Embodied Models of Language Learning and Use

Session 4: Embodied language learning

Srini Narayanan and Nancy Chang
{snarayan,nchang}@icsi.berkeley.edu
UC Berkeley / International Computer Science Institute

Course Overview

- **Session 1: Foundations of embodied language**
  - Introduction to NTL: language, neural computation
- **Session 2: Embodied representations**
  - Modeling action and perception
  - Simulative inference
- **Session 3: Language understanding**
  - Construction Grammar
  - Metaphor
- **Session 4: Grammar learning**
  - Modeling child language acquisition

Session 4 outline

1. **Language acquisition: the problem**
2. **Child language acquisition**
3. **Usage-based construction learning model**
4. **Recapitulation:**
   Embodied cognitive models

From single words to complex utterances

FATHER: Nomi are you climbing up the books?
NAOMI: **up.**
NAOMI: **climbing.**
NAOMI: **books.**
1;11.3

MOTHER: what are you doing?
NAOMI: **I climbing up.**
MOTHER: you're climbing up?
2;0.18

FATHER: what's the boy doing to the dog?
NAOMI: **squeezing his neck.**
NAOMI: **and the dog climbed up the tree.**
NAOMI: **now they're both safe.**
NAOMI: **but he can climb trees.**
4;9.3

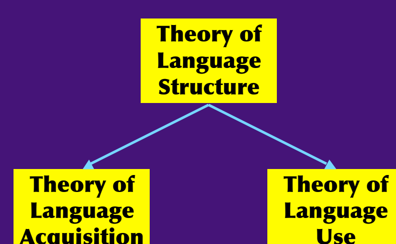
Sachs corpus (CHILDES)

How do they make the leap?

0-9 months
- Smiles
- Responds differently to intonation
- Responds to name and "no"

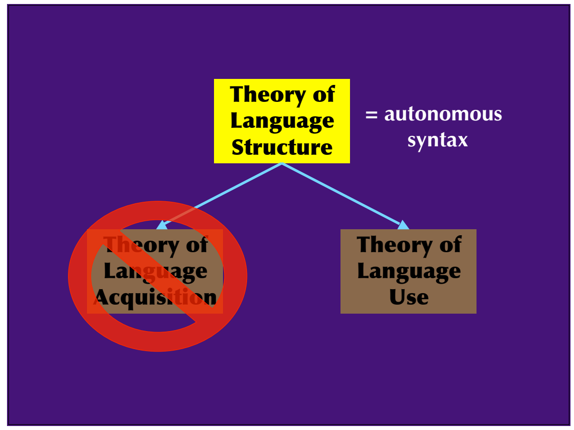
9-18 months
- First words
- Recognizes intentions
- Responds, requests, calls, greets, protests

18-24 months
- agent-object
  - Daddy cookie
  - Girl ball
- agent-action
  - Daddy eat
  - Mommy throw
- action-object
  - Eat cookie
  - Throw hat
- entity-attribute
  - Daddy cookie
- entity-locative
  - Doggie bed



The logical problem of language acquisition

- Gold's Theorem: Identification in the limit
  - No superfinite class of language is identifiable in the limit from positive data only
- The logical problem of language acquisition
  - Natural languages are not finite sets.
  - Children receive (mostly) positive data.
  - But children acquire productive language abilities quickly and reliably, with little overgeneralization!
- One (not so) logical conclusion:
  - **THEREFORE**: there must be strong innate biases restricting the search space
    - Universal Grammar + parameter setting



What is knowledge of language?

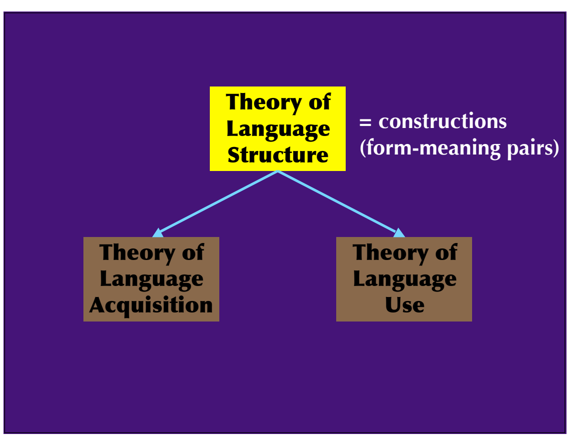
- Basic sound patterns (Phonology)
- How to make words (Morphology)
- How to put words together (Syntax)
- What words (etc.) mean (Semantics)
- How to do things with words (Pragmatics)
- Rules of conversation (Pragmatics)

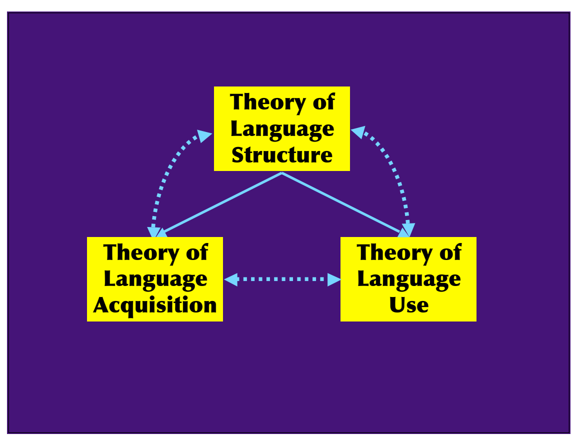
Hypothesis

Grammar learning is driven by
meaningful language use in context.

All aspects of the problem should reflect this assumption:

- Target of learning: a **construction** (form-meaning pair)
- Prior knowledge: rich **conceptual structure**, pragmatic inference
- Training data: pairs of utterances / situational **context**
- Performance measure: success in **communication** (comprehension)

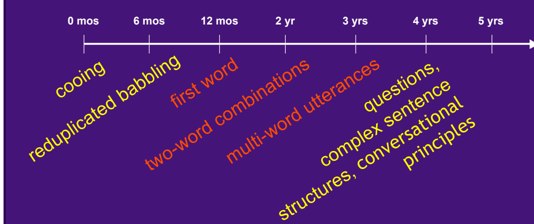




Session 4 outline

1. Language acquisition: the problem
2. Child language acquisition
3. Usage-based construction learning model
4. Recapitulation:
   Embodied cognitive models

The course of development

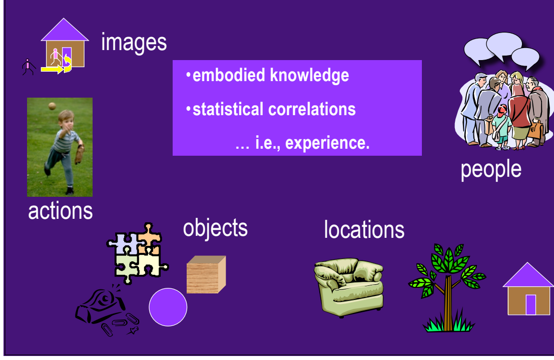


Incremental development

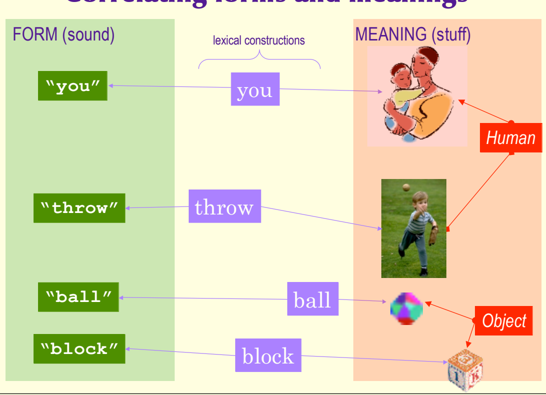
throw	
throw	1;8.0
throw off	1;8.0
I throwded (= *I fell*)	1;10.28
I throw it.	1;11.3
throwing in.	1;11.3
throw it.	1;11.3
throw frisbee.	1;11.3
can I throw it?	2;0.2
I throwed Georgie.	2;0.2
you throw that?	2;0.5
gonna throw that?	2;0.18
throw it in the garbage.	2;1.17
throw in there.	2;1.17
throw it in that.	2;5.0
throwed it in the diaper pail.	2;11.12

fall	
fell down.	1;6.16
fall down.	1;8.0
I fall down.	1;10.17
fell out.	1;10.18
I fell it.	1;10.28
fell in basket.	1;10.28
fall down boom.	1;11.11
almost fall down.	1;11.11
toast fall down.	1;11.20
did Daddy fall down?	1;11.20
Kangaroo fall down	1;11.21

Children in one-word stage know a lot!

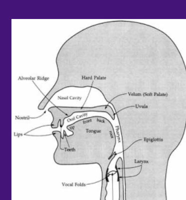


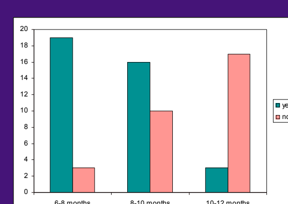
Correlating forms and meanings



Phonology: Non-native contrasts

- Werker and Tees (1984)
- Thompson: velar vs. uvular, /`ki/-/`qi/.
- Hindi: retroflex vs. dental, /t.a/-/ta/





Finding words: Statistical learning

- Saffran, Aslin and Newport (1996)

pretty baby

- /bidaku/, /padoti/, /golabu/
- /bidakupadotigolabubidaku/
- 2 minutes of this continuous speech stream
- By 8 months infants detect the words (vs non-words and part-words)

Language Acquisition

- Opulence of the substrate
  - Prelinguistic children already have rich sensorimotor representations and sophisticated social knowledge
  - intention inference, reference resolution
  - language-specific event conceptualizations

(Bloom 2000, Tomasello 1995, Bowerman & Choi, Slobin, et al.)

- Children are sensitive to statistical information
  - Phonological transitional probabilities
  - Most frequent items in adult input learned earliest

(Saffran et al. 1998, Tomasello 2000)

	cow								
apple	ball								yes
juice	bead		girl				down		no more
bottle	truck		baby	woof	yum	go	up	this	more
spoon	hammer	shoe	daddy	moo	whee	get	out	there	bye
banana	box	eye	momy	choo-choo	uhoh	sit	in	here	hi
cookie	horse	door	boy	boom	oh	open	on	that	no
food	*toys*	*misc.*	*people*	*sound*	*emotion*	*action*	*prep.*	*demon.*	*social*

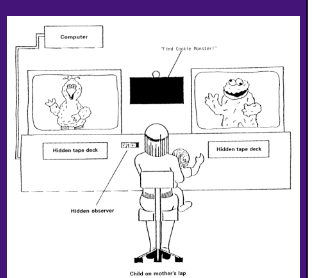
Words learned by most 2-year olds in a play school (Bloom 1993)

Early syntax

- agent + action 'Daddy sit'
- action + object 'drive car'
- agent + object 'Mommy sock'
- action + location 'sit chair'
- entity + location 'toy floor'
- possessor + possessed 'my teddy'
- entity + attribute 'crayon big'
- demonstrative + entity 'this telephone'

Word order: agent and patient

- Hirsch-Pasek and Golinkoff (1996)
- 1;4-1;7
  - mostly still in the one-word stage
  - Where is CM tickling BB?



Language Acquisition

- Basic Scenes
  - Simple clause constructions are associated directly with scenes basic to human experience

(Goldberg 1995, Slobin 1985)

- Verb Island Hypothesis
  - Children learn their earliest constructions (arguments, syntactic marking) on a verb-specific basis

(Tomasello 1992)

throw frisbee	get ball
throw ball	get bottle
⋮	⋮
throw OBJECT	get OBJECT

Children generalize from experience

push3 (force=high) push12 (force=low) • • • push34 (force=?)

Specific cases are learned before general cases.

throw frisbee throw ball • • • throw OBJECT

drop ball drop bottle • • • drop OBJECT

Earliest constructions are **lexically specific (*item-based*)**.
(Verb Island Hypothesis, Tomasello 1992)

Development Of *Throw*

1;2.9	don't throw the bear.	
1;8.0	throw	
	throw off	
1;10.11	don't throw them on the ground.	
1;10.28	I throwded it.	(= I fell)
	I throwded.	(= I fell)
1;11.3	Nomi don't throw the books down.	
	what do you throw it into?	
1;11.3	I throw it.	
	what did you throw it into?	
	I throw it ice.	(= I throw the ice)
1;11.9	they're throwing this in here.	
	throwing the thing.	
	throwing in.	
	throwing.	

Contextually grounded

Parental utterances more complex

Development Of *Throw (cont'd)*

2;0.3	don't throw it Nomi.
	can I throw it?
	I throwed Georgie.
	could I throw that?
	Nomi stop throwing.
2;0.5	throw it?
	well you really shouldn't throw things Nomi you know. remember how we told you you shouldn't throw things.
	you throw that?
2;0.18	gonna throw that?
2;1.17	throw it in the garbage.
	throw in there.
2;5.0	throw it in that.
2;11.12	I throwed it in the diaper pail.

Session 4 outline

1. **Language acquisition: the problem**
2. **Child language acquisition**
3. **Usage-based construction learning model**
4. **Recapitulation: Embodied cognitive models**

How do children make the transition from single words to complex combinations?

- Multi-unit expressions with **relational** structure
  - Concrete word combinations
    - *fall down, eat cookie, Mommy sock*
  - Item-specific constructions (limited-scope formulae)
    - X *throw* Y, *the* X, X*'s* Y
  - Argument structure constructions (syntax)
  - Grammatical markers
    - Tense-aspect, agreement, case

Language learning is structure learning

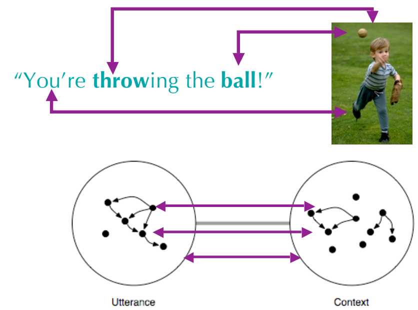
"You're **throw**ing the **ball**!"

- Intonation, stress
- Phonemes, syllables
- Morphological structure
- Word segmentation, order
- Syntactic structure
- Sensorimotor structure
- Event structure
- Pragmatic/informational structure: attention, intention, perspective
- Statistical regularities

Making sense: structure begets structure!

- Structure is cumulative
  - Object recognition → scene understanding
  - Word segmentation → word learning
- **Language learners** exploit existing structure to **make sense** of their environment
  - Achieve **communicative** goals
  - Infer **communicative** intentions

Exploiting existing structure



Comprehension
is
partial.

(not just for dogs)

What we say to kids…

what do you throw it into?
they're throwing this in here.
do you throw the frisbee?
they're throwing a ball.
don't throw it Nomi.

well you really shouldn't throw things Nomi you know.
remember how we told you you shouldn't throw things.

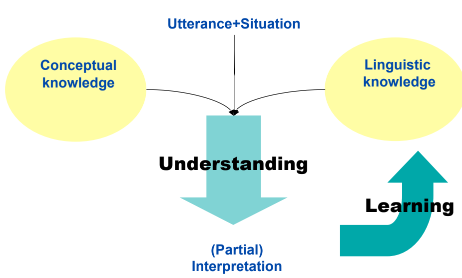
What they hear…

blah blah **YOU THROW** blah?
blah **THROW** blah blah **HERE**.
blah **YOU THROW** blah blah?
blah **THROW** blah blah **BALL**.
DON'T THROW blah **NOMI**.

blah **YOU** blah blah **THROW** blah **NOMI** blah blah.
blah blah blah blah **YOU** shouldn't **THROW** blah.

But children also have rich situational context/cues they can use to fill in the gaps.

Understanding drives learning



Potential inputs to learning

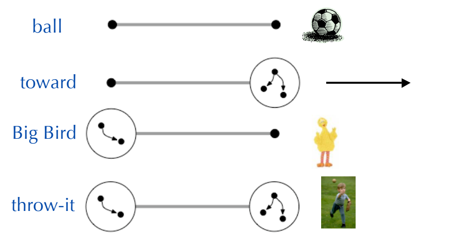
- Genetic language-specific biases
- Domain-general structures and processes
  - Embodied representations
    - …grounded in action, perception, conceptualization, and other aspects of physical, mental and social experience
      Talmy 1988, 2000; Glenberg and Robertson 1999; MacWhinney 2005; Barsalou 1999; Choi and Bowerman 1991; Slobin 1985, 1997
  - Social routines
    - Intention inference, reference resolution
  - Statistical information
    - transition probabilities, frequency effects

Usage-based approaches to language learning
(Tomasello 2003, Clark 2003, Bybee 1985, Slobin 1985, Goldberg 2005)

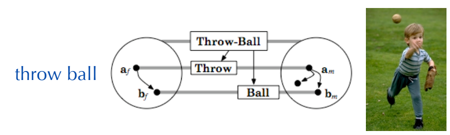
…the opulence of the substrate!

Representation: **constructions**

- The basic linguistic unit is a <form, meaning> pair
  (Kay and Fillmore 1999, Lakoff 1987, Langacker 1987, Goldberg 1995, Croft 2001, Goldberg and Jackendoff 2004)



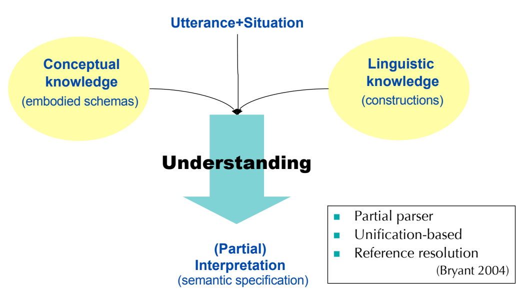
Relational constructions



construction THROW-BALL
constituents
t : THROW
o : BALL
form
t_f before o_f
meaning
t_m.throwee ↔ o_m

Embodied Construction Grammar
(Bergen & Chang, 2005)

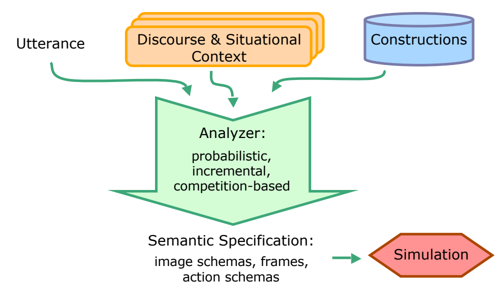
Usage: Construction analyzer



- Partial parser
- Unification-based
- Reference resolution

(Bryant 2004)

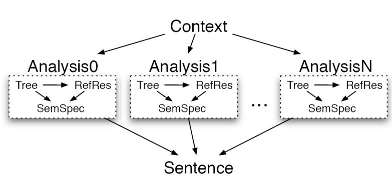
Usage: best-fit constructional analysis



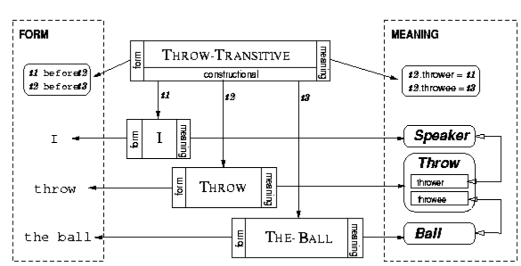
Competition-based analyzer finds the best analysis

- An analysis is made up of:
  - A constructional tree
  - A set of resolutions
  - A semantic specification

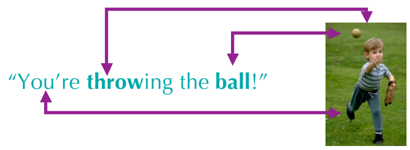
The best fit has the highest combined score



An analysis using THROW-TRANSITIVE



Usage: Partial understanding



ANALYZED MEANING

Participants: ball, Ego

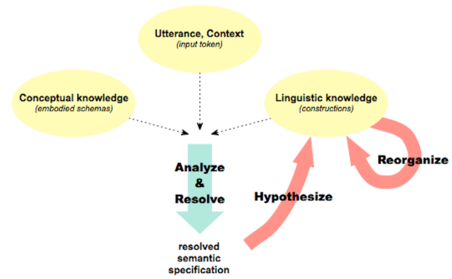
Throw-Action
  thrower = ?
  throwee = ?

PERCEIVED MEANING

Participants: my_ball, Ego

Throw-Action
  thrower = Ego
  throwee = my_ball

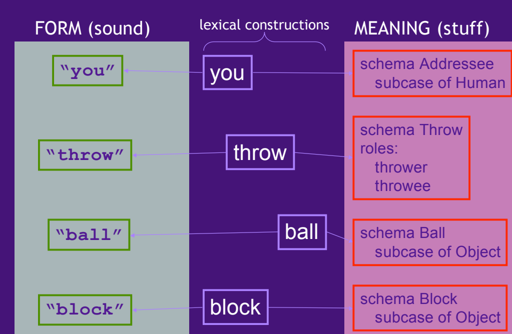
Construction learning model: search



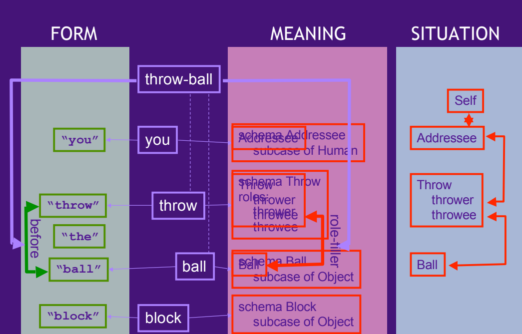
Proposing new constructions

- Relational Mapping — context-dependent
- Reorganization — context-independent
  - Merging (generalization)
  - Splitting (decomposition)
  - Joining (compositon)

Initial Single-Word Stage



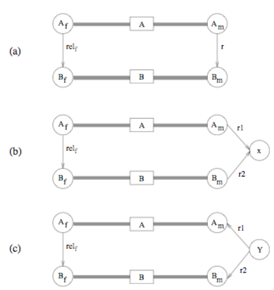
New Data: "You Throw The Ball"



New Construction Hypothesized

construction THROW-BALL
 constructional
  constituents
   t : THROW
   b : BALL
 form
  t_f before b_f
 meaning
  t_m.throwee ↔ b_m

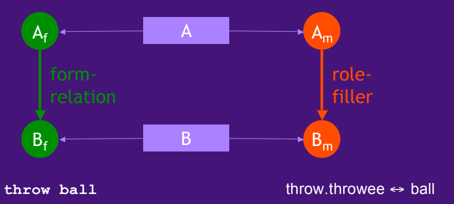
Meaning Relations: pseudo-isomorphism



- strictly isomorphic:
  - B_m fills a role of A_m
- shared role-filler:
  - A_m and B_m have a role filled by X
- sibling role-fillers:
  - A_m and B_m fill roles of Y

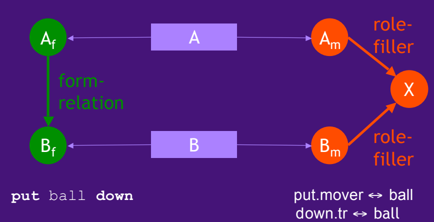
Relational mapping strategies

- **strictly isomorphic:**
  - B_m is a role-filler of A_m (or vice versa)
  - A_m.r1 ↔ B_m



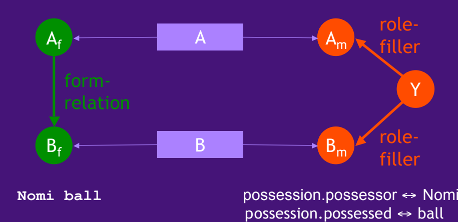
Relational mapping strategies

- **shared role-filler:**
  - A_m and B_m each have a role filled by the same entity
  - A_m.r1 ↔ B_m.r2



Relational mapping strategies

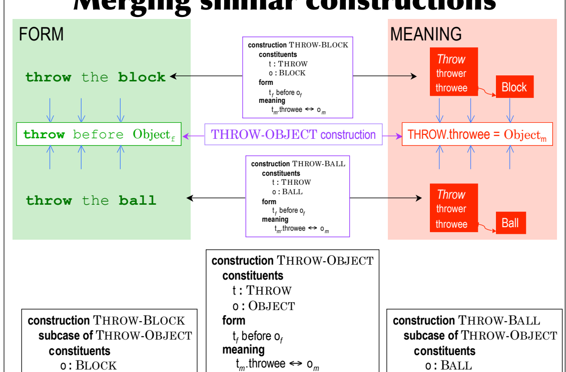
- **sibling role-fillers:**
  - A_m and B_m fill roles of the same schema
  - Y.r1 ↔ A_m, Y.r2 ↔ B_m



Overview of learning processes

- **Relational mapping** — THROW < BALL
  - **throw** the **ball**
- **Merging** — THROW < OBJECT
  - **throw** the **block**
  - **throw**ing the **ball**
- **Joining** — THROW < BALL < OFF
  - **throw** the **ball**
  - **ball off**
  - you **throw** the **ball off**

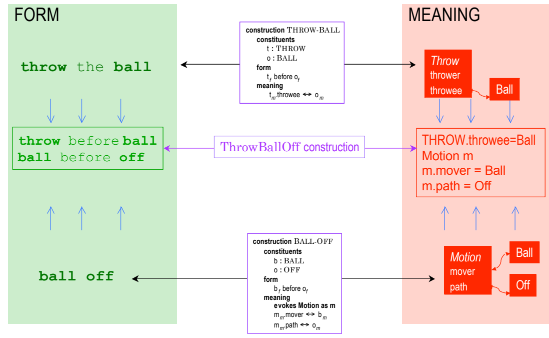
Merging similar constructions



Overview of learning processes

- **Relational mapping**
  - **throw** the **ball** — THROW < BALL
- **Merging**
  - **throw** the **block**
  - **throw**ing the **ball**
  - — THROW < OBJECT
- **Joining**
  - **throw** the **ball**
  - **ball off**
  - you **throw** the **ball off**
  - — THROW < BALL < OFF

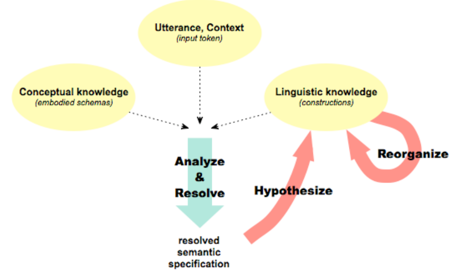
Joining co-occurring constructions



Joined construction

construction THROW-BALL-OFF
constructional
constituents
t : THROW
b : BALL
o : OFF
form
t_f before b_f
b_f before o_f
meaning
evokes MOTION as m
t_m.throwee ↔ b_m
m.mover ↔ b_m
m.path ↔ o_m

Construction learning model: evaluation



Heuristic: minimum description length (MDL: Rissanen 1978)

Learning: usage-based optimization

- **Grammar learning = search for (sets of) constructions**
  - Incremental improvement toward best grammar given the data
- **Search strategy**: usage-driven learning operations
- **Evaluation criteria**: simplicity-based, information-theoretic
  - Minimum description length: most compact encoding of the grammar and data
  - Trade-off between storage and processing

Minimum description length

(Rissanen 1978, Goldsmith 2001, Stolcke 1994, Wolff 1982)

- Seek most compact encoding of data in terms of
  - Compact representation of **model** (i.e., the **grammar**)
  - Compact representation of **data** (i.e., the **utterances**)
- Approximates Bayesian learning (Bailey 1997, Stolcke 1994)
- Exploit tradeoff between preferences for:

smaller grammars	simpler analyses of data
Fewer constructions	Fewer constructions
Fewer constituents/constraints	More likely constructions
Shorter slot chains (more local concepts)	Shallower analyses
Pressure to compress/generalize	Pressure to retain specific constructions

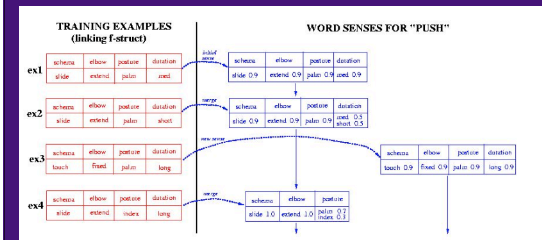
MDL: details

- Choose grammar G to minimize length(G|D):
  - **length(G | D) = *m* • length(G) + *n* • length(D | G)**
  - Bayesian approximation:
    length(G|D) ≈ posterior probability P(G|D)

- **Length of grammar = length(G) ≈ prior P(G)**
  - favor fewer/smaller constructions/roles
  - favor shorter slot chains (more familiar concepts)

- **Length of data given grammar =**
  length(D | G) ≈ likelihood P(D | G)
  - favor simpler analyses using more frequent constructions

Flashback to verb learning: Learning 2 senses of **PUSH**



Model merging based on Bayesian MDL

Experiment: learning verb islands

- **Question:**
  - **Can the proposed construction learning model acquire English item-based motion constructions?** (Tomasello 1992)

- Given: initial lexicon and ontology
- Data: child-directed language annotated with contextual information

Form: text : throw the ball
intonation : falling
Participants : Mother, Naomi, Ball
Scene : Throw
thrower : Naomi
throwee : Ball
Discourse : speaker :Mother
addressee Naomi
speech act : imperative
activity : play
joint attention : Ball

Experiment: learning verb islands

Subset of the CHILDES database of parent-child interactions (MacWhinney 1991; Slobin et al.)

- **coded by developmental psychologists for**
  - form: particles, deictics, pronouns, locative phrases, etc.
  - meaning: temporality, person, pragmatic function, type of motion (self-movement vs. caused movement; animate being vs. inanimate object, etc.)
- **crosslinguistic (English, French, Italian, Spanish)**
  - English motion utterances: 829 parent, 690 child utterances
  - English all utterances: 3160 adult, 5408 child
  - age span is 1;2 to 2;6

Annotated Childes Data

- **765 Annotated Parent Utterances**
- **Annotated for the following scenes:**
  - CausedMotion : "Put Goldie through the chimney"
  - SelfMotion : "did you go to the doctor today?"
  - JointMotion : "bring the other pieces Nomi"
  - Transfer :"give me the toy"
  - SerialAction: "come see the doggie"
- **Originally annotated by psychologists**

An Annotation (Bindings)

- **Utterance: Put Goldie through the chimney**
- **SceneType: CausedMotion**
- **Causer: addressee**
- **Action: put**
- **Direction: through**
- **Mover: Goldie (toy)**
- **Landmark: chimney**

Learning *throw*-constructions

INPUT UTTERANCE SEQUENCE	LEARNED CXNS
1. Don't throw the bear.	throw-bear
2. you throw it	you-throw
3. throw-ing the thing.	throw-thing
4. Don't throw them on the ground.	throw-them
5. throwing the frisbee.	throw-frisbee
MERGE	throw-OBJ
6. Do you throw the frisbee? COMPOSE	you-throw-frisbee
7. She's throwing the frisbee. COMPOSE	she-throw-frisbee

Example learned *throw*-constructions

- Throw bear
- You throw
- Throw thing
- Throw them
- Throw frisbee
- Throw ball
- You throw frisbee
- She throw frisbee
- <Human> throw frisbee
- Throw block
- Throw <Toy>
- Throw <Phys-Object>
- <Human> throw <Phys-Object>

Early talk about *throwing*

Sample input prior to 1;11.9:
don't throw the bear.
don't throw them on the ground.
Nomi don't throw the books down.
what do you throw it into?

Sample tokens prior to 1;11.9:
throw
throw off
I throw it.
I throw it ice. *(= I throw the ice)*

Transcript data, Naomi 1;11.9

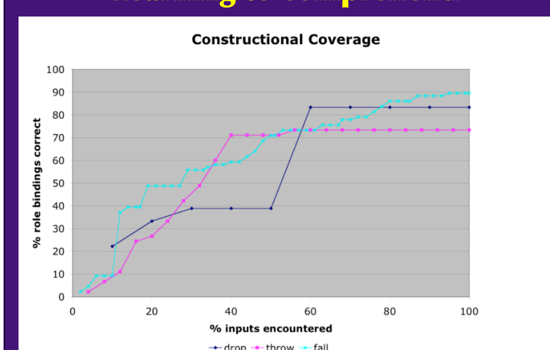
Par: they're throwing this in here.
Par: throwing the thing.
Child: throwing in.
Child: throwing.
Par: throwing the frisbee. …
Par: do you throw the frisbee?
do you throw it?
Child: throw it.
Child: I throw it. …
Child: throw frisbee.
Par: she's throwing the frisbee.
Child: throwing ball.

Sachs corpus (CHILDES)

A quantitative measure: coverage

- **Goal: incrementally improving comprehension**
  - At each stage in testing, use current grammar to analyze test set
- **Coverage = % role bindings correctly analyzed**
- **Example:**
  - Grammar: **throw-ball, throw-block, you-throw**
  - Test sentence: **throw the ball.**
    - Bindings: **scene=Throw**, **thrower=Nomi, throwee=ball**
    - Parsed bindings: **scene=Throw**, **throwee=ball**
  - Score for test grammar on sentence: 2/3 = 66.7%

Learning to comprehend



Principles of interaction

- Early in learning: no conflict
  - Conceptual knowledge dominates
  - More lexically specific constructions (no cost)

throw	want
throw off	want cookie
throwing in	want cereal
you throw it	I want it

- Later in learning: pressure to categorize
  - More constructions = more potential for confusion during analysis
  - Mixture of lexically specific and more general constructions

throw OBJ	want OBJ
throw DIR	I want OBJ
throw it DIR	ACTOR want OBJ
ACTOR throw OBJ	

Experiment: learning verb islands

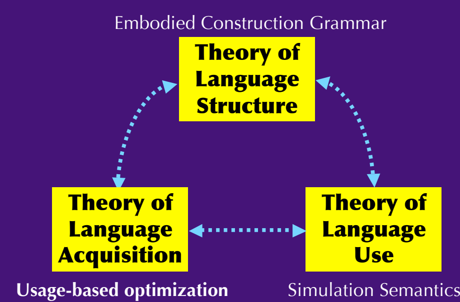
- Individual verb island constructions learned
  - Basic processes produce constructions similar to those in child production data.
  - System can generalize beyond encountered data given enough pressure to merge specific constructions.
  - Differences in verb learning lend support to verb island hypothesis.
- **Future directions**
  - full English corpus: non-motion scenes, argument structure cxns
  - Crosslinguistic data: Russian (case marking), Mandarin Chinese (directional particles, aspect markers)
  - Morphological constructions
  - Contextual constructions; multi-utterance discourse (Mok)

Summary

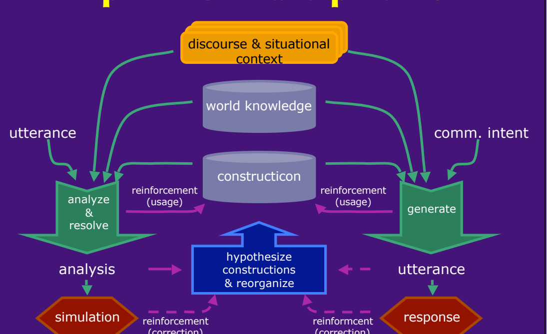
- **Model satisfies convergent constraints from diverse disciplines**
  - Crosslinguistic developmental evidence
  - Cognitive and constructional approaches to grammar
  - Computationally precise grammatical representations and data-driven learning framework for understanding and acquisition
- **Model addresses special challenges of language learning**
  - Exploits structural parallels in form/meaning to learn relational mappings
  - Learning is usage-based/error-driven (based on partial comprehension)
- **Minimal specifically linguistic biases assumed**
  - Learning exploits child's rich experiential advantage
  - Earliest, item-based constructions learnable from utterance-context pairs

Key model components

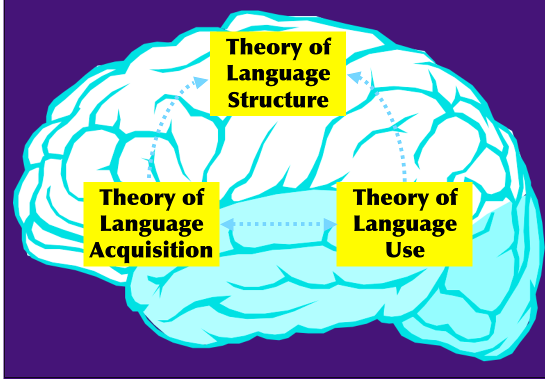
- **Embodied representations**
  - Experientially motivated rep'ns incorporating meaning/context
- **Construction formalism**
  - Multiword constructions = **relational form-meaning correspondences**
- **Usage 1: Learning tightly integrated with comprehension**
  - New constructions bridge gap between **linguistically analyzed** meaning and **contextually available** meaning
- **Usage 2: Statistical learning framework**
  - Incremental, specific-to-general learning
  - **Minimum description length** heuristic for choosing best grammar



Usage-based learning: comprehension and production



Recapitulation



Turing's take on the problem

"Of all the above fields the **learning of languages** would be the most impressive, since it is the most human of these activities.

This field seems however to depend rather too much on **sense organs and locomotion** to be feasible."

Alan M. Turing
Intelligent Machinery (1948)

Five decades later...

- Sense organs and locomotion
  - Perceptual systems (especially vision)
  - Motor and premotor cortex
  - Mirror neurons: possible representational substrate
  - Methodologies: fMRI, EEG, MEG
- Language
  - Chomskyan revolution
  - ...and counter-revolution(s)
  - Progress on cognitively and developmentally plausible theories of language
  - Suggestive evidence of embodied basis of language

...it may be more feasible than Turing thought!

(Maybe language depends *enough* on sense organs and locomotion to be feasible!)

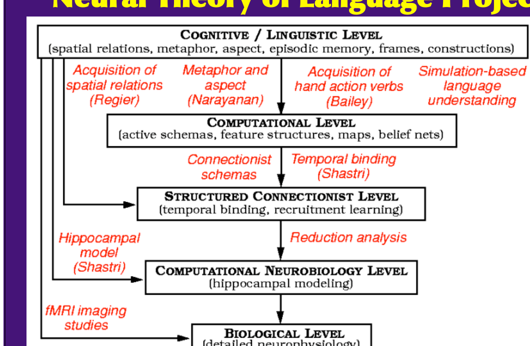
Motivating assumptions

- Structure and process are linked
  - Embodied language use constrains structure!
- Language and rest of cognition are linked
  - All evidence is fair game
- Need computational formalisms that capture embodiment
  - Embodied meaning representations
  - Embodied grammatical theory

Embodiment and Simulation: Basic NTL Hypotheses

- Embodiment Hypothesis
  - Basic concepts and words derive their meaning from embodied experience.
  - Abstract and theoretical concepts derive their meaning from metaphorical maps to more basic embodied concepts.
  - Structured connectionist models provide a suitable formalism for capturing these processes.
- Simulation Hypothesis
  - Language exploits many of the same structures used for action, perception, imagination, memory and other neurally grounded processes.
  - Linguistic structures set parameters for simulations that draw on these embodied structures.

The ICSI/Berkeley Neural Theory of Language Project

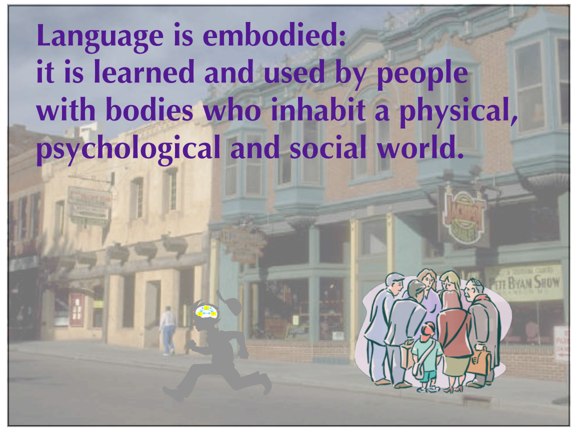


Jerome Feldman

From Molecule to Metaphor:

The Neural Basis of Language and Thought

MIT Press, 2006



How does the brain compute the mind?

How can a mass of chemical cells give rise to language and (the rest of) cognition?

Will computers think and speak?

How much can we know about our own experience?

How do we learn new concepts?

Does our language determine how we think?

Is language Innate?

How do children learn grammar?

How did languages evolve?

Why do we experience everything the way that we do?