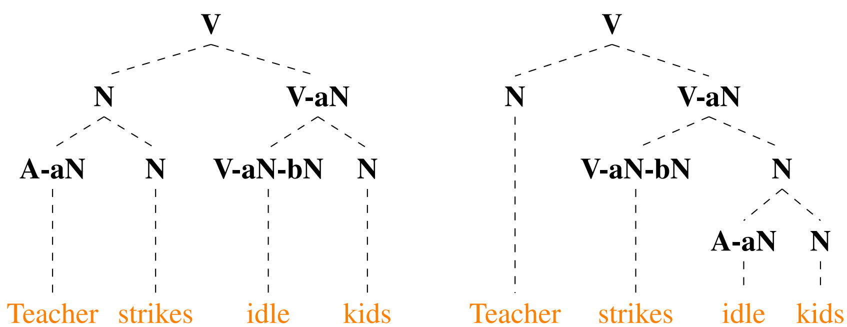
LING5702: Lecture Notes 14
Syntactic Ambiguity

Contents

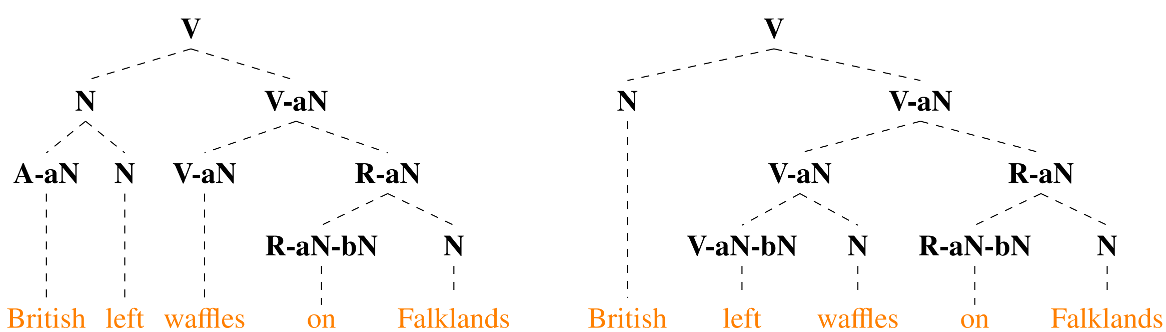
14.1 Syntactic Ambiguity

Ambiguity occurs in syntax too, even though it must propagate for several words:

- (headline) *'Teacher strikes idle kids'*



- (headline) *'British left waffles on Falklands'*



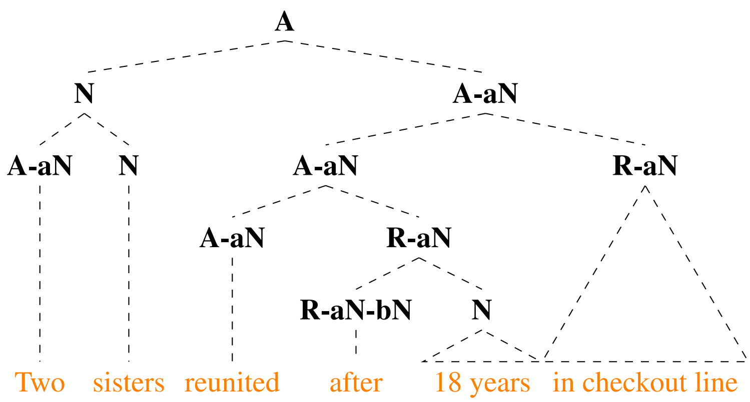
These are examples of ***category*** ambiguity (word can be **N** or **V-aN-bN**).

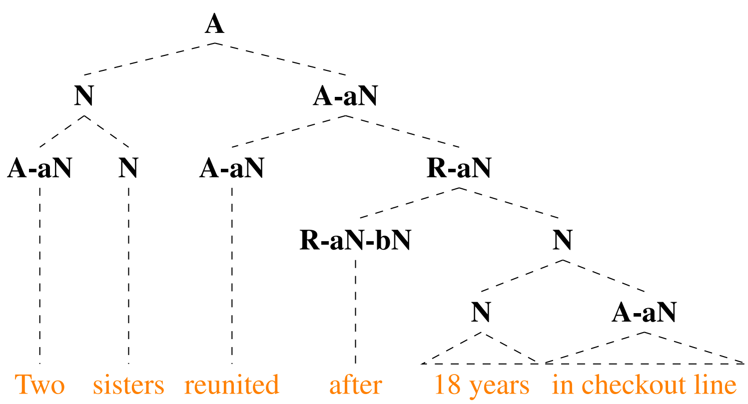
We can also have ***scope*** ambiguity with same categories (more common). . .

14.2 Scope Ambiguity

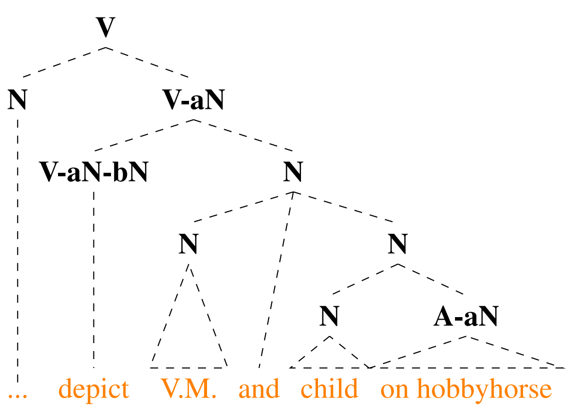
Example of scope ambiguity:

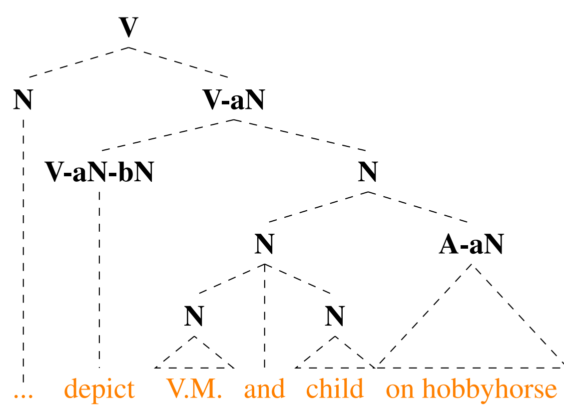
- (headline) *'Two sisters reunited after 18 years in checkout line'*





- (Headline) *'Holiday stamps depict Virgin Mary and child on hobbyhorse'*



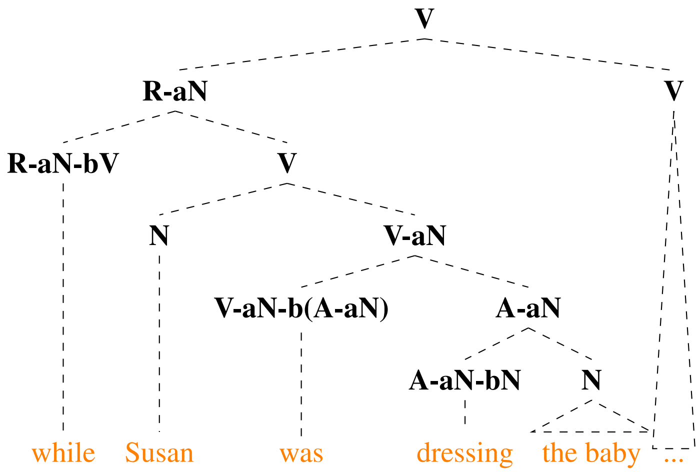


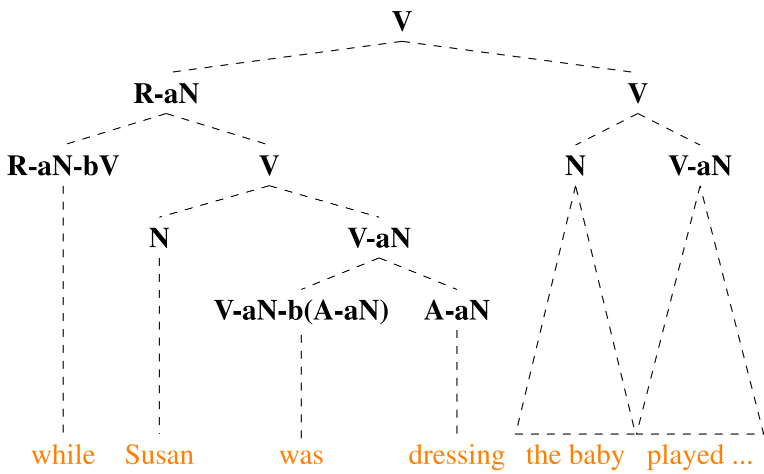
This kind of ambiguity is hard to avoid — we would need 'audible brackets.'

14.3 Local Ambiguity

Ambiguity not always *global* (stays to end), can also be *local* (incremental).

- *'While Susan was dressing the baby ... '* (prior to end of sentence)



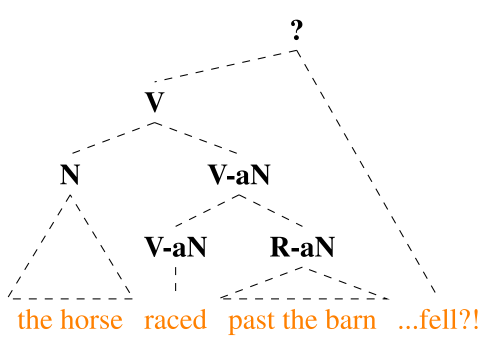


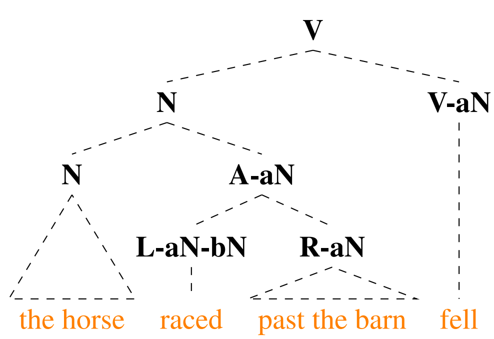
'The baby' has to be reanalyzed as the subject of the main clause.

14.4 Catastrophic Local Ambiguity [Bever, 1970]

Local ambiguity can sometimes bring parsing to a halt!

- (Bever '70) *'The horse raced past the barn fell.'* (trouble at end)





'Raced ...' must be reanalyzed as a passive (reduced relative) modifier.

Often readers don't even see this second possible analysis. Perhaps it comes too late.

14.5 Garden Path Model [Frazier, 1979]

Why does this happen?

[Frazier, 1979]: Garden Path model — a serial processor with reanalysis

- readers maintain a **single** partial parse and **reanalyze** when predicted words are not observed;
- readers examine multiple options at each word, but then **commit** to a single best;
- reanalysis takes time and can cause a complete failure to comprehend.

The Garden Path model also ignores lexical preferences for subcategories.

Phases of the Garden Path parser:

1. a **lexical processor** produces lexical categories (but is **V** always treated as **V-aN-bN**)
2. a **syntactic parser** produces syntactic structure
3. a **thematic (dependency) interpreter** produces sentence meaning

This is justified by results of self-paced reading experiments [Mitchell, 1987]:

- **stimuli:** sentences are resented block by block, with transitive/intransitive verbs:
  (a) *'After the audience had applauded the actors ... (sat down for a drink).'*
  (b) *'After the audience had departed the actors ... (sat down for a drink).'*
- **measure:** self-paced reading time for each segment
- **results:** intransitive sentences (b) are slower, despite lack of ambiguity

Mitchell, Frazier conclude the processor ignores subcategory preferences of verbs.

The Garden Path model makes parse decisions using ***heuristics*** (rules of thumb):

- **late closure**: the parser prefers to keep constituents open, attach low
- **minimal attachment**: the parser prefers to build simpler structure with fewer nodes
  (but you have to providently define verb phrases to have fewer nodes)
- **main assertion**: the parser prefers to modify the main assertion

14.6 Problems with Garden Path Model

[Gorrell, 1991]: Mitchell's sentences had suggestive segmentations, lacked punctuation.

[Trueswell et al., 1993]: eye-tracking experiments show rapid lexical subcategory effects

- **stimuli:** sentences read with eye-tracking headset
  (a) *'The student forgot the solution was in the book.'* (prefer noun phrase complement)
  (b) *'The student hoped the solution was in the book.'* (prefer sentential complement)
- **measure:** eye-tracking fixation positions and durations
- **results:** large delay at *'in'* for verbs preferring nominal (a)

[Trueswell et al., 1994]: eye-tracking experiments show rapid semantic effects

- **stimuli:** sentences read with eye-tracking headset
  (a) *'The defendant examined by the lawyer turned out to be unreliable.'*
  (b) *'The evidence examined by the lawyer turned out to be unreliable.'*
- **measure:** eye-tracking fixation positions and durations
- **results:** large delay at *'by the lawyer'* for animate subject of examine (a)

[Tanenhaus et al., 1995]: eye-tracking experiments show rapid referent effects

- **stimuli:** in context of the following scenes:
  1. one-apple: apple on towel, empty towel, empty box, pencil
  2. two-apple: apple on towel, empty towel, empty box, apple on napkin

  subjects were presented with the following spoken directives:
  (a) unambiguous: *'Put the apple that's on the towel in the box.'*
  (b) ambiguous: *'Put the apple on the towel in the box.'*
- **measure:** location and duration of eye fixations
- **results:** subjects look at empty towel after *'towel'* in (1b)
  subjects look at apple on towel after *'towel'* in (2b)

  In one-apple scene, subjects initially interpret *'on the towel'* as a ***goal***,

  In two-apple scene, subjects initially interpret *'on the towel'* as a ***restriction***.

[Kjelgaard & Speer, 1999]: word-naming experiments show rapid prosody effects

- **stimuli:** manipulated speech; visual request to pronounce *'is'* after *'house'*

(a) cooperative: *'When Roger leaves – the house is dark.'* ('–' = pause)

(b) cooperative: *'When Roger leaves the house – it's dark.'*

(c) uncooperative: *'When Roger leaves the house – is dark.'*

(d) uncooperative: *'When Roger leaves – the house it's dark.'*

- **measure:** duration of word-naming from onset of request
- **results:** delay for uncooperative sentences (c,d)

So we have rapid effects due to several different factors:

1. lexical subcategory (transitive / sentential complement) of verb
2. semantic class (animate / inanimate) of verb subject
3. prosody of verb phrase
4. number of referents in visual scene

These support an interactive (constraint-based) vs. 'modular' model like GP.

- readers constrain structural decisions with subcategories, other factors
- when high probability analyses fail to predict observations, they reallocate probability
- this reallocation causes delays, very large reallocations may cause failure

14.7 Problems for Competition in Constraint-based Models

Constraint-based models usually define delays as a function of competition.

This is problematic because it incorrectly predicts delays during ambiguities.

[van Gompel et al., 2001]:

1. **stimuli:** sentences read with eye-tracking headset

   (a) *'The hunter killed the dangerous leopard with the rifle.'* (attach modifier to verb)

   (b) *'The hunter killed the dangerous leopard with the scars.'* (attach modifier to noun)

   (c) *'The hunter killed the dangerous poacher with the rifle.'* (ambiguous)

2. **measure:** location and duration of eye fixations
3. **results:** ambiguous sentence processed faster

These results argue against either parallelism or competition.

van Gompel et al. throw out both: **race-based model**

- readers choose a single analysis **randomly**, based on multiple interactive factors;
- readers stick with it, then reanalyze if wrong;
- processing is interactive, but not parallel (i.e. propagated along multi-word sequences).

14.8 Surprisal [Hale, 2001, Levy, 2008]

The problem with constraint-based models may be just with the role of competition.

Maybe processing is parallel, but competition is not what causes delays.

John Hale '01: **surprisal** account

- readers pursue **multiple** hypotheses **propagated in parallel**;
- activation of each hypothesis is weighted by the probability it assigns to words;
- probabilities are defined using weights on grammar rules:

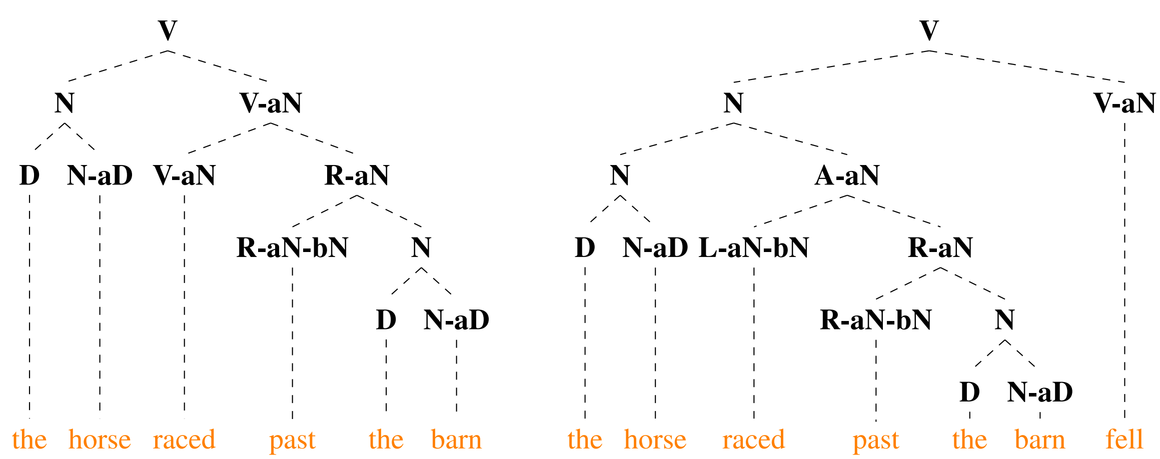
  grammatical rules:

P(**V** → **N V-aN** | **V**) = 1.0	– a sentence may be noun phrase, verb phrase
P(**N** → **D N-aD** | **N**) = .5	– a noun phrase may be determiner, common noun
P(**N** → **N A-aN** | **N**) = .1	– a noun phrase may be a noun phrase, adj phrase
P(**V-aN** → **V-aN R-aN** | **V-aN**) = .05	– a verb phrase may be verb phrase, adv phrase
P(**A-aN** → **L-aN-bN R-aN** | **A-aN**) = .1	– an adj phrase may be trans participial, adv phrase
P(**R-aN** → **R-aN-bN N** | **R-aN**) = 1.0	– an adv phrase may be preposition, noun phrase

  lexical rules:

P(**L-aN-bN** → **raced** | **L-aN-bN**) = .0001	– a transitive participial verb may be 'raced.'
P(**V-aN** → **raced** | **V-aN**) = .001	– an intransitive finite verb may be 'raced.'
P(**V-aN** → **fell** | **V-aN**) = .001	– an intransitive finite verb may be 'fell.'
P(**D** → **the** | **D**) = .5	– a determiner may be 'the.'
P(**N-aD** → **horse** | **N-aD**) = .001	– a common noun may be 'horse.'
P(**N-aD** → **barn** | **N-aD**) = .001	– a common noun may be 'barn.'
P(**R-aN-bN** → **past** | **R-aN-bN**) = .1	– an adverbial preposition may be 'past.'

  The probability of a tree is then product of probabilities of its rules.



(First tree: $\overbrace{1\cdot.5\cdot.05\cdot1\cdot.5}^{\text{grammatical rules}}\cdot\overbrace{.5\cdot.001\cdot.001\cdot.1\cdot.5\cdot.001}^{\text{lexical rules}} = .\overbrace{000000000000}^{12}3125$.)

(Second: $\underbrace{1\cdot.1\cdot.5\cdot.1\cdot1\cdot.5}_{\text{grammatical rules}}\cdot\underbrace{.5\cdot.001\cdot.0001\cdot.1\cdot.5\cdot.001\cdot.001}_{\text{lexical rules}} = .\underbrace{000000000000}_{12}00000625$.)

The probability of the observed words is the sum of the probabilities of possible trees:

$.\underbrace{000000000000}_{12}3125 + .\underbrace{000000000000}_{12}00000625 = .\underbrace{000000000000}_{12}31250625$.

- When hypotheses don't predict observed words, activation is reallocated

  (the wrong tree and its probability is removed when the first incompatible word is reached).

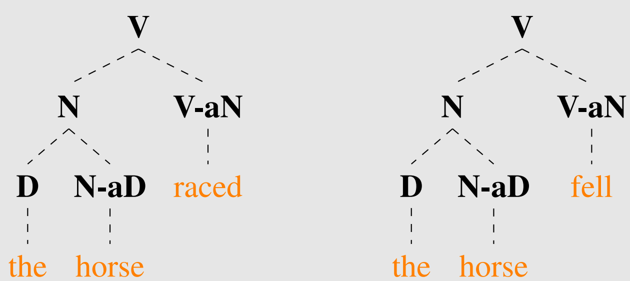
- Reading delay is proportional to the fraction of remaining activation (probability) over total:

  $\frac{.00000000000000000625}{.00000000000031250625} = 0.0000196 = 2^{-15.639}$ (surprisal is written as $-\log_2$ of this, so 15.639).

  (The delay may also be proportional to the *log* of this fraction.)

Practice 14.1:

Using the above rules, what's the probability of 'the horse raced,' and 'the horse fell':



Practice 14.2:

What share of activation (probability) remains, above, when 'raced' is encountered?

(You may limit your consideration to just the above two trees.)

(The probability before 'raced' is the sum of both trees; the probability after just contains the first.)

References

[Bever, 1970] Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279–362). New York: Wiley.

[Frazier, 1979] Frazier, L. (1979). *On comprehending sentences: syntactic parsing strategies*. PhD thesis, University of Connecticut.

[Gorrell, 1991] Gorrell, P. (1991). Subcategorization and sentence processing. In R. Berwick, S. Abney, & C. Tenny (Eds.), *Principle-based parsing*, Studies in linguistics and philosophy ; 44. Dordrecht: Kluwer.

[Hale, 2001] Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics* (pp. 159–166). Pittsburgh, PA.

[Kjelgaard & Speer, 1999] Kjelgaard, M. M. & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, 40(2), 153–194.

[Levy, 2008] Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.

[Mitchell, 1987] Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII: The Psychology of Reading* (pp. 601–618). Hillsdale, NJ: Erlbaum.

[Tanenhaus et al., 1995] Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.

[Trueswell et al., 1994] Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33(3), 285 – 318.

[Trueswell et al., 1993] Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference form garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(3), 528–553.

[van Gompel et al., 2001] van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing. evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, (pp. 225–258).