

Bloomberg, 8 July 2011

At once too many and too few readings

Listen….

Bloomberg, 8 July 2011

At once too many and too few readings

Listen….

OK, robustness is important, but enough is enough.

Bloomberg, 8 July 2011

Global and local ambiguity

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

  I saw the man with the telescope.
  Time flies.

  - See which interpretation(s) people prefer.

- ***Local ambiguity:*** Resolved by later input.
  - *The horse raced…*
    Mary expected the woman…

Syntactic sources of ambiguity 1

- **Derived from PoS ambiguity:**

  Time flies.

- **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

Syntactic sources of ambiguity 2

- **Internal structure of a phrase:**

  winter boot sale

  airport long term car park courtesy vehicle pickup point

- **Alternative semantic role of subconstituent:**

  The tourists objected to the guide that they couldn't hear.

  I want the music box on the table.

What do people do? 1

- Look at human behaviour:
  - Expected / preferred interpretations.
  - Clues for successfully pruning parses.
- Some human strategies: …

What do people do? 2

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

Karen knew the schedule …

❶ [S [NP [PN *Karen*]] [VP [V *knew* [NP *the schedule …*

❷ [S [NP [PN *Karen*]] [VP [V *knew* [S [NP *the schedule …*

Fits ❶.

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

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

W.D.Marslen-Wilson *et al.* Prosodic effects in minimal attachment. *Quarterly Journal of Experimental Psychology*, 45A(1), 73–87, 1992.

What do people do? 3

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

Karen met the mother of a singer who …

1. [NP *the mother* [PP [P *of*] [NP *a singer* [S *who …*
2. [NP *the mother* [PP [P *of*] [NP *a singer*]] [S *who …*

- Notice that this might contradict minimal attachment. When?

What do people do? 4

- ***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 PP attachment 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.

(1) Brewbaker, James L. and Marshall, William B. Liquid crystalline copolyesters of 4-hydroxybenzoic acid and substituted 4-hydroxybenzoic acids. U.S Patent 5268443.
(2) V. B. Mikhailik. XEOL studies of impurity core-valence luminescence in mixed rubidium caesium chloride crystals. *Journal of Physical Studies*, 9 (2005) 182–184.
(3) P. Wang, X.D. Kanga, H.M. Cheng. Dependence of H-storage performance on preparation conditions in TiF3 doped NaAlH4. *J of Alloys & Compounds*, 2006, 217–22.

When is PP attachment easy (-ier)? 1

- Many unambiguous cases:
  - *The man with the telescope saw me.*
    The signals were analyzed with the oscilloscope.
- Sometimes syntax really should be able to say ‘no.’.

When is PP attachment easy (-ier)? 2

- More often, syntax can say ‘probably not’:
  - The preposition *of* rarely 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 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.
  - When it's the head of an NP in a PP, noun n_1 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 preferences

- 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 disambiguation

- Gather statistics for lexical usages from a ***corpus***.
  - That means the corpus gives the preferred structure
  - That means that the corpus may have to be manually annotated by people (expensive)
- Use statistics to numerically estimate the parameters of a model.
- Apply model 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.
- *Wall Street Journal* Corpus 1987–92 (80M words).
- 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

- Can count linguistic phenomena in corpus.
  - *E.g.,* count how many times a *with*-PP is noun-attached or verb-attached in Penn Treebank.
- Problems:
  - Sparse data — even with large corpora.
  - Required information may not be explicit in corpus.

Corpora with Grammars

- Conventional view: Use these counts to estimate numerical parameters of prior, otherwise discrete-looking grammars.
- Avant-garde view: treat corpora themselves as a means of specifying the grammars.
  - the phrase-structure rules are *grounded* in actual data.
  - the phrase-structure rules are specified in *context*.

Statistical pattern recognition algorithms 1

- Use corpus statistics to ***train*** an algorithm — *i.e.*, set parameters of an underlying model.
  - 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 NP?

Statistical pattern recognition algorithms 2

- Types of training:
  - ***Supervised***: Learn from data with known answers: From set of pairs {*input, output*}, learn to classify new inputs.
  - ***Unsupervised***: Given only inputs and maybe possible outcomes.
  - 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!)