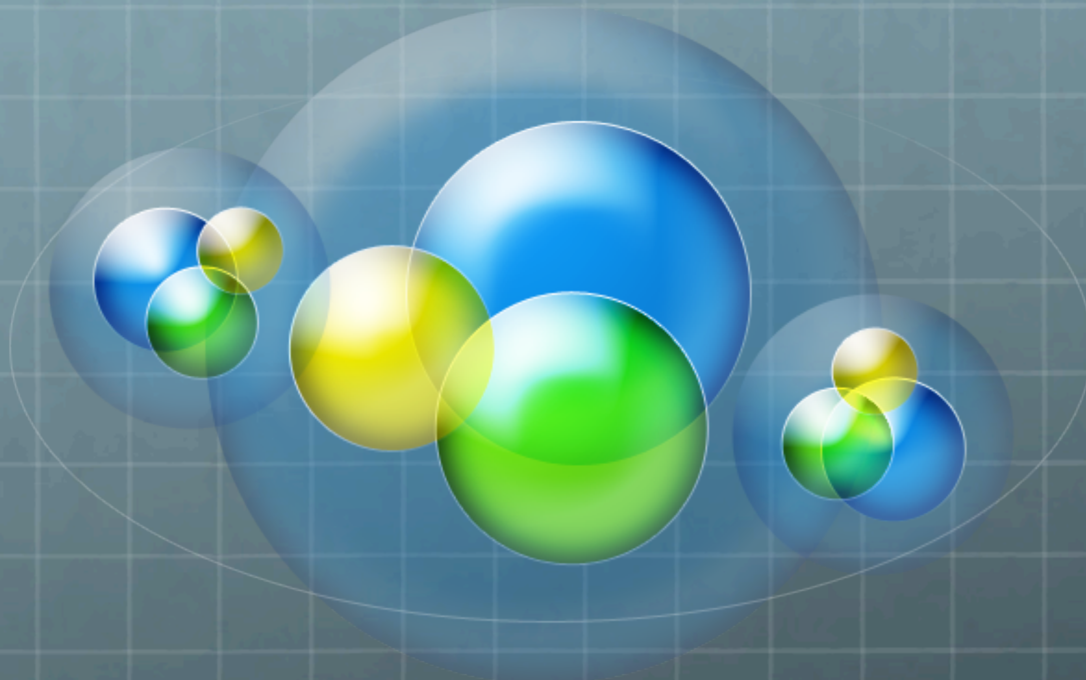


Object-Oriented Semantics:

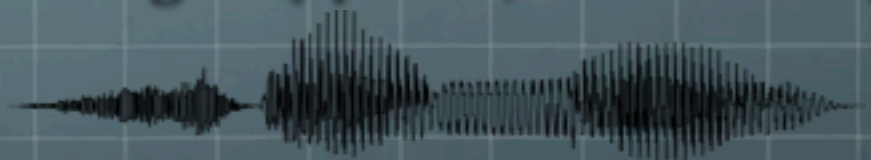
a visual approach to meaning

Daniel Ross with Ed Schade

Nov. 4, 2013 – UIUC Syntax and Semantics Reading Group

“Form and meaning”

- Linguistics is the study of the relationship between form and meaning. We know that. But what exactly does that mean?
- Form: linear signal (spoken, written or signed)



- Meaning: mental representations, among other perspectives. This talk focuses on *mental images.*

Mental Images

- We often think with images (or at least we are under the impression that we do so).
- So metaphorically at least, representing linguistic meaning visually may be useful.
- Other senses are not theoretically excluded, but for practical reasons set aside for now.
- Little previous research has considered a relationship between vision and semantics (but see Cresswell 1983; Arsenijevic 2008; Bolender, Erdeniz & Kerimoğlu 2008; Henderson 2009; Ursini 2011).

Illusions in both domains

“More people have been to Russia than I have.”

“Colorless green ideas sleep furiously.”

Biological basis?

- Language is often seen as a critical difference between humans and other animals.
- Vision occupies a very large part of the brain.
  - In itself, vision is impressive and shares properties with language (such as underdetermined input).
  - We can assume vision is complex.
- It is not inconceivable that there is some relationship, either neural or evolutionary, between the systems.

Images and meaning: a computational benefit

- Computers do not talk, and while they can to some degree work with formal semantic systems, pragmatics is a major obstacle in applying the results.
- However, computers can generate images.

car ←→ 

- By using images to represent meaning, a computer can create relevant output for humans and demonstrate, by application, that parsing works.

Images and meaning: a computational benefit

- If accurate, the generated image is one possible representation of the meaning of the linguistic form.
- This simulates pragmatic variation: much of the generated image is not determined by the input, but certain aspects are.
- The image is not *the* meaning, but *a* meaning. Consider it to be a Possible World. *The* meaning is hidden inside the program: it is the relationship between linguistic form and possible images.

Images and meaning: a computational benefit

- The constraints imposed by the linguistic input and rules in the program guarantee a semantically appropriate image, but never the same image:
- Pragmatic variation is simulated by random values.
- The semantic content is represented by what is consistent across all of the potential images.
- But what does linguistic input actually contribute to the image output?

An experiment

Describe what you see,
and draw what you read.

- How much information is conveyed by ‘translation’ from images to language or language to images?
- If a computer performed at human-like levels, what would it be expected to do?
- Goal: generate an objective corpus of meaning.
- Result: pairs of images and descriptions.

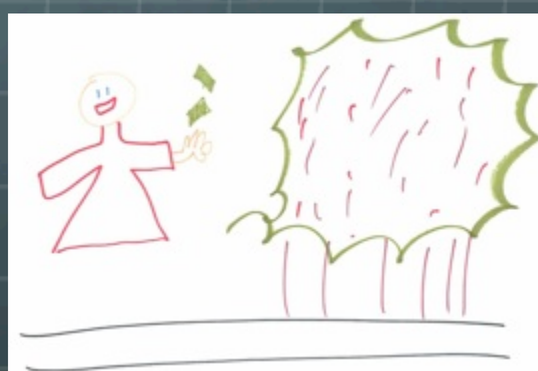
Methodology

- Each subject saw one set of five video clips.
  - For each, they drew an image and wrote a paragraph.
- Then they saw five paragraphs and drew images.
  - These were descriptions of the other set of five videos they had not seen, as described by other subjects.
- From this, we can analyze what information is conveyed by the paragraph by comparing the original drawing to the drawing from the paragraph.

Video stimulus

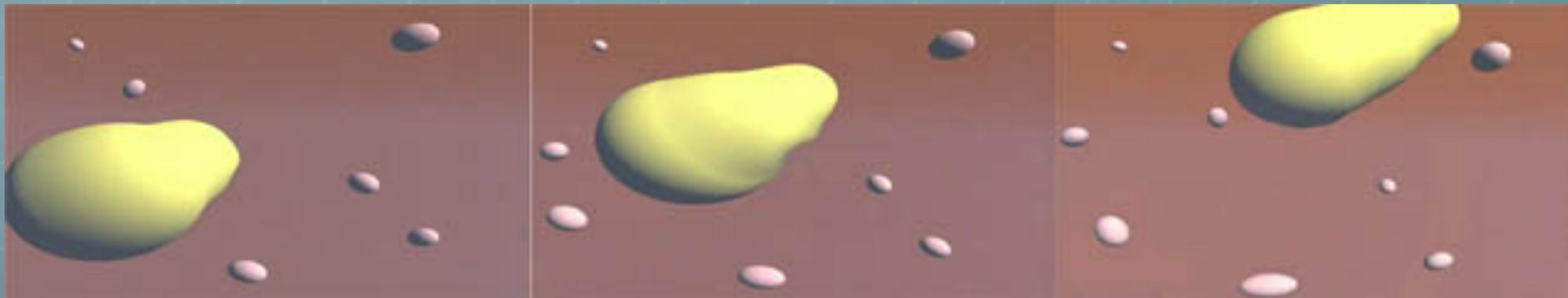
- Subjects were encouraged to use descriptive language rather than rely on assumptions or shared background knowledge.
- The videos ranged from concrete to abstract in order to include unusual images in the study and avoid reliance on pragmatics.
- Drawings were compared to each other to eliminate factors related to ignoring details in the description or artistic ability.

"A young boy is standing in front of withering shrubbery on a seemingly autumn day. The clue as to the season is given not from the sky, but rather the autumn colors of the shrubbery; colors such as green with some scattered brown. The boy is wearing a red winter jacket as he outstretches his hand in an attempt to catch money that is falling down on him. The expression on his face is one of happiness and excitement."

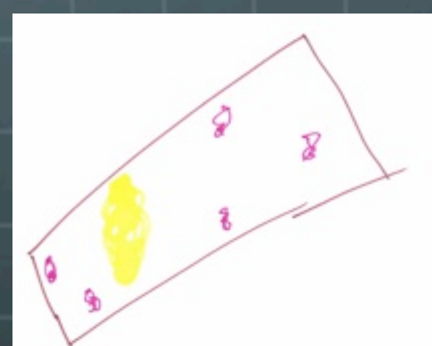


Interpreting the Results

- Similarities represent inherent semantic content.
- Differences represent variable pragmatic content.
- We can approximate concepts (that is, mental images) from the drawings, with a large enough corpus.
- The approach represents Possible World Semantics:
  - each image is a possible world, varying with pragmatics, and the shared properties of all of the images together, the set of possible worlds, determine the semantics.



"A big yellow blob oozing diagonally across a light brown background that also has small pink blobs moving randomly around. "



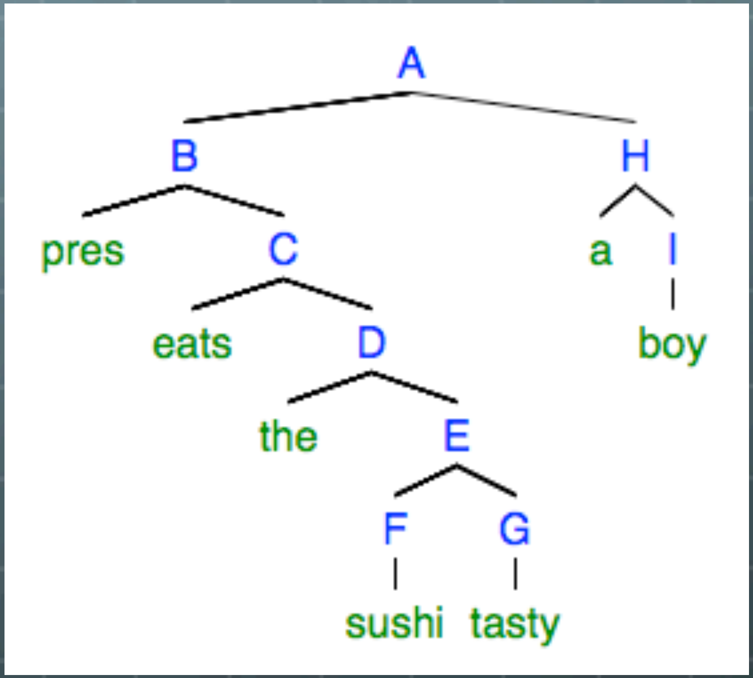
Objects in Semantics

- The process of drawing an image based on a description involves interpreting objects and events and relationships between them.
- This is the core of semantics: a hierarchical structure involving objects and relationships.
- By using a certain color in the drawing, the subjects are modifying the object they are drawing: they are applying a property, and establishing a relationship between that color and the object.

"An explosion of white light with purple highlights against a solid black background. The explosion of light forms a shape with numerous arms of light extending from the middle."

Semantic Structure

```
A:B H
B:pres C
C:eats D
D:the E
E:F G
F:sushi
G:tasty
H:the I
I:boy
```



"A boy eats the tasty sushi."

Computational Implementation

- The current version is written in PHP (mainly a web design language), and has a limited lexicon but is fully recursive within that.
- Images are generated from input sentences/phrases.
- The program consists of:
  - A very basic syntactic parser (rule-based linear parse).
  - Object-oriented semantic structures.
    - Modification function allows binary branching.
  - Lexicon (class, argument structure, images, etc.).

Goals

- Every image conforms to semantics, but pragmatic variation is simulated by random aspects:
  - Varied images for lexical items, and changing background for context. (And more in the future.)
  - Images are from Google Image Search: they represent world-knowledge and lexical semantics.
- Truth-conditions are not generated; instead, the set of all possibly generated images represent this.
- Compositionality is based on modification.
- Object-Oriented Semantics: structured meanings.

Modification

- Modification and argument selection are based on the same principles.
- Each node is a modifier or a head.
  - The head goes on the left.
  - English word order is discarded (via syntactic parser).
- Any modifiers or arguments go on the right.
  - Lexical semantics (whether there is an open argument) determines whether it is a modifier or argument.

Lexical entry: ‘bike’

```
if ($h=='bike') {
    $n->wordclass = 'N';
    $n->args = array();
    $n->img = 'bike';
    $n->m = array();
}
```

Lexical entry: ‘a’

```
if ($h=='a') {
    $n->wordclass = 'D';
    $n->args = array(
        array('N','copy')
    );
    $n->img = FALSE;
    $n->m = array();
}
```

Lexical entry: ‘green’

```
else if ($h=='green') {
      $n->wordclass = 'A';
      $n->args = array();
      $n->img = 'green';
      $n->m = array('N','texture');
}
```

Lexical Entries

- Word class is specified for syntactic reasons (unimportant in semantics for the most part).
- Arguments (if any) are included so it can “look for” other nodes that fill those argument places.
- Images are assigned here also (future: externalize).
- Modification properties determine (if any) how this lexical entry could potentially modify another node when combined with it (inverse argument selection).

Lexical entry: ‘eats’

```
if ($h=='eats') {
        $n->wordclass = 'V';
        $i = mt_rand(1,3); //select one of three possible schemas below
        if ($i==1) { $s1 = 100; $x1 = 150; $y1 = 230; $s2 = 150; $x2 = 150; $y2 = 80; }
        if ($i==2) { $s1 = 200; $x1 = 160; $y1 = 275; $s2 = 150; $x2 = 150; $y2 = 80; }
        if ($i==3) { $s1 = 50; $x1 = 55; $y1 = 75; $s2 = 200; $x2 = 200; $y2 = 150; }
        $n->args = array(
                array('D','put',$s1,$x1,$y1,0)//directobjectforimg,N,expectedfirst
        );
        $n->img = 'eat'.$i; //select the image for the schema selected above
        $n->m = array(
                array('T','givearg',
                        array('D','put',$s2,$x2,$y2,0) //argumenttogivetoaTnode
                )
        );
}
```

Object-Oriented

- Semantics can be seen as the study of the way objects are assigned properties and related to one another.
- A “flat” meaning at the top of the tree isn’t the full picture.
- Object-Oriented Programming creates objects that can be modified, rather than procedurally processing based on functions.
- Syntactic theory is inspired by on pre-Object-Oriented programming.

Object-Oriented

- Basic Procedural:

```
$x=2;
$y = $x*$x;
print $y; //4
```

- Functional

```
$x=2;
$y = square($x);
print $y; //4
```

- Object-Oriented

```
$x = new Number;
$x->value = 2;
$x->square();
print $x->value; //4
```

Results

- Currently the program only has the following lexical items defined:

Nouns — *people*: boy, girl; *food*: burger, salad, sushi;
clothing: hat, jacket, sock; *vehicles*: bike, car
Adjectives — red, green, striped
Verbs — eats, sleeps, wears
Determiners — a, the

- All possible syntactic structures formed with them [excluding nominal compounds] can be parsed.

Results

- "The boy eats the sushi."

Results

- "The boy eats the sushi."



Results: variation

- “The boy eats the sushi.”







Results

- At the moment, a limited number of images are defined in each lexical entry, but that is not a theoretical limitation.
- Images are from Google Images. Nouns are represented by objects, and verbs by schemas with coordinates for arguments.
- Random factors simulate pragmatic variation and clearly patterns begin to emerge that represent the core semantic meaning of the sentence.

Results: productivity

- "A girl sleeps."

- "A car sleeps."

Results: no pragmatics

- "The boy wears the red striped jacket."



- Intentionally, no pragmatic filter for expectations.
- Also, modification with adjectives, compositionally.

Results: compositionality

- “A girl eats the bike.”

- “a girl eats”

- “eats the bike”

- “a bike”

Results: recursion

- "a hat"

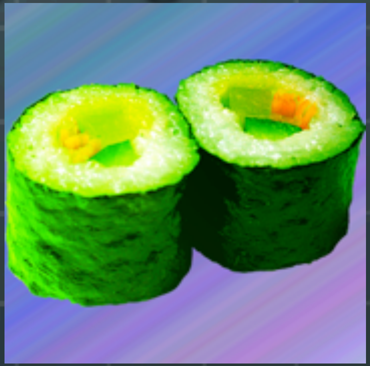
- "a striped hat"

- "green sushi"

- "a red hat"

- "a red striped hat"

- "green green sushi"



Object-Oriented Semantics

- Every node is a separate object.
  - Each can be analyzed independently with connections.
  - The (relevant) original structure is preserved.
- This is preferable for several reasons, such as providing the entire parsed, structured sentence at the pragmatics interface, in addition to establishing the foundation for a theory of discourse, which can be thought of as the process of manipulating or modifying parts of mental images.

Future Directions

- (Automatically) add new lexical items.
  - For example, search for them on Google Images.
  - Likewise, include idioms and collocations as phrases by not treating them compositionally when pragmatic information (=Google Images) provides enough information for the whole phrase.
- Operators (conjunctions, quantifiers, negation, etc.)
- Statistical methods to simulate pragmatics based on frequency.

Future Directions

- Improved graphics.
  - Image recognition would allow inverse (generating sentences from images).
  - 3D graphics would allow manipulation of visual objects, rather than pixel-based images.
- Eventually, the theory/program should be applicable to any input/output.
- Additionally, gathering more experimental results would be be useful to plan the ideal program.

Bibliography

Arsenijevic, B. (2008). From Spatial Cognition to Language. *Biolinguistics*, *2*(1), 003–023.

Bolender, J., Erdeniz, B., & Kerimoğlu, C. (2008). Human Uniqueness, Cognition by Description, and Procedural Memory. *Biolinguistics*, *2*(2-3), 129–151.

Cresswell, M. J. (1983). A highly impossible scene. The semantics of visual contradictions. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, Use, and Interpretation of Language* (pp. 62–78). Walter de Gruyter.

Henderson, B. (2009). Symmetry in Visual and Linguistic Perception. *Biolinguistics*, *2*(4), 291–316.

Ursini, F.-A. (2011). Space and the Vision–Language Interface: A Model-Theoretic Approach. *Biolinguistics*, *5*(3), 170–225.