Language Acquisition
Fall 2010/Winter 2011

Syntax-Semantics Interface

Afra Alishahi, Heiner Drenhaus
Computational Linguistics and Phonetics
Saarland University
Linking Syntax to Semantics

• How is the surface structure (syntax) linked to the underlying meaning (semantics)?
  • **Alternative 1:** syntax is learned independently of semantics; later the two are linked together
  • **Alternative 2:** syntax and semantics are learned simultaneously

• Central unit: **verb argument structure**
  • Relationship btw the semantics of a verb and its syntactic form
    • Number and type of the arguments that the verb takes
    • Semantic roles that the arguments receive in an event
    • Syntactic realization of the verb and its arguments
How to Convey a Relational Meaning?

chimp

apple

eat

The chimp is eating an apple
How to Convey a Relational Meaning?

[Fisher’94]

This is blicking!

The rabbit is blicking the duck
How to Convey a Relational Meaning?

She is dropping the vase.
The vase is falling.
*She is falling the vase.
AGENT *is verbing* THEME

- running
- looking
- reading
- drinking
- eating
- breaking
- rolling
- dancing
- falling
THEME is VERBing

- running
- drinking
- breaking
- looking
- reading
- eating
- rolling
- dancing
- falling
Acquisition of Verb Argument Structure

• General patterns
  • Young children are sensitive to argument structure regularities

  \[\text{bunny gorped duck} \implies \text{causal action?}\]
  \[\text{kitty blicked down the street} \implies \text{manner of motion?}\]

• Idiosyncrasies
  • Semantically similar verbs can have different syntactic behaviour

  \[\text{I filled the glass with water, } *\text{I filled water into the glass}\]
  \[*\text{They loaded the truck with hay, They loaded hay into the truck}\]

• A U-shaped behavioural pattern is observed for children’s argument structure acquisition
Semantic Bootstrapping

• Semantic Bootstrapping (Pinker, 1984)

• Syntactic behaviour of a verb is innately determined by the decompositional representation of its meaning

  Agent is 1st argument of **CAUSE**

  Patient is 2nd argument of **CAUSE**

  Theme is 1st argument of **GO** and **BE**

• With the innate knowledge of mapping between semantics and syntax, a child can predict the correct mapping once she knows what a verb means
Argument Structure Constructions

• Construction Grammar
  • Meaning may be directly associated with syntactic forms
  • Lakoff 1987, Fillmore et al. 1988, Langacker 1999

• Argument structure construction (Goldberg, 1995)
  • A mapping between underlying verb-argument relations and the syntactic form that is used to express them

Subj V Obj Obj2 ⇔ X cause Y receive Z
Example: Pat faxed Bill the letter.

Subj V Oblique ⇔ X move Y
Example: The fly buzzed into the room.
How are Constructions Learned?

• Tomasello (1991):
  • Argument structure patterns are acquired on a verb-by-verb basis
  • Abstract constructions learned through categorization and generalization of common patterns

• Goldberg (1995):
  • Constructional meaning is formed around the meanings of highly frequent light verbs
  • E.g., the construction "Subj V Oblique" paired with the meaning "X moves Y" corresponds to the light verb go
Computational Modeling of Constructions

- FrameNet (Baker, Fillmore, Low, 1998)
  - A database of lexical constructions (or frames)

- The acquisition of constructions
  - Learning lexical constructions via structure mapping (Chang, 2004)
  - Learning verb meaning from image data (Dominey, 2003; Dominey & Inui, 2004)
  - Learning abstract constructions from verb usage data (Alishahi & Stevenson, 2008)
Case Study: Chang (2004)

- Learning lexical-based constructions from child-directed data
  - Goal: learning associations between form relations (word order) and meaning relations (role-filler bindings)
  - Search space: grammars defined by a unification-based formalism (Embodied Construction Grammar, ECG)
  - Form and meaning representations: subgraphs of elements and relations among them
  - Construction representation: a mapping between two subgraphs
  - Learning task: finding the best grammar to fit the observed data
Case Study: Chang (2004)

- Learning lexical-based constructions from child-directed data

**Construction** THROW-BALL

**Constituents**
- $t_1$: THROW
- $t_2$: BALL

**Form**
- $t_1 f$ before $t_2 f$

**Meaning**
- $t_1 m. \text{throwee}$, $t_2 m$
Case Study: Chang (2004)

- The model makes generalizations at the lexical level:

  \[
  \text{construction } \text{THROW-BLOCK} \\
  \text{constituents} \\
  t1: \text{THROW} \\
  t2: \text{BLOCK} \\
  \text{form} \\
  t1_f \text{ before } t2_f \\
  \text{meaning} \\
  t1_m.\text{throwee } t2_m
  \]

  \[
  \text{construction } \text{THROW-BALL} \\
  \text{constituents} \\
  t1: \text{THROW} \\
  t2: \text{BALL} \\
  \text{form} \\
  t1_f \text{ before } t2_f \\
  \text{meaning} \\
  t1_m.\text{throwee } t2_m
  \]

  \[
  \text{construction } \text{THROW-OBJECT} \\
  \text{constituents} \\
  t1: \text{THROW} \\
  t2: \text{OBJECT} \\
  \text{form} \\
  t1_f \text{ before } t2_f \\
  \text{meaning} \\
  t1_m.\text{throwee } t2_m
  \]
The model makes generalizations at the lexical level:

- **The model makes generalizations at the lexical level:**

  
  ![Graph showing percent correct bindings against percent training examples encountered for different verbs (drop, throw, fall)].

  The graph illustrates the comprehension over time for different verbs: drop (n=10, b=18), throw (n=25, b=45), and fall (n=53, b=86). The x-axis represents the percent training examples encountered, while the y-axis shows the percent correct bindings.

  **Case Study: Chang (2004)**

  - The model first learned item-specific representations and the current set of constructions, which provides support for the verb island hypothesis. The results indicate that the model is able to acquire useful item-based constructions, and there is evidence that multi-word constructions could be learned using the same approach.

  A quantitative measure of comprehension over time, shown in Figure 8. The graph reflects the incremental improvement in comprehension. To the extent possible, the assumptions in which meaning in context plays a pivotal role lend support to the verb island hypothesis.

  Differences in verb learning trajectories are interesting to note, as the learning model shows distinct trajectories for the different verbs. For example, throwing is pragmatically more likely to be done on command than throwing, and misses the association of an imperative speech act with the lack of an expressed agent, providing a natural formulation of the grammar learning problem.
Case Study: Alishahi & Stevenson (2008)

- A Bayesian model of early argument structure acquisition
  - Each verb usage is represented as a set of features

<table>
<thead>
<tr>
<th>head verb</th>
<th>eat</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb semantic primitives</td>
<td>[act, consume]</td>
</tr>
<tr>
<td>argument roles</td>
<td>&lt;Agent, Theme&gt;</td>
</tr>
<tr>
<td>argument categories</td>
<td>&lt;human, food&gt;</td>
</tr>
<tr>
<td>syntactic pattern</td>
<td>arg1 verb arg2</td>
</tr>
</tbody>
</table>
• A Bayesian model of early argument structure acquisition
  • Each verb usage is represented as a set of features
  • Each construction is a cluster of verb usages
  • A probability distribution over feature values
  • The best construction is found for each new usage through a Bayesian approach

Case Study: Alishahi & Stevenson (2008)
Sample Constructions

• Verb semantic primitives for Transitive Construction:

<table>
<thead>
<tr>
<th>Simulation</th>
<th>50 input pairs</th>
<th>500 input pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation 1</td>
<td>[Diagram]</td>
<td>[Diagram]</td>
</tr>
<tr>
<td>Simulation 2</td>
<td>[Diagram]</td>
<td>[Diagram]</td>
</tr>
<tr>
<td>Simulation 3</td>
<td>[Diagram]</td>
<td>[Diagram]</td>
</tr>
<tr>
<td>Simulation 4</td>
<td>[Diagram]</td>
<td>[Diagram]</td>
</tr>
<tr>
<td>Simulation 5</td>
<td>[Diagram]</td>
<td>[Diagram]</td>
</tr>
</tbody>
</table>
Sample Constructions

Verb semantic primitives for Intransitive Construction:

<table>
<thead>
<tr>
<th>Simulation</th>
<th>50 input pairs</th>
<th>500 input pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation 1</td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
<tr>
<td>Simulation 2</td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
<tr>
<td>Simulation 3</td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
<tr>
<td>Simulation 4</td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
<tr>
<td>Simulation 5</td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
</tbody>
</table>

Note: Darker squares correspond to higher probabilities.
First eight usages of *fall* by Adam (CHILDES) and by one of the simulations of our model

<table>
<thead>
<tr>
<th>Adam</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>go fall!</em></td>
<td><em>John fall ball</em></td>
</tr>
<tr>
<td><em>no no fall no!</em></td>
<td><em>toy fall</em></td>
</tr>
<tr>
<td><em>no fall!</em></td>
<td><em>Mary fall book</em></td>
</tr>
<tr>
<td><em>oh Adam fall.</em></td>
<td><em>toy fall</em></td>
</tr>
<tr>
<td><em>Adam fall toy.</em></td>
<td><em>cookie fall</em></td>
</tr>
<tr>
<td><em>Adam fall toy.</em></td>
<td><em>kitty fall</em></td>
</tr>
<tr>
<td><em>oh fall.</em></td>
<td><em>spoon fall</em></td>
</tr>
<tr>
<td><em>I not fall.</em></td>
<td><em>ball and toy fall</em></td>
</tr>
</tbody>
</table>
Learning Curves

- Learning phases are successfully simulated:
  - Imitation
  - Overgeneralization and recovery
  - Productive generalization

![Graph showing learning curves]

Fig. 13. Cumulative accuracy of the model for predicting syntactic pattern.

Note: Solid lines show performance with noiseless input data, and dashed lines show performance with noisy input data.

With testing of individual verbs, the variability in the performance of the model early on is more pronounced, due to variations in exposure to that verb in the random corpora. For frequent verbs with various argument structures, such as *go*, a U-shaped curve is often observed. The learning curve for *fall*, which is less frequent, shows a delay compared to more frequent verbs such as *go*, but the U-shaped curve can still be observed in most simulations.

Figures 13 and 14 show that model generally exhibits the characteristics of a U-shaped learning curve observed for children. In the following subsections, we turn to a more detailed description of the behavior of the model as it uses the knowledge it has acquired at various stages of learning along the “U”: specifically, during generalization (including an initial stage of imitation), possible over-generalization and recovery, and productive generalization.
Verb Semantic Roles

- Semantic (thematic) roles indicate the relations of the participants in an event to the main predicate.

\[ \text{Pat gave the hammer to Matt.} \]

\[ \text{Give}_{cause,possess}(\text{Pat, Hammer, To(Matt)}) \]

Diagram:
- Subject
- Direct Object
- PP Phrase
- Agent
- Theme
- Recipient
Theoretical Questions

• What is the nature of semantic roles?
  • Traditional view: roles are atomic and universal, such as Agent, Theme, Goal, … (e.g., Jackendoff 1990)
  • Proto-role Hypothesis (Dowty, 1991): roles are a set of properties, such as volitional, affecting, animate

• Where do they come from?
  • Traditional view: roles and their link to syntactic positions are innate (e.g., Pinker 1989)
  • Alternative view: they are gradually learned from verb usages (e.g., Tomasello 2000)
AGENT is VERBing THEME

- running
- looking
- reading
- eating
- dancing
- breaking
- rolling
- falling
AGENT is VERBing THEME

reader is reading text
eater is eating food
drinker is drinking liquid
Learnability of Semantic Roles

- Usage-based account: verb-specific roles change to general roles over time

- Experimental evidence confirms that access to general roles such as Agent and Theme is age-dependent (Shayan & Gershkoff-Stow, 2007)
Linking Semantic Roles to Grammatical Functions

• Semantic roles are linked to syntactic positions early on
  • Children are sensitive to the association between semantic roles (e.g. Agent) and grammatical functions (e.g. Subject) from an early age (Fisher 1994, 1996; Nation et al., 2003)

• Nativist account
  • Innate linking rules that map roles to sentence structure enable children to infer associations between role properties and syntactic positions (e.g., Pinker, 1989)
Computational Studies of Roles

• Assignment of general pre-defined roles to sentence constituents
  • E.g., McClelland and Kawamoto (1986), Allen (1997)

• Role learning
  • Learning verb-specific roles from annotated data (Chang 2004)
  • Discovering relational concepts from unstructured examples (Kemp et al., 2006; Doumas et al., 2008)
  • Acquiring semantic profiles for general roles from verb usages (Alishahi & Stevenson, 2008)
Case Study: Allen (1997)

- A connectionist model of thematic role assignment
  - Integrates syntax, semantics and lexical information
  - Is trained on usages of most frequent verbs in CHILDES
  - Predicts semantic properties of verbs and nouns
  - Simulates grammaticality judgment

![Diagram](Image)
Role Representation

Each basic role was elaborated by a set of subroles, or proto-role properties:

- **Cause**
  - patient
  - change of state
  - motion
  - travel
  - location
  - experiencer
  - possessor
  - instrument
  - path

**Causation Subtypes**
- apply force
- action
- direct cause
- allow
- help
- impede
- instrumental
- author
- agent
- internal cause
Semantic Prediction & Grammaticality

Figure 4: Violation of what is traditionally known as the double object form during traveling. Comparing the nets performance with the verb kick results in a mis-

The net shows an interesting capacity to generalize such that some features are more important than others which do alternate from those that do not. Why does the net discriminate between and with its performance with the verb kick in the double object form produces that assignment associated with a particular set of role assignments, rather than about any specific word or words. 

Pro-

viding feedback concerning the meaning of the utter-

In addition to organizing the semantic space to al-

influences how a verb is used.

There are two well known attempts to incorporate formation, such that the set of forms that a verb is used with only the number of arguments in the novel sentence.

Insofar as that outlined in most theories of acquisition, insofar as able the intended meaning of some utterances, and it includes the assumption that the learner has avail-

John carry Mary basket

OUTPUT

John Kick Mary Ball

INPUT

Verb Input

Noun Input

OUTPUT

Argument Roles Computed

Roles Computed

Core Semantics Computed

Time Step

Time Step

123456789101112131415161718

123456789101112131415161718

15161718
Case Study: Alishahi & Stevenson (2010)

• A Bayesian model of early verb learning can learn
  • general conceptions of roles based only on exposure to individual verb usages
  • associations between general semantic properties and the syntactic positions of the arguments

• The acquired semantic roles
  • naturally metamorphose from verb-specific to general properties
  • are an intuitive match to the expected properties of various roles
  • are useful in guiding comprehension in the face of ambiguity
Distributional Representations of Roles

Lexical properties:
{living thing, animal, chimp, ...}

Event-based properties:
{volitional, affecting, animate, ...}

The chimp is eating an apple

Number of arguments: 2

Lexical properties:
{entity, object, fruit, ...}

Event-based properties:
{non-independently exist, affected, change, ...}

Event (Verb): Eat

Event properties:
action, consumption
Event-based Properties of Transitive Arguments

### Argument 1 (Agent)

<table>
<thead>
<tr>
<th>Probability</th>
<th>Event-based property</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.048</td>
<td>independently exist</td>
</tr>
<tr>
<td>0.048</td>
<td>sentient</td>
</tr>
<tr>
<td>0.035</td>
<td>animate</td>
</tr>
<tr>
<td>0.035</td>
<td>change</td>
</tr>
<tr>
<td>0.035</td>
<td>affected</td>
</tr>
<tr>
<td>0.035</td>
<td>change emotional</td>
</tr>
<tr>
<td>0.035</td>
<td>becoming</td>
</tr>
<tr>
<td>0.013</td>
<td>volitional</td>
</tr>
<tr>
<td>0.013</td>
<td>possessing</td>
</tr>
<tr>
<td>0.013</td>
<td>getting</td>
</tr>
</tbody>
</table>

### Argument 2 (Theme)

<table>
<thead>
<tr>
<th>Probability</th>
<th>Event-based property</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.086</td>
<td>state</td>
</tr>
<tr>
<td>0.031</td>
<td>independently exist</td>
</tr>
<tr>
<td>0.031</td>
<td>change</td>
</tr>
<tr>
<td>0.031</td>
<td>change possession</td>
</tr>
</tbody>
</table>
Lexical Properties of Transitive Arguments

<table>
<thead>
<tr>
<th>Probability</th>
<th>Lexical property</th>
<th>Probability</th>
<th>Lexical property</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.054</td>
<td>entity</td>
<td>0.056</td>
<td>entity</td>
</tr>
<tr>
<td>0.040</td>
<td>object</td>
<td>0.037</td>
<td>object</td>
</tr>
<tr>
<td>0.040</td>
<td>physical object</td>
<td>0.037</td>
<td>physical object</td>
</tr>
<tr>
<td>0.026</td>
<td>being</td>
<td>0.023</td>
<td>unit</td>
</tr>
<tr>
<td>0.026</td>
<td>organism</td>
<td>0.023</td>
<td>artifact</td>
</tr>
<tr>
<td>0.026</td>
<td>living thing</td>
<td>0.023</td>
<td>artefact</td>
</tr>
<tr>
<td>0.026</td>
<td>animate thing</td>
<td>0.023</td>
<td>whole</td>
</tr>
<tr>
<td>0.015</td>
<td>person</td>
<td>0.023</td>
<td>whole thing</td>
</tr>
<tr>
<td>0.015</td>
<td>individual</td>
<td>0.018</td>
<td>abstraction</td>
</tr>
<tr>
<td>0.015</td>
<td>someone</td>
<td>0.014</td>
<td>being</td>
</tr>
<tr>
<td>0.015</td>
<td>somebody</td>
<td>0.014</td>
<td>organism</td>
</tr>
<tr>
<td>0.015</td>
<td>mortal</td>
<td>0.014</td>
<td>living thing</td>
</tr>
<tr>
<td>0.015</td>
<td>human</td>
<td>0.014</td>
<td>animate thing</td>
</tr>
<tr>
<td>0.015</td>
<td>soul</td>
<td>0.014</td>
<td>person</td>
</tr>
<tr>
<td>0.015</td>
<td>causal agent</td>
<td>0.014</td>
<td>individual</td>
</tr>
</tbody>
</table>

Figure 5. Semantic profiles of argument positions Agent and Theme in a transitive construction.
Learning Curves for Semantic Profiles

To observe the trend of moving from a more specific to a more general semantic profile for each argument position, we need to compare the semantic profile for an argument position at a given point in learning, and the profile for that position that the model eventually converges to at the end of each simulation. More technically, we need to measure the divergence between the two probability distributions represented by these semantic profiles. We use a standard divergence measure, Relative Entropy, for this purpose.

Relative Entropy \( D(P \| Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \) where \( P \) and \( Q \) are probability distributions.

Figure 9 shows the profile divergence for Subject and Object positions of a transitive construction after every 5 input items over a total of 200 items, averaged over 5 simulations. The divergence between the lexical portion of the profiles is shown by solid lines, and the divergence between the event-based portion of the profiles is shown by dashed lines. Figure 9 shows that the profile for the Subject position (i.e., the Agent) is learned faster than the profile for the Object position (i.e., the Theme), which is a much less constrained role.
Verb Selectional Restrictions

• Most verbs impose semantic restrictions on the arguments that they take, which
  • affect the acceptability of natural language sentences: eating food, drinking water, *eating water, *drinking food
  • facilitate language comprehension and word learning

• Earlier theories view selectional constraints as defining features of the arguments:

  \[ \text{hit} \left( \text{Subj}, \text{Obj} \right) \]

  \( \text{Subj}: \) HUMAN or HIGHER ANIMAL

  \( \text{Obj}: \) PHYSICAL OBJECT

• Identifying necessary and sufficient restrictions is a challenge
Verb Selectional Preferences

- Resnik (1993) proposed an alternative view: verbs have preferences for the type of arguments they allow for
  - World knowledge is represented as a semantic class hierarchy
  - Selectional preferences are viewed as probability distributions over various semantic classes

- Verbs have different degrees of preference
  - e.g. *eat* and *sing* have strong preferences for the direct object position, but *put* and *make* do not
Computational Modeling of Selectional Preferences

• Most of the existing computational models are influenced by the information-theoretic model of Resnik (1993, 1996)
  • Represent preference for an argument position of a verb as a mapping of each semantic class to a real number
  • Model the induction of a verb’s preferences as estimating that number, using a training data set

• Different approach: Erk (2007)
  • Estimate preferences for a head word based on the similarities between that word and other head words observed in a corpus.
Cognitive Modeling and NLP

• Early NLP viewed itself as building models of human understanding

• Modern NLP has shifted emphasis
  • Focus on applications: do limited tasks accurately and robustly, often without real understanding
  • Emphasis on representations, coverage and efficiency, not concerned with cognitive plausibility

• However, cognitive modeling of language is heavily informed by research in NLP
  • Modeling of human language acquisition is influenced by specialized machine learning techniques
Open Questions

• How various aspects of language acquisition interact with each other?
  • Various learning procedures are most probably interleaved (e.g., word leaning and syntax acquisition)
  • Most of the existing models of language acquisition focus on one aspect, and simplify the problem

• How to evaluate the models on realistic data?
  • Large collections of child-directed utterances/speech are available, but no such collection of semantic input
  • A wide-spread evaluation approach is lacking in the community