

January 13, 2004

Chapter 2.1-2.3

Sentence Structure, Word-level categories

Administrivia

- Web page: <http://courses.washington.edu/ling461>
- HW1 due at the beginning of class on Thursday
- Note cards:
 - Name
 - Email address
 - Major
 - Other linguistics classes taken
 - Languages spoken/studied
 - What you hope to get out of this class

The Big Picture (1/3)

- Big claim #1: Sentences are not merely flat strings of words, but have hierarchical structure.
- Speakers have some intuitions about this structure, but it is more clearly established through linguistic analysis.
- Linguists have fairly firm intuitions about this structure (why might that be?).

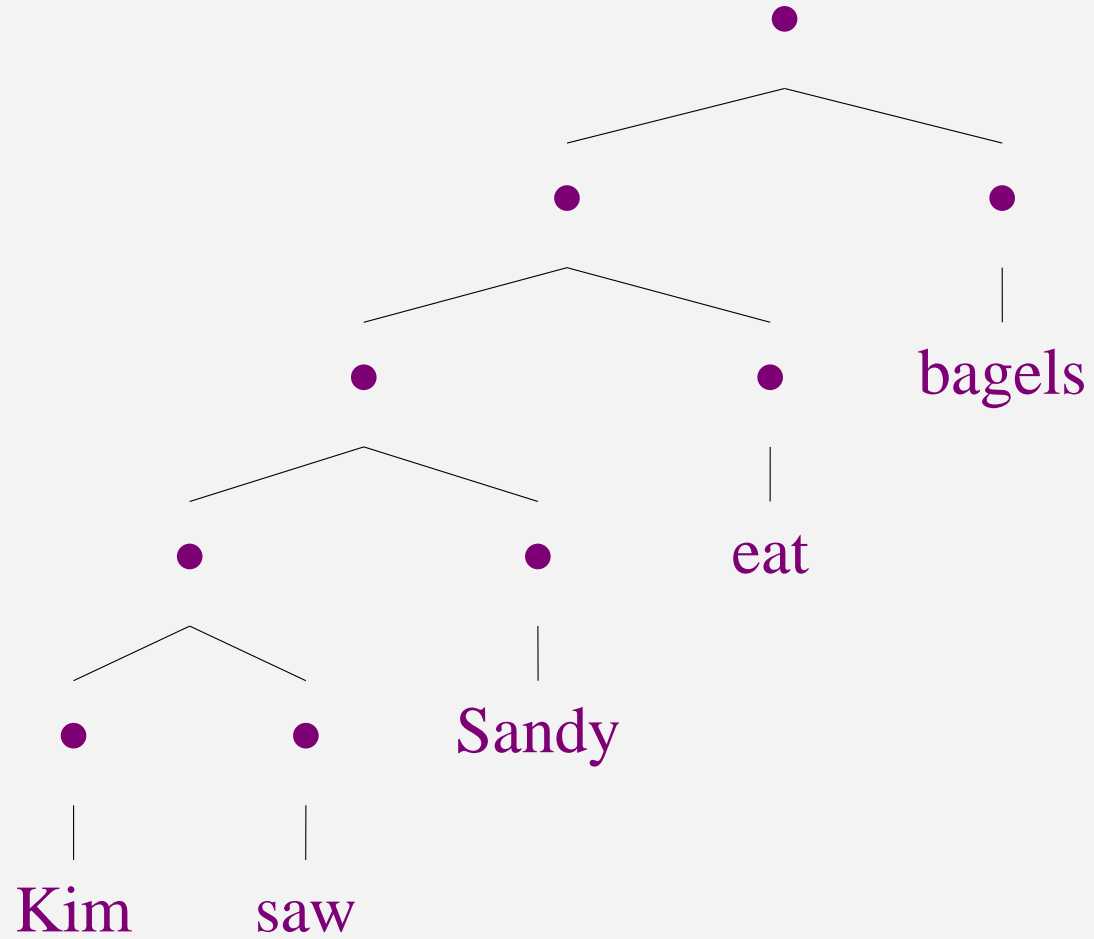
The Big Picture (2/3)

- Big claim #2: Words and phrases can be grouped into *equivalence classes* called categories.
- Once again, speakers have some intuitions about equivalence classes, but they are more clearly established through linguistic analysis.
- Speakers don't have intuitions about names of classes, unless they've studied them somewhere.

The Big Picture (3/3)

- Today:
 - A little bit on constituent structure
 - A lot on word classes
- Next time:
 - More on constituent structure
 - Phrase classes

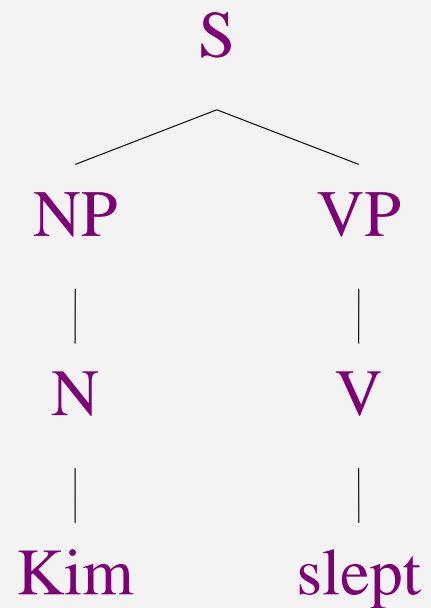
Is this structure right?



Trees and labeled bracketings (1/3)

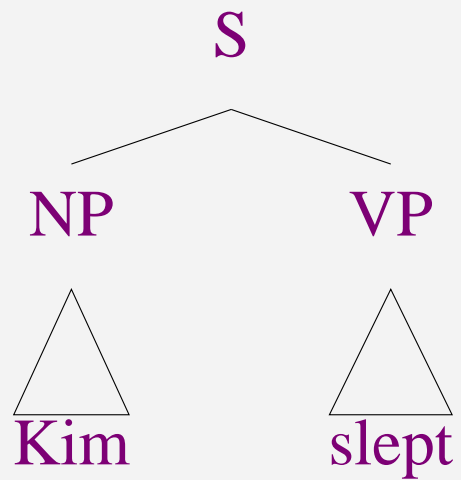
- Both show constituent structure.
- No partially overlapping (i.e., crossing) constituents.
- In most cases, both show the same information, although trees tend to be easier to read.
- In some cases, trees actually do encode more information.

Trees and labeled bracketings (2/3)



[S [NP [N Kim]] [VP [V slept.]]]

Trees and labeled bracketings (3/3)



[S [NP Kim] [VP slept.]]

Exercise: How consistent are our intuitions?

- I saw a bird.
- I saw a bird yesterday.
- I have never seen any birds.
- Hopefully, I will see a bird tomorrow.
- You saw a bird, didn't you?
- I gave up on seeing any more birds.

Word classes: Types of evidence/arguments

- Phonological (stress assignment)
- Semantic (ambiguity)
- Morphological (inflectional classes)
- Syntactic (distributional)
- (Language acquisition)
- (Psycholinguistics)

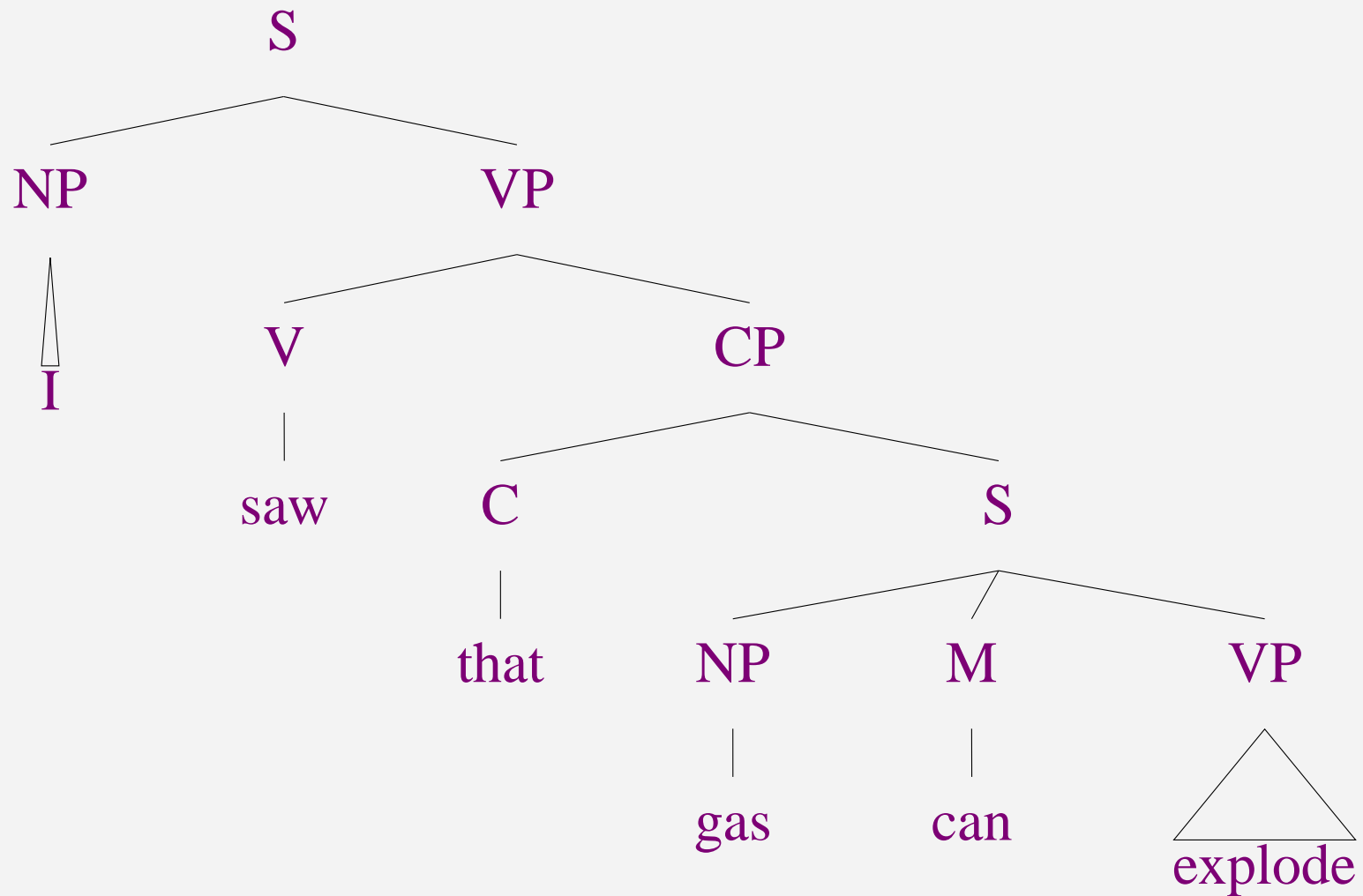
Phonological evidence: Stress assignment

- Stress assignment in certain Latinate borrowings into English is sensitive to word class.
- What's the pattern here?
 - I like réconds./Let's recórd that.
 - I made good prógress./It didn't progréss very far.
 - We présented them with a présent.
 - I won't permít you to do that without a pérmit.
- Further examples of/counter-examples to this pattern?
- Other phonological phenomena (English or otherwise) which are sensitive to word class?

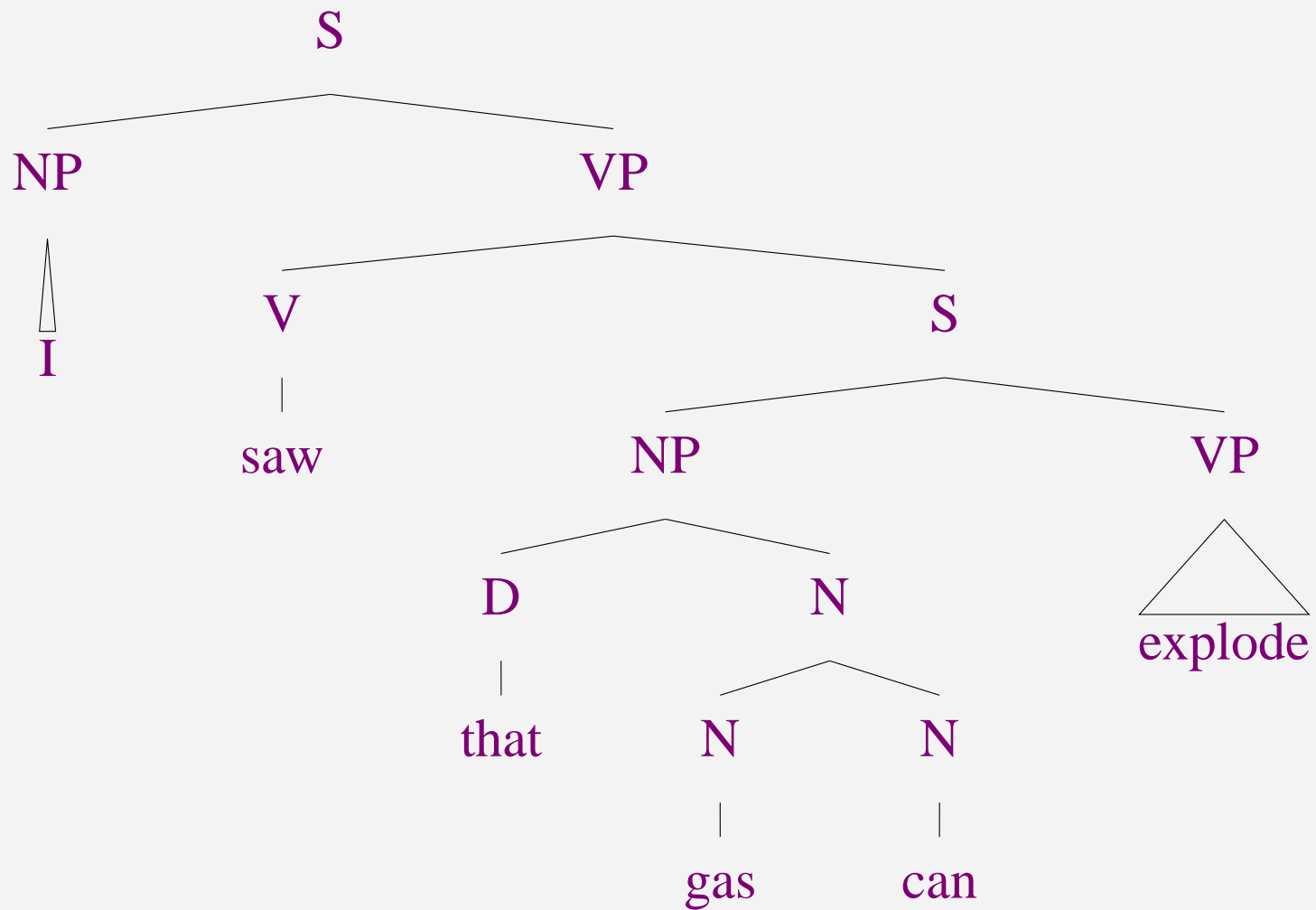
Semantic evidence: Ambiguity (1/4)

- Ambiguous (as opposed to vague) sentences have two different meanings which correspond to two different ‘parses’ or ‘P-markers’.
- Two kinds of ambiguity: lexical and structural.
- A single ambiguous sentence might involve both.
 - Lexical only: *The bank is crumbling.*
 - Structural only: *Kim saw the astronomer with the telescope.*
 - Both: *I saw that gas can explode.*
 - Flamboyantly both: *Have that report on my desk by Friday.*

Semantic evidence: Ambiguity (2/4)



Semantic evidence: Ambiguity (3/4)



Semantic evidence: Ambiguity (4/4)

- Because the words *that* and *can* stand in different relationships to the other words in the sentence in the two structures, we conclude that they belong to different word classes.
- What assumptions underlie that conclusion?

Getting down to the nitty gritty...

- The phonological and semantic evidence motivate the need for word classes.
- They won't get us very far in figuring out all the word classes we'll need.
- This is because word classes are fundamentally (morpho)syntactic.
- Phonological and semantic facts reflect them only some of the time.
- Thus *inflectional* and *distributional* evidence will be what we use to establish the specific categories we use.

Morphological evidence: Inflection

- Morphological processes are sensitive to word class.
- No morphological process applies to all words.
- In order to state the rules for the processes, we need to classify words.

Inflecting verbs (1/2)

- English verbs have the following forms:
 - Base form (used after modals)
 - Present tense, with 3rd person subjects
 - Past tense
 - Past participle (used after *have* and in passives)
 - Present participle (used after *be* and *keep*)
- For example:
do, does, did, done, doing

Inflecting verbs (2/2)

- Not all verbs distinguish all forms, that is, for many verbs, two or more of the forms are homophonous.
- Examples?
- One verb distinguishes more forms. Which?
- Not all verbs have all of the forms, that is, some verbs are disallowed from some contexts.
- Examples?

Inflecting modals

- Modals (*can, will, might, etc.*) are a subclass of verbs.
- They contrast with *main verbs*.
- They have fewer inflectional forms (in fact only one), and different distributional properties.

Inflecting other word classes

- What kind of inflectional affixes to nouns take?
- Adjectives?
- Adverbs?
- Prepositions?
- Determiners?
- Number names?
- Hesitation forms?

Derivational morphology

- Certain *derivational* morphological processes change the word class of a form.
- For example *-ation* creates nouns out of some verbs.
- How do you make a verb out of a noun?
- A verb out of an adjective?
- A noun out of an adjective?
- An adverb out of an adjective?
- A noun out of an adverb?

Distributional analysis

- Distributional analysis distinguishes categories by the syntactic contexts they can show up in.
- Why does distribution across syntactic contexts motivate word classes?
- The resulting categories correspond fairly well (but not perfectly!) to those defined by the inflectional system.

Distributional analysis: Exercise 1

- Here's a sentence frame for nouns:
Kim has always liked __ .
- Can all nouns show up there?
- Find other sentence frames for nouns.
- Can all nouns show up there?

Subclasses: Count v. mass nouns (1/2)

- Count nouns (*cat, dog, toothbrush*) generally refer to countable objects.
- Mass nouns (*information, furniture, rice, water*) generally refer to uncountable substances.
- Singular count nouns cannot generally appear without a determiner.
- Mass nouns do not require a determiner.

Subclasses: Count v. mass nouns (2/2)

- Count nouns can appear with *a* and *many*.
- Mass nouns can appear with *much*.
- Either kind can appear with *the*.
- In the right contexts, mass nouns can be used as count nouns and vice versa:
 - I'd like a water, please.
 - There was cat all over the driveway.

Distributional analysis: Exercise 2

- Main verbs (*eat*) and modals (*will*) are subclasses of verb.
- Find a sentence frame with accepts either.
- Find a sentence frame which only accepts main verbs.
- Find a sentence frame which only accepts modals.

So how many classes are there? (1/2)

- Tagsets from computational linguistics provide one estimate.
- They are designed for tagging each word in a large corpus with its part of speech....
- ... and therefore can't set aside phenomena as 'peripheral'.

So how many classes are there? (2/2)

- A sampling:

Brown	227	ICE	205
London-Lund	210	LOB	153
Polytechnic of Wales	66	Penn Treebank	46

- For more information:

<http://www.comp.leeds.ac.uk/amalgam/tagsets/tagmenu.html>

Aligning word classes cross-linguistically

- For any pair of languages, there are probably word classes that are not shared between them.
- There are certainly subclasses that aren't.
- Examples?

Aligning word classes cross-linguistically

- Nonetheless, it seems intuitive to say that all languages have nouns, and (probably) verbs, etc.
- Radford dismisses the ‘notional’ definitions of the word classes...
- ... but something along the lines of ‘canonical nouns are used to refer to entities in the world’ is necessary to draw equivalences across unrelated languages.

Distinguishing the model from the data

- The model includes word classes, assignments of words to classes, and rules (phonological, morphological, syntactic) which refer to the word classes.
- The data are merely sets of sentences and non-sentences, words and non-words.

Summary

- Constituent structure
- Trees and labeled bracketings
- Evidence for word classes:
 - Phonological
 - Semantic
 - Morphological
 - Syntactic
- Next time:
 - More on constituent structure
 - Categories of phrases