

# **HG2002 Semantics and Pragmatics**

## **Introduction, Organization What does it mean to mean?**

Francis Bond

**Division of Linguistics and Multilingual Studies**

<http://www3.ntu.edu.sg/home/fcbond/>  
[bond@ieee.org](mailto:bond@ieee.org)

Lecture 1

Location: HSS Auditorium

Creative Commons Attribution License: you are free to share and adapt as long as you give appropriate credit and add no additional restrictions:

<https://creativecommons.org/licenses/by/4.0/>.

HG2002 (2018)

# Overview

---

- Syllabus; Administrivia
- What is semantics?
- Why should we be interested in semantics?
- What is meaning?
- Meaning as an open ended conceptual system
- Semantic problems and solutions?
- Information Theory (new!)

# Self Introduction

---

- BA in Japanese and Mathematics
- BEng in Power and Control 
- PhD in English on *Determiners and Number in English contrasted with Japanese, as exemplified in Machine Translation*
- 1991-2006 NTT (Nippon Telegraph and Telephone)
  - Japanese - English/Malay Machine Translation
  - Japanese corpus, grammar and ontology (Hinoki)
- 2006-2009 NICT (National Inst. for Info. and Comm. Technology)
  - Japanese - English/Chinese Machine Translation
  - Japanese WordNet
- 2009- NTU

# Administrivia

---

**Coordinator** Francis Bond <bond@ieee.org> !<fcbond@ntu.edu.sg>

Details about the tutor, lecture and tutorial times and locations are online:

<http://compling.hss.ntu.edu.sg/courses/hg2002/>

# 100% Continuous Assessment

---

- Class Participation (10%)
- Assignment (30%)
  - The assignment involves some annotation  
— determining the meaning of words in context
- Mid-term Quiz (20%)
- Final Exam (40%)

# Guidelines for Written Work in LMS

---

- All assignments must follow the *Guidelines to Submitting Written Work for the Division of Linguistics and Multilingual Studies*
  - You can get it from: <http://linguistics.hss.ntu.edu.sg/CurrentStudents/Pages/Resources.aspx>
  - I am **strict** about assignment length and format  
learning to read and follow instructions is a **very** useful skill
  - Useful advice on citation, transcription, formatting
  - I also recommend my own (*Computational*) *Linguistics Style Guide*: [www3.ntu.edu.sg/home/fcbond/data/ling-style.pdf](http://www3.ntu.edu.sg/home/fcbond/data/ling-style.pdf)
  - Proper citation is important  
— failure to cite is plagiarism — **fail subject**  
See the NTU code of academic integrity  
<http://academicintegrity.ntu.edu.sg/>

## What do you learn?

---

Students will learn semantics at an introductory level and they will acquire semantic analysis skills. With these skills and their knowledge of semantic approaches, students will be able to approach natural language data, as well as develop awareness of the inherent connections between semantics and other branches of linguistics.

# Textbook and Readings

---

- Textbooks
  - Saeed, John (2009). *Semantics*. 3rd Edition. Wiley-Blackwell. (**required**)
  - Lyons, John (1977) *Semantics*. Cambridge University Press (**recommended**)
  
- From next week, I expect you to read all chapters assigned before class.
  
- Ideas from the book will be pursued in parallel with the topics given above.

# Student Responsibilities

---

By remaining in this class, the student agrees to:

1. Make a good-faith effort to learn and enjoy the material.
2. Read assigned texts and participate in class discussions and activities.
  - Question marks mean I will ask you something
3. Submit assignments on time.
4. Attend class at all times, barring special circumstances (see below).
5. Get help early: approach us when you first have trouble understanding a concept or homework problem rather than complaining about a lack of understanding afterward.

?

---

6. Treat other students with respect in all class-related activities, including on-line discussions.

# Attendance

---

1. You are expected to attend all classes.
2. Be on time - lateness is disruptive to your own and others' learning.
3. Valid reasons for missing class include the following:
  - (a) A medical emergency (including mental health emergencies)
  - (b) A family emergency (death, birth, natural disaster, etc).

You must provide documentation to the student office who then contacts me.
4. There will be significant material covered in class that is not in your readings. You cannot expect to do well without coming to class.

- 
5. If you miss a class, it is your responsibility to get the notes, any handouts you missed, schedule changes, etc. from a classmate.

# Remediation and Academic Integrity

---

1. No late work will be accepted, except in the case of a documented excuse.
2. For planned, justified, absences on class days or days on which assignments are due, advance notice must be provided.
3. Cheating will not be tolerated. Violations, including plagiarism, will be seriously dealt with, and could result in **a failing grade for the entire course**.
4. For all other issues of academic integrity, refer to the University Honour Code:  
<http://academicintegrity.ntu.edu.sg/>
5. As always, use your common sense and conscience.

# The winning strategy

---

- Read the books before class (and after again, if necessary)
- Work together: make study groups
- Homework/Tutorials: Discuss as much as you want, write up your own answers
- Exams: No discussion
- Ask questions ... early and often!

---

# Introduction to Semantics

# What is Semantics

---

- Very broadly, semantics is the study of meaning
  - Word meaning
  - Sentence meaning
- Why do we want to study meaning?
- What kind of knowledge does it take for a speaker to produce language and for a hearer to comprehend language?

# Layers of Linguistic Analysis

---

1. Phonetics & Phonology
2. Morphology
3. Syntax
4. Semantics
5. Pragmatics

## Two theories

- Semantics is **autonomous**, a separate module
- Semantics is **integrated** with other knowledge, inseparable
  - linguistic knowledge is inseparable from encyclopedic knowledge

# Do we share a common conceptual system?

---

- What is a *high school*?
- What color is *blue*?
- What is *carrot cake*?
- What does *verb* mean?



# Meaning is an open-ended conceptual system

---

➤ Lexical innovation

➤ *Meritocracy* (1958)

➤ *LASER* (1960)

➤ *WWW*

...

➤ Is this association between creating new words and creating new concepts justified?

➤ Can we have a new concept without a new word?

?

➤ Can we have a new word without a new concept?

?

➤ More creativity

➤ *I am so hungry I can eat ten million elephants.*

# More Meaning

---

We can define the meaning of a speech form accurately when this meaning has to do with some matter of which we possess scientific knowledge. We can define the names of minerals, for example, in terms of chemistry and mineralogy, as when we say that the ordinary meaning of the English word *salt* is 'sodium chloride (NaCl)', and we can define the names of plants or animals by means of the technical terms of botany or zoology, but we have no precise way of defining words like *love* or *hate*, which concern situations that have not been accurately classified – and these latter are in the great majority.

*Language* (Bloomfield, 1933)

- But is *salt* really just NaCl?
- And what about non-scientific speakers?

# Determining meaning

---

Some useful concepts

- **Synonymy**:  $A$  means the same as  $B$
- **Contradiction**:  $A$  and  $B$  cannot both be true
- **Entailment**: if  $A$  is true then  $B$  must also be true
- **Ambiguous**:  $A$  has more than one meaning

# Meaning in the larger context

---

- Semiotics is the study of interpreting symbols, or **signification**
  - We refer to the **signified**
  - Using a **signifier**

Saussure

- Signs can be more or less related to their objects

**icon**      map or diagram      Children Crossing



**index**      closely represented      Roundabout



**symbol**      arbitrary      Stop



- Are words icons, indexes or symbols?

?

# Problems with defining meaning

---

- The **grounding** problem and **circularity**
- The boundaries of meaning: **linguistic** vs **encyclopedic knowledge**
- Regional variation in meaning: **dialects** “the usage or vocabulary that is characteristic of a specific group of people”
- Individual variation in meaning: **idiolects** “the language or speech of one individual at a particular period in life”
- There is a shared usage of words and meanings that defines a language

# Metalanguages and Notational Conventions

---

We use language to talk about language, which can get messy. So we try to use certain words with very specific technical senses.

- **technical term** ← remember me!
- *word* “gloss” or *utterance*
- *lexeme*
- *predicate*

# Word Meaning and Sentence Meaning

---

- We store information about words in our **mental lexicon**
  - It is still unclear what exactly a word is!
- Words can be combined to form an infinite number of expressions
  - This building up of meaning is referred to as **composition**
  - If the meaning of the whole can be deduced from the parts then it is **compositional**

# Reference and Sense

---

- Words **refer** to things in the world (like **unicorns**)
- The meaning of a word across different contexts is often referred to as its **sense**
  - Same word can refer to different things
    - \* English: *I put my money in the bank*
    - \* English: *I fell asleep at the river bank*
  - Same basic concept can have different boundaries
    - \* French: *mouton* “sheep/mutton”
    - \* English: *sheep* vs *mutton*
    - \* Japanese: *hato* “dove/pigeon”
    - \* English: *dove* vs *pigeon*

# Utterances, Sentences and Propositions

---

- **utterance**: an actual instance of saying (or writing or ...) something
  
- **sentence**: an abstraction, the type of what was said
  - (1) Caesar invades Gaul
  
- **proposition**: a further abstraction, normally ignoring some non-literal meaning
  - (2) `invade(Caesar, Gaul)`
  
- **information structure**: what part of a proposition is emphasized
  - (3) *Caesar invaded Gaul*
  - (4) *Gaul was invaded by Caesar*

- 
- (5) *It was Gaul that Caesar invaded*
  - (6) *It was Caesar who invaded Gaul*

# Propositions

---

- A logical construct
- Abstracts away from grammatical differences
  - (7) *John kicked the dog*
  - (8) *The dog was kicked by John*
  - (9) ジョンが犬を蹴った . *John-ga inu-wo ketta.*
- Can be reasoned over (**logic**)
- Can be formalized
  - $x, y(\text{named}(\text{John}, x), \text{dog}(y), \text{the}(y), \text{kick}(e, x, y), \text{past}(e))$

# Non-literal meaning

---

- Consider “That Mitchell and Webb Look”  
Season One Episode Two 1:10–
  - Why aren't they more direct? ?
  - Is the meaning clear anyway? ?
  
- Can you give some examples of non-literal meaning? ?

# Teaching Meaning through Annotating Text

---

- Read a Sherlock Holmes story
  - So far SPEC, DANC, REDH and SCAN  
(also news, essays and Japanese short stories)
- Look at every content word
  - Find its meaning in a dictionary (wordnet)
  - Or write a new definition if needed
  - Say if it has positive or negative connotation (for some)

this shows how little you need to know to understand and enjoy

# Defining Meaning

---

- When we use a word, we don't have to know everything about the referent
  - A *dog-cart* is a kind of CART
  - ⇒ you can ride it
  - ⇒ it has wheels
  - ⇒ it has something to do with a dog
- We infer that it has many of the same properties as its hypernym, even though this is not always true
  - A *hover-car* is a kind of CAR
  - ⇒ you can ride it
  - ⇏ it has wheels
- Many of the properties may be irrelevant to the story at hand, and irrelevant to the syntax of the language

## How do we learn?

---

*You shall know a word by the company it keeps* (Firth, 1957, p11)

- You see a new word **in context**  
*buttoning up his pea-jacket,*
  - ⇒ it is a kind of jacket (green jacket?)
  - ⇒ with buttons
  - ? it is thick material (they are going to a stake out)
  - ? it has something to do with peas  
not true (from the West Frisian word *pijjaker*, in which *pij* referred to the type of cloth used, a coarse kind of twilled blue cloth)
- And you deduce information from the context
- We are getting better at doing this with computers
  - but people use more than words: eyes, noses and other senses

## How else do we learn?

---

- From word internal cues
  - *Television* “far vision”
  - *iphone* “internet phone” (also individual, instruct, inform, inspire)
  - 鯖 *saba* “mackerel” = 魚 fish; 青 blue
- From the sound
  - *bouba/kiki* ★ or ♣
  - *banged, beaten, battered, bruised, blistered, bashed*
  - mouth shape for *teeny weeny* vs *large*



- From images:

*Magnifying Glass*

## Words are related in many other ways

---

- Domains: *ball, racket, net, love, ace*
- Origin: *chew, eat, drink* vs *masticate, consume, imbibe*
- ? come up with some words with different origins  
English or another language!
- Dialect: *ripper, bonza, sickie, no worries*
- Part-of-speech: *die, live* vs *death, life*
- When you learned them!
- and many more

?

All of these relations affect how you use and understand language.

# Idioms

---

- Some expressions clearly involve more than one orthographic word
  - compound noun
    - \* grass snake; grass and tree snakes
  - verb-particle
    - \* I looked it up vs I looked up the very long word
  - idiom
    - \* going great guns, give the Devil his due
    - \* jog someone's memory
    - \* blow one's top, cast one's eyes
- Knowing the individual words is not enough to know the meaning (or usage)

---

# Information Theory: another way of looking at meaning

# A gentle introduction to Information Theory

---

- Language has many uses, only one of which is to convey information  
— but surely transferring information is important
- How can we measure information?  
Shannon, C.E. (1948), "A Mathematical Theory of Communication", Bell System Technical Journal, 27, pp. 379–423 & 623–656, July & October, 1948. <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>
- How can we get our message across efficiently and safely?

# Consider an abstract case

34

*The Mathematical Theory of Communication*

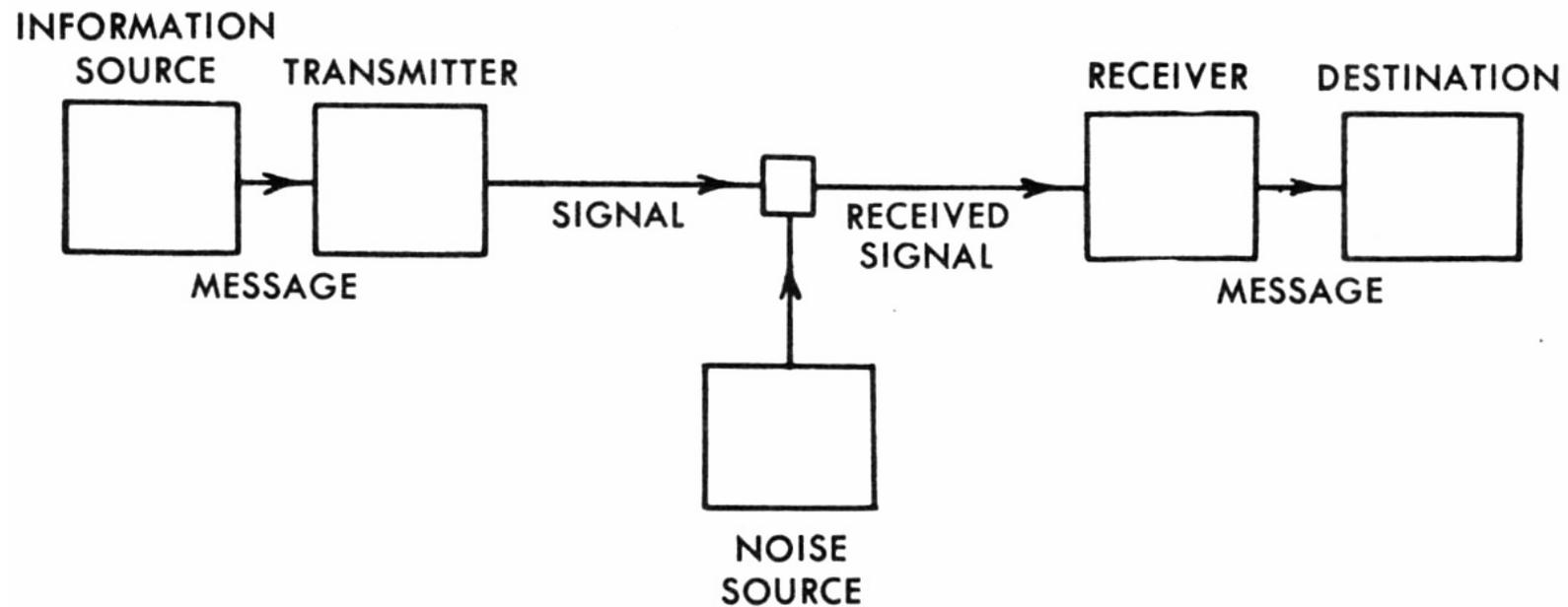


Fig. 1. — Schematic diagram of a general communication system.

# Information as bits

---

➤ Suppose we have 8 equally likely facts (A, B, C, ...H).  
How many yes-no questions does it take to pinpoint one fact?

?

➤ This is the information in **bits**  
you need  $n$  bits of information

➤ Technically the **Entropy**

$$H(\mathbf{p}) = - \sum_{x \in X} p(x) \log_2 p(x)$$

➤ We won't do all the maths

➤ Just think of things in terms of questions

## What if some facts are more common?

---

- Simplified Polynesian (6 letters: p, k, i, u, t, a):  
Distribution: p ( $\frac{1}{8}$ ); k ( $\frac{1}{8}$ ); i ( $\frac{1}{8}$ ); u ( $\frac{1}{8}$ ); t ( $\frac{1}{4}$ ); a ( $\frac{1}{4}$ )
- We can do one letter in  $2\frac{1}{2}$  bits
  - Is it (a or t) or (p, k, i, u), ...
- We can define a  $2\frac{1}{2}$  bit code
  - p (100); k (101); i (110); u (111); t (00); a (01)
- Which codes should be longer — frequent or infrequent letters?
  - What does this imply for language?

# What if there is mutual information?

---

- **Mutual information** measures the information that  $X$  and  $Y$  share: how much does knowing one variable reduce our uncertainty about the other.
  - What is the next letter?
  - What is the next letter following  $t$ ?
  - What is the next letter following  $a$ ?
  - What is the next letter following  $q$ ?
  - What is the next letter following  $th$ ?
  - What is the next letter following  $as$ ?
  - What is the next letter following  $qu$ ?
  - What is the next letter following *semanti*?
- A language model and more context improves our guess

# Different Models for English

---

Consider only 26 lowercase letters and a space, and a language model based on probability (Hidden Markov Model). How many guesses do we need on average to guess the next letter?

- Zeroth order (random) =  $\log_2 27 = 4.76$
- First order (frequency) = 4.03 (pick e)
- Second order (one previous letter) = 2.8
- Human (two previous letters) = 1.34

Surrounding context helps interpretation

# What if there is noise?

---

- Imagine you want to send a signal, but randomly a bit gets flipped (noise)
  - Original message /pa/ 100 01
  - Received message /ka/ 101 01
- A very efficient model is weak to noise
- If we make the message longer, we can guard against this
  - Original message /pa/ 100 100 100 01 01 01
  - Received message /ka/ 101 100 100 01 01 01
  - We add **redundancy** to the signal
  - There are much better encodings than this (**Hamming codes**)
- Hmn lngge s vr rdndnt

# Bandwidth

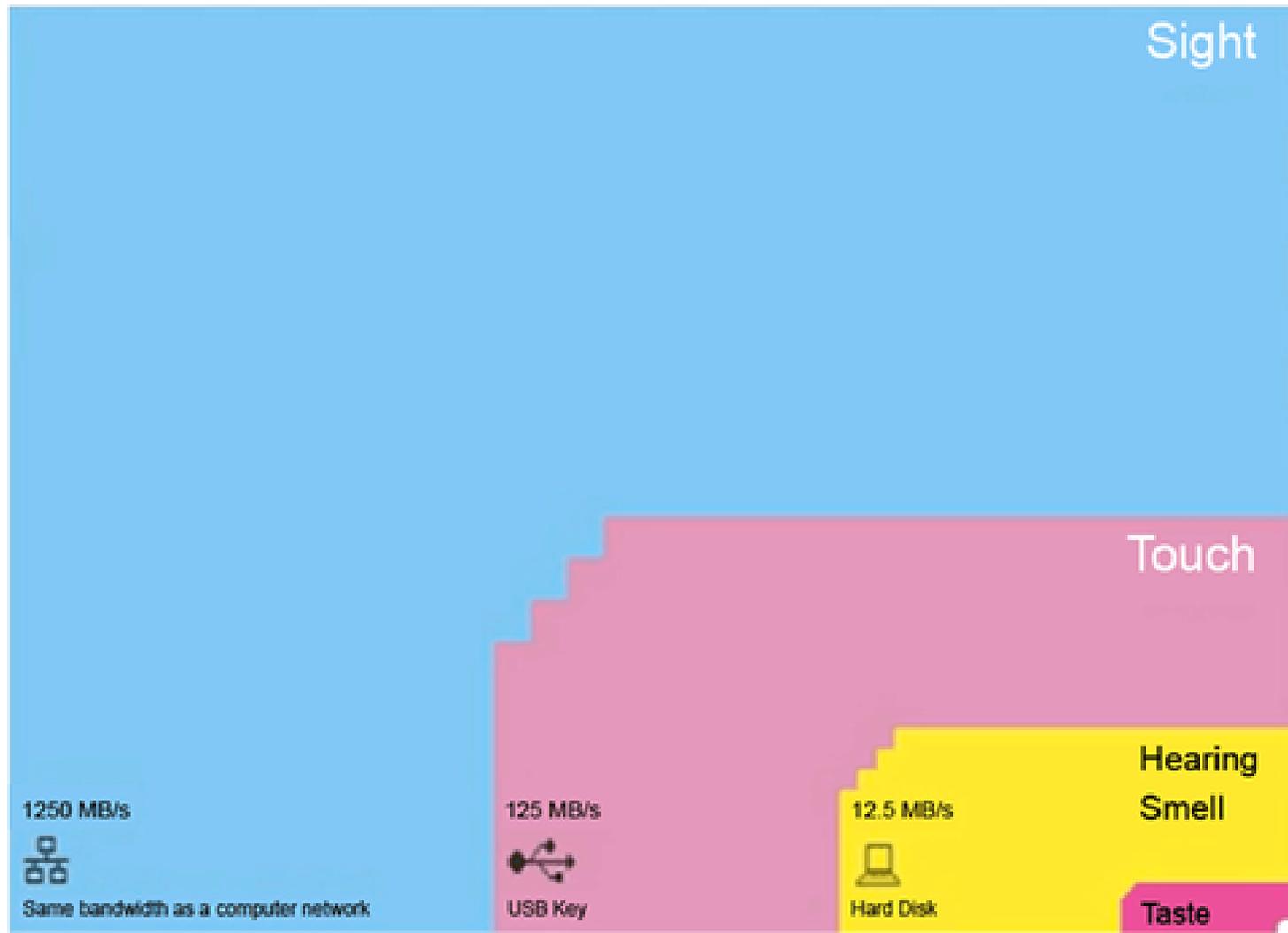
---

- **Bandwidth** (or **bit-rate**) is the maximum throughput of a logical or physical communication path in a digital communication system: e.g. how much information can be conveyed
- Human Communication: words per minute (one word is 6 characters)  
(English, computer science students, various studies)

Modality	Normal	Peak	Comment
Reading	300	200	(proof reading is slower)
Writing	31	21	(composing)
Speaking	150		(plus all kinds of tone/nuance)
Hearing	150	210	(speeded up)
Typing	33	19	(composing)

- We adapt our communication depending on the bandwidth available

# Bandwidth of our Senses



# Representing meaning

---

- One of our goals will be to represent meaning
- There are various ways to do this
  - Syntactic trees
  - Logical forms
  - Thesauri and Ontologies
  - Translation
  - Paraphrasing

Can you think of others?

- At the end of this course you should be able to use these to describe many aspects of meaning

# Language is normally under-specified

---

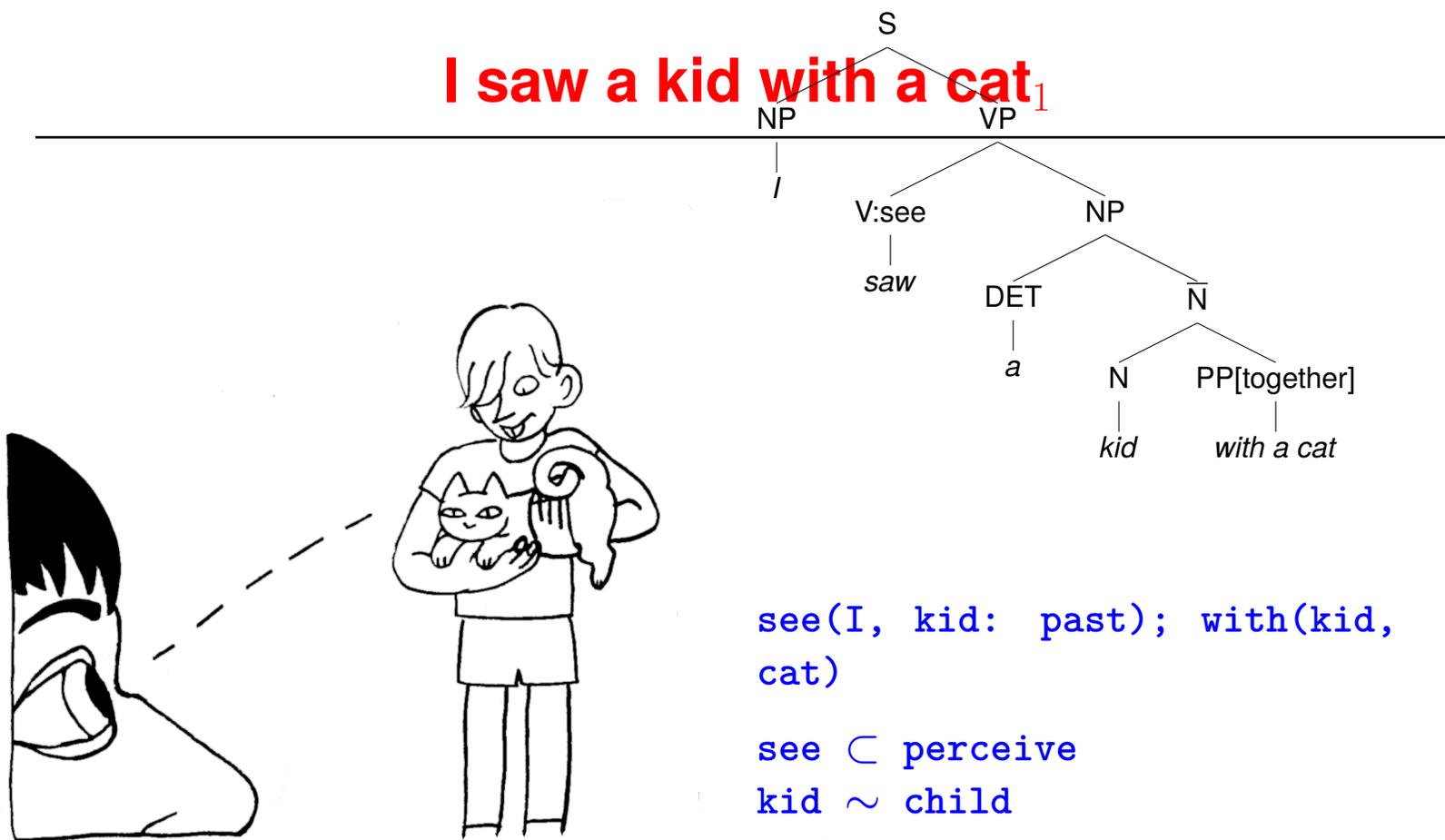
We get **words**:

*I saw a kid with a cat.*

We want **meaning**:



# I saw a kid with a cat<sub>1</sub>



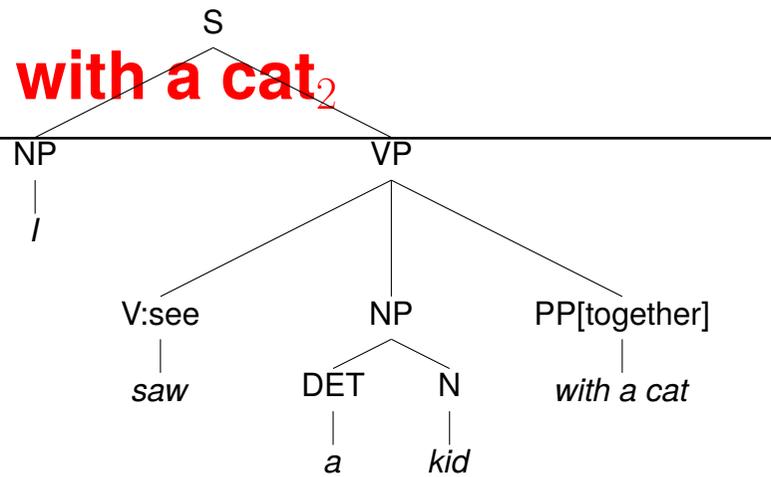
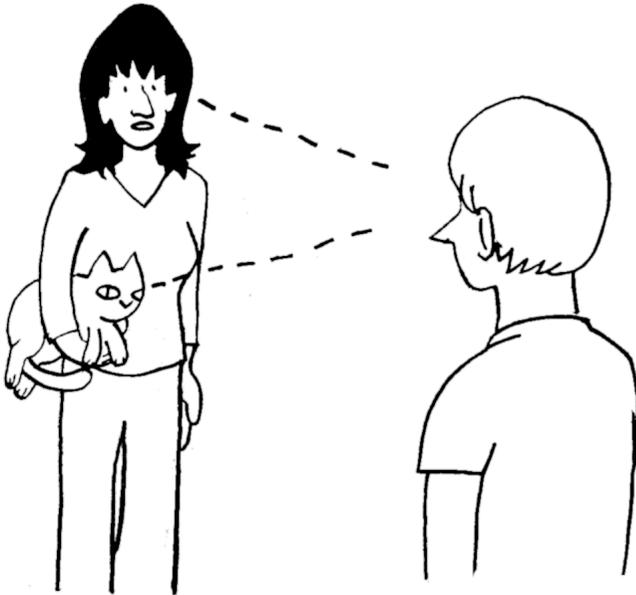
see(I, kid: past); with(kid, cat)

see  $\subset$  perceive

kid  $\sim$  child

with  $\subset$  together

# I saw a kid with a cat<sub>2</sub>



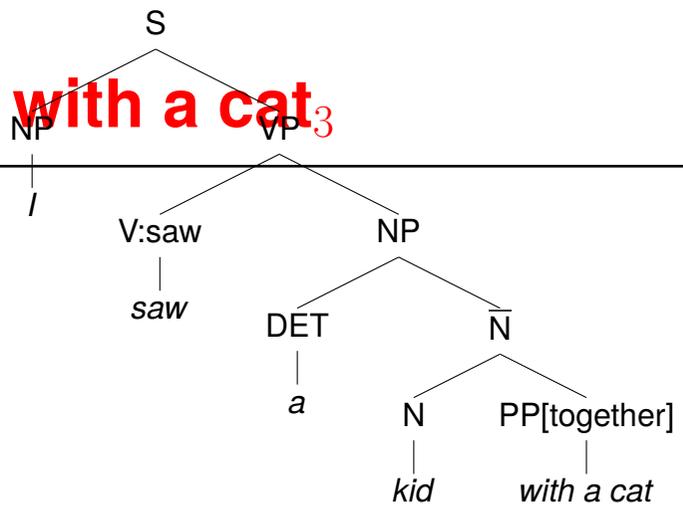
see(I, kid: past) with(I, cat)

see  $\subset$  perceive

kid  $\sim$  child

with  $\subset$  together

# I saw a kid with a cat



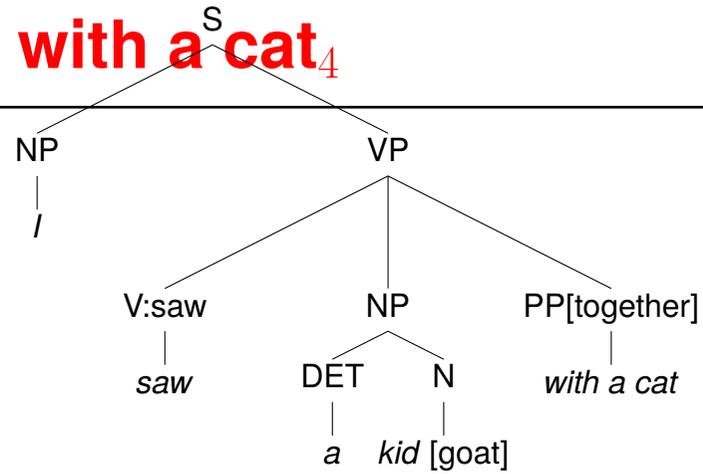
saw(I, kid: pres); with(kid, cat)

saw  $\subset$  cut

kid  $\sim$  child

with  $\subset$  together

# I saw a kid with a cat<sub>4</sub>



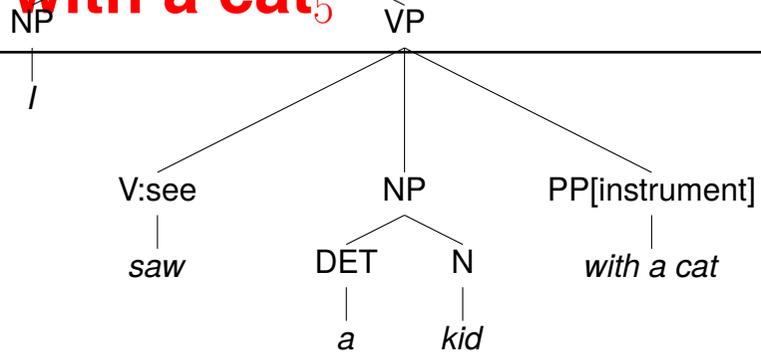
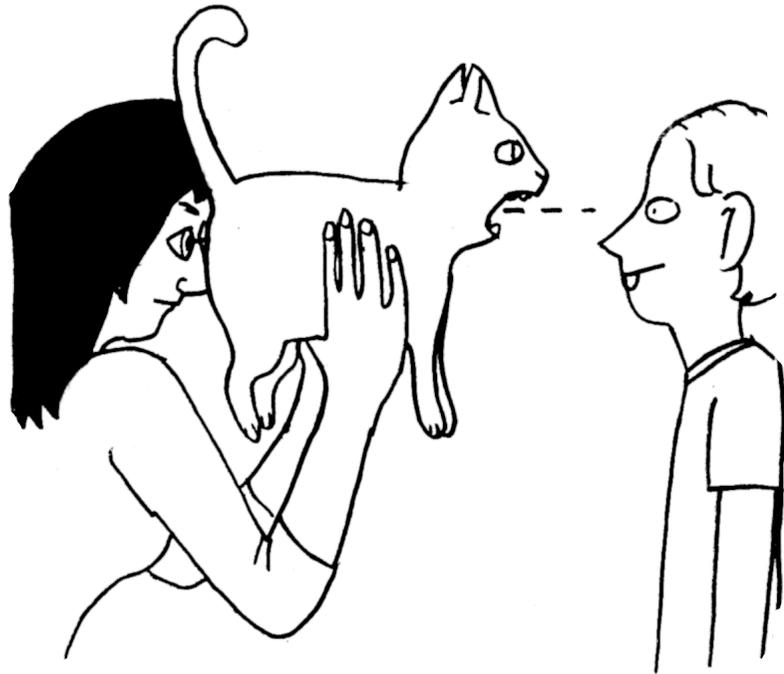
saw(I, kid: present) with(I, cat)

saw  $\subset$  cut

kid  $\sim$  young goat

with  $\subset$  together

# I saw a kid with a cat



see(I, kid: past) with(I, cat)

see  $\subset$  perceive

kid  $\sim$  child

with  $\subset$  instrumental

## We can also use translations

---

- (10) 我 看到了 一个抱着 猫 的 孩子  
wǒ kàndào le yīgè bào zhe māo de hái zi.  
I saw one holding cat 's child  
I did see a child holding a cat
- (11) 我 抱着 猫 看到了 一个 孩子  
wǒ bào zhe māo kàndào le yīgè hái zi  
I holding cat saw one child  
I holding a cat did see a child
- (12) 我 锯锯 一个 孩子 和 他/她 的 猫  
wǒ jù yīgè hái zi hé tā/tā de māo  
I saw one child and he/she 's cat  
I saw a child and their cat
- (13) 我 和 一只 猫 锯锯 一只 小 山羊  
wǒ hē yīzhǐ māo jù yīzhǐ xiǎo shānyáng  
I and one cat saw one small goat  
I and a child saw a young goat

---

(14) 我 用 一 只 猫 看 到 了 一 个 孩 子  
wǒ yòng yīzhǐ māo kàndào le yīgè hái zi  
I use one cat saw one child  
Using a cat, I did see a child

Your turn: try to paraphrase — translate into English  
aim to be unambiguous, even if slightly disfluent

# Summary

---

- Syllabus; Administrivia
- What is semantics?
- Why should we be interested in semantics?
- What is meaning?
- Meaning as an open ended conceptual system
- Semantic problems and solutions?
- Information Theory (new!)

**Next Week** Chapter 2: Meaning, Thought and Reality

# Acknowledgments and References

---

- Course design and slides inherit from Nala Lee's HG202 course, back in the depths of time (2009).
- Thanks to Na-Rae Han for inspiration for the student policies (from *LING 2050 Special Topics in Linguistics: Corpus linguistics*, U Penn; adapted).
- Further Reading:
  - Shannon, C.E. (1948), "A Mathematical Theory of Communication", Bell System Technical Journal, 27, pp. 379–423 & 623–656, July & October, 1948. <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>

# Bibliography

---

Leonard Bloomfield. 1933. *Language*. Henry Holt.

J. R. Firth. 1957. *Papers in Linguistics 1934-1951*. OUP.