Extracting Conceptual Knowledge from Text Corpora

Marco Baroni

Center for Mind/Brain Sciences
(University of Trento)

Napoli
April 16, 2008
Collaborators

- Brian Murphy, Eduard Barbu, Massimo Poesio (CIMeC)
- Alessandro Lenci (University of Pisa and ILC/CNR)
- Marco Baroni, Eduard Barbu, Brian Murphy and Massimo Poesio, in preparation. StruDEL: A distributional semantic model based on properties and types
Outline

Introduction

Distributional semantics
   The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
   Property generation
   Categorization
   Other tasks

The Human Experience
Text corpora

- Corpora are large electronic collections of texts produced in natural communicative settings
- Popular in language engineering, lexicography, language teaching/learning since (at least) the nineties
- Typology and some famous exemplars:
  - Balanced, representative, ‘reference’ corpora: Brown/LOB (1M tokens), COBUILD (10M, ...), BNC (100M)
  - Opportunistic: WSJ, la Repubblica-SSLMIT, Gigaword (1B)
  - Web-derived corpora (WaCky project)
  - Parallel: Hansard, OPUS, EuroParl
  - Specialized (CHILDES), comparable, diachronic...
Computational modelers in cognitive science (e.g., Rogers and McClelland 2004) typically work with hand-crafted input.

Corpora are “real”, natural input, akin to what humans hear/read, with same problems of noise and skewed input distribution (Zipf’s law) humans must face.

Computer seen as statistics-driven agent that “learns” from its environment: can it teach us something about human learning?
Zipf’s law
The contextual view meaning

- Acquisition/representation of meaning/conceptual knowledge is core issue in cognitive science
- Corpus-based simulations can help!
The contextual view meaning

- Acquisition/representation of meaning/conceptual knowledge is core issue in cognitive science
- Corpus-based simulations can help!
- “You should tell a word by the company it keeps” (Firth, 1957)
- “[T]he semantic properties of a lexical item are fully reflected in appropriate aspects of the relations it contracts with actual and potential contexts [...] [T]here are are good reasons for a principled limitation to linguistic contexts” (Cruse, 1986)
Outline

Introduction

**Distributional semantics**
   The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
   Property generation
   Categorization
   Other tasks

The Human Experience
Distributional semantics
Word space models (WSMs)

- Meaning of word/concept defined by *set of contexts* in which word occurs in corpus
- Similarity of words represented as *geometric distance* among *context vectors*
The dog barked in the park. The owner of the dog put him on the leash since he barked.

- bark
- park
- owner
- leash
The dog barked in the park. The owner of the dog put him on the leash since he barked.
Distributional semantics
Co-occurrence extraction for target word dog

The **dog** barked in the **park**. The owner of the dog put him on the leash since he barked.
The dog barked in the park.
The owner of the dog put him on the leash since he barked.

<table>
<thead>
<tr>
<th></th>
<th>bark</th>
<th>park</th>
<th>owner</th>
<th>leash</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
The dog barked in the park.
The owner of the dog put him on the leash since he barked.
Distributional semantics
Co-occurrence extraction for target word dog

The dog barked in the park. The owner of the dog put him on the leash since he barked.

<table>
<thead>
<tr>
<th></th>
<th>bark</th>
<th>park</th>
<th>owner</th>
<th>leash</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
## Distributional semantics

Meaning as co-occurrence

<table>
<thead>
<tr>
<th></th>
<th>leash</th>
<th>walk</th>
<th>run</th>
<th>owner</th>
<th>pet</th>
<th>bark</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>cat</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>lion</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>light</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>bark</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>car</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Distributional semantics
Similarity in word space
Distributional semantics
Similarity in word space

pet

owner

car (3,0)  
cat (2,3)  
dog (5,3)
Distributional semantics
Which context?

- Documents/large textual spans
- All words in a narrow window
- Lemmatized content words in a narrow window
- Content words in specific syntactic constructions or specific surface patterns
- Context needs not be linguistic! Vectors could include, e.g., co-occurrence counts with sensory stimuli
Distributional semantics
Success in cognitive simulations

- synonym identification (Landauer and Dumais 1997)
- text coherence (Landauer and Dumais 1997)
- categorization (Burgess and Lund 1997)
- semantic priming (Lowe 2000, McDonald and Brew 2002, Vigliocco et al. 2004)
- substitution errors (Vigliocco et al. 2004)
- child lexicon acquisition (Li et al. 2004, Baroni et al. 2007)
Introduction

**Distributional semantics**
   The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
   Property generation
   Categorization
   Other tasks

The Human Experience
The TOEFL synonym match task

- 80 items
The TOEFL synonym match task

- 80 items
- Target: *levied*
  Candidates: *imposed, believed, requested, correlated*
The TOEFL synonym match task

- 80 items
- Target: *levied*
  Candidates: *imposed, believed, requested, correlated*
Human performance on the synonym match task

- Average foreign test taker: 64.5%
- Macquarie University staff (Rapp 2004):
  - Average of 5 non-natives: 86.75%
  - Average of 5 natives: 97.75%
TOEFL results

- Humans:
  - Foreign test takers: 64.5%
  - Macquarie non-natives: 86.75%
  - Macquarie natives: 97.75%
TOEFL results

- **Humans:**
  - Foreign test takers: 64.5%
  - Macquarie non-natives: 86.75%
  - Macquarie natives: 97.75%

- **Machines:**
  - Rapp’s 2003 SVD-based model trained on lemmatized BNC: 92.5%
Outline

Introduction

Distributional semantics
  The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
  Property generation
  Categorization
  Other tasks

The Human Experience
“Semantic similarity” is a multi-faceted notion, but a single WSM provides only one way to rank a set of words (according to distance measure of choice).

Nearest neighbours of *motorcycle* in a standard WSM:

- motor → component
- car → co-hyponym
- diesel → component?
- to race → proper function
- van → co-hyponym
- bmw → hyponym
- to park → proper function
- vehicle → hypernym
- engine → component
- to steal → frame?
“Although jugs might be related to both vinegar and bottles, these relations are extremely different, and an overall similarity score does not represent these differences.”

In order to distinguish how jugs are related to vinegar from how they are related to bottles, one needs to know what are the properties of these concepts:

“[S]ince one’s concept of a jug, say, would include detailed information about its origins, parts, materials, functions and so on, the concept is more than sufficient to distinguish the meaning of jugs from that of vinegar and, for that matter, bottles.”
Semantic relations in cognitive and applied tasks

- Property generation: humans can easily produce coherent lists of typical properties of concepts (*norms* of McRae et al., Vinson/Vigliocco and others)
- Humans are able to distinguish different *types* of relations between properties and concepts, e.g., between formal and functional properties
  - A *dish* looks like a *CD* but its function is more similar to that of a *bottle*
Semantic relations in cognitive and applied tasks

- Different relations must be extracted and identified for modeling semantic interpretation, e.g.:
  - Telic quale relation in type coercion: finish the book (to read) vs. the ice-cream (to eat) (Pustejovsky 1995)
  - Salient properties in compound interpretation: a zebra cup is a cup with stripes
  - Parts in co-reference bridging: The building faced a dark alley. A window opened
- Specific types of properties (e.g., visual vs. functional) play crucial role in neural organization of concepts and semantic deficits (Martin 2007, Vigliocco et al. 2004)
- Semantic relations needed in practical applications, in particular development of lexical resources
Outline

Introduction

Distributional semantics
  The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
  Property generation
  Categorization
  Other tasks

The Human Experience
StruDEL
Structured Dimensions Extraction and Labeling

- Tries to build contextual vectors that represent *typed properties of concepts*
StruDEL
Structured Dimensions Extraction and Labeling

- Tries to build contextual vectors that represent *typed properties of concepts*
- Concept: *motorcycle*
- (Target) representation:
  - *for* traveling
  - *for* running
  - *is-a* vehicle
  - *has* motor
  - *has* two wheels
  - ...
Strategies for typed property extraction

- Automated identification of plausible concept-property connectors:
  - lice in a large number of dogs → YES
  - lice and leeches → NO
Strategies for typed property extraction

▶ Automated identification of plausible concept-property connectors:
  ▶ lice in a large number of dogs → YES
  ▶ lice and leeches → NO

▶ (Weighted) number of distinct connectors between concept and property is better indicator of true semantic relation than absolute co-occurrence frequency
  ▶ year of the tiger is very frequent, but following are not attested: year of some tigers, the tigers have years, etc.
  ▶ Vice versa, no tail/tiger connector is very frequent, but there are many of them: tail of the tiger, tail of some tigers, the tigers have tails, etc.
Strategies for typed property extraction

- Automated identification of plausible concept-property connectors:
  - lice in a large number of dogs → YES
  - lice and leeches → NO

- (Weighted) number of distinct connectors between concept and property is better indicator of true semantic relation than absolute co-occurrence frequency
  - year of the tiger is very frequent, but following are not attested: year of some tigers, the tigers have years, etc.
  - Vice versa, no tail/tiger connector is very frequent, but there are many of them: tail of the tiger, tail of some tigers, the tigers have tails, etc.

- The type of a relation can be extracted by generalizing over the connectors: of, with, of some, have point, together, to a part/whole relation
StruDEL baking

- Concept-property-type tuples extracted from ukWaC, a corpus of random Web pages including 2.25 billion tokens
- Property lists extracted for 1,234 (concrete) concepts
- Compared to various state-of-the-art WSMs, including SVD-based model and model using dependency parses (DV)
The StruDEL description

<table>
<thead>
<tr>
<th>property</th>
<th>type</th>
<th>$LL$</th>
</tr>
</thead>
<tbody>
<tr>
<td>to read</td>
<td>verb obj</td>
<td>3941.3</td>
</tr>
<tr>
<td>author</td>
<td>c from p</td>
<td>3772.8</td>
</tr>
<tr>
<td>to write</td>
<td>verb obj</td>
<td>2399.5</td>
</tr>
<tr>
<td>reader</td>
<td>c for p</td>
<td>2298.5</td>
</tr>
<tr>
<td>chapter</td>
<td>p in c</td>
<td>2259.8</td>
</tr>
<tr>
<td>library</td>
<td>c in p</td>
<td>2222.4</td>
</tr>
<tr>
<td>to publish</td>
<td>verb obj</td>
<td>1907.7</td>
</tr>
<tr>
<td>reading</td>
<td>p from c</td>
<td>1296.8</td>
</tr>
<tr>
<td>publisher</td>
<td>c from p</td>
<td>1258.0</td>
</tr>
<tr>
<td>review</td>
<td>p about c</td>
<td>1156.4</td>
</tr>
</tbody>
</table>
### Tiger

**The StruDEL description**

<table>
<thead>
<tr>
<th>property</th>
<th>type</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>jungle</td>
<td>c from p</td>
<td>132.2</td>
</tr>
<tr>
<td>cat</td>
<td>p as c</td>
<td>94.1</td>
</tr>
<tr>
<td>species</td>
<td>p as c</td>
<td>89.4</td>
</tr>
<tr>
<td>stripes</td>
<td>p as c</td>
<td>84.1</td>
</tr>
<tr>
<td>animal</td>
<td>p as c</td>
<td>75.6</td>
</tr>
<tr>
<td>to maul</td>
<td>subj verb</td>
<td>63.7</td>
</tr>
<tr>
<td>habitat</td>
<td>c in p</td>
<td>63.4</td>
</tr>
<tr>
<td>lion</td>
<td>p as c</td>
<td>56.0</td>
</tr>
<tr>
<td>to tame</td>
<td>verb obj</td>
<td>53.2</td>
</tr>
<tr>
<td>zoo</td>
<td>c in p</td>
<td>51.4</td>
</tr>
<tr>
<td>property</td>
<td>type</td>
<td>LL</td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>to ride</td>
<td>verb obj</td>
<td>345.8</td>
</tr>
<tr>
<td>rider</td>
<td>p on c</td>
<td>199.8</td>
</tr>
<tr>
<td>vehicle</td>
<td>p as c</td>
<td>103.7</td>
</tr>
<tr>
<td>motorbike</td>
<td>p for c</td>
<td>100.2</td>
</tr>
<tr>
<td>street</td>
<td>c on p</td>
<td>71.3</td>
</tr>
<tr>
<td>to park</td>
<td>verb obj</td>
<td>69.3</td>
</tr>
<tr>
<td>scooter</td>
<td>p over c</td>
<td>51.6</td>
</tr>
<tr>
<td>car</td>
<td>p as c</td>
<td>45.7</td>
</tr>
<tr>
<td>to insure</td>
<td>p for c</td>
<td>39.8</td>
</tr>
<tr>
<td>bike</td>
<td>p out c</td>
<td>37.7</td>
</tr>
</tbody>
</table>
Outline

Introduction

Distributional semantics
  The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
  Property generation
  Categorization
  Other tasks

The Human Experience
Outline

Introduction

Distributional semantics
  The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
  Property generation
  Categorization
  Other tasks

The Human Experience
725 participants rating 541 concepts, 30 subjects per concept

Subjects produce list of properties that describe concept

Manual normalization: *loud, noise, noisy* mapped to *is loud*

NB: NORMS is a (conscious) *product* of human semantic representations, not a direct window into these representations

In other words, NORMS is useful comparison point, but not necessarily a “gold standard”
motorcycle

The NORMS description

<table>
<thead>
<tr>
<th>property</th>
<th>productions</th>
</tr>
</thead>
<tbody>
<tr>
<td>has wheels</td>
<td>22</td>
</tr>
<tr>
<td>has 2 wheels</td>
<td>20</td>
</tr>
<tr>
<td>is dangerous</td>
<td>18</td>
</tr>
<tr>
<td>has engine</td>
<td>13</td>
</tr>
<tr>
<td>is fast</td>
<td>12</td>
</tr>
<tr>
<td>used with helmet</td>
<td>11</td>
</tr>
<tr>
<td>made by Harley Davidson</td>
<td>10</td>
</tr>
<tr>
<td>is loud</td>
<td>10</td>
</tr>
<tr>
<td>used by 1 or 2 people</td>
<td>9</td>
</tr>
<tr>
<td>is a vehicle</td>
<td>9</td>
</tr>
</tbody>
</table>
Property analysis

- On average, NORMS and StruDEL share 2.4 of the 10 most salient properties of a concept.
- No other distributional model we tested shares more than 1.5/10 properties with NORMS.
Property analysis

- On average, NORMS and StruDEL share 2.4 of the 10 most salient properties of a concept.
- No other distributional model we tested shares more than 1.5/10 properties with NORMS.
- Systematic analysis of different property types privileged by NORMS vs. StruDEL (Baroni and Lenci 2008).
- E.g., for car:
  - Shared: *engine, gasoline, transportation*
  - NORMS only: *wheels, 4 wheels, doors, steering wheel, expensive, for passengers, vehicle*
  - StruDEL only: *it is driven, driver, it is parked, road, garage, race, parking*
Property types
Wu and Barsalou (Submitted), McRae et al. (2005), simplified

C Category: dog-animal, airplane-vehicle
P Parts: dog-tail, airplane-wing
Q Qualities: dog-brown, airplane-fast
A Typical Activities and behaviours: dog-barks, airplane-flies
F Function: dog-pet, dog-hunting, airplane-transportation
E Related Entities: dog-cat, airplane-pilot
L Location: dog-kennel, airplane-sky
Property types in NORMS and StruDEL
Properties by categories

NORMS

<table>
<thead>
<tr>
<th>C</th>
<th>P</th>
<th>Q</th>
<th>A</th>
<th>F</th>
<th>E</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grAnim</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fruit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>veggie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tool</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vehicle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

StruDEL

<table>
<thead>
<tr>
<th>C</th>
<th>P</th>
<th>Q</th>
<th>A</th>
<th>F</th>
<th>E</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grAnim</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fruit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>veggie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tool</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vehicle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Outline

Introduction

Distributional semantics
   The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
   Property generation
   Categorization
   Other tasks

The Human Experience
Data-set
44 concrete concepts

- 24 natural concepts
  - 15 animals: 7 birds, 8 ground animals
  - 9 vegetables: 4 fruits, 5 greens
- 20 artifacts
  - 13 tools
  - 7 vehicles
Hierarchical categorization

- 6-way: birds, ground animals, fruits, greens, tools, vehicles
- 3-way: animals, vegetables, man-made
- 2-way: natural, man-made
Categorization as clustering in semantic space
Categorization as clustering in semantic space
## Results

Percentage purity of clusters

<table>
<thead>
<tr>
<th>space</th>
<th>6 categories</th>
<th>3 categories</th>
<th>2 categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORMS</td>
<td>91</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>StruDEL</td>
<td>79</td>
<td>91</td>
<td>98</td>
</tr>
<tr>
<td>DV</td>
<td>73</td>
<td>89</td>
<td>95</td>
</tr>
<tr>
<td>SVD</td>
<td>79</td>
<td>75</td>
<td>59</td>
</tr>
</tbody>
</table>
### “Animals”
Typical properties in the 3-way solution

<table>
<thead>
<tr>
<th>NORMS</th>
<th>StruDEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>it breeds</td>
</tr>
<tr>
<td>legs</td>
<td>is seen</td>
</tr>
<tr>
<td>beak</td>
<td>is shot</td>
</tr>
<tr>
<td>eggs</td>
<td>is rescued</td>
</tr>
<tr>
<td>bird</td>
<td>dies</td>
</tr>
</tbody>
</table>
“Vegetables”
Typical properties in the 3-way solution

<table>
<thead>
<tr>
<th>NORMS</th>
<th>StruDEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ vegetable</td>
<td>▶ is sliced</td>
</tr>
<tr>
<td>▶ sweet</td>
<td>▶ is minced</td>
</tr>
<tr>
<td>▶ on trees</td>
<td>▶ is eaten</td>
</tr>
<tr>
<td>▶ fruit</td>
<td>▶ it grows</td>
</tr>
<tr>
<td>▶ edible</td>
<td>▶ slice</td>
</tr>
</tbody>
</table>
“Man-made”
Typical properties in the 3-way solution

<table>
<thead>
<tr>
<th>NORMS</th>
<th>StruDEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>metal</td>
<td>is used</td>
</tr>
<tr>
<td>plastic</td>
<td>in hands</td>
</tr>
<tr>
<td>handle</td>
<td>powered</td>
</tr>
<tr>
<td>for transportation</td>
<td>has use</td>
</tr>
<tr>
<td>wood</td>
<td>makes things</td>
</tr>
</tbody>
</table>
Outline

Introduction

Distributional semantics
   The TOEFL synonym match task

From distributional semantics to conceptual knowledge

StruDEL

Testing StruDEL
   Property generation
   Categorization
   Other tasks

The Human Experience
Other tasks
StruDEL is consistently best performer

- Predicting free association (*cat* . . . *dog*)
- Modeling prototypicality ratings (*a sparrow* is a more typical bird than *a penguin*)
- Generating specific properties (*what is the typical* location *of hammers? what is their typical* function?)
From word spaces to brain spaces
Work in progress

Subject 5, image, animal vs manipulate object

pet

owner

car (3,0)

dog (5,3)

cat (2,3)
my brother
and I got dried bananas when war broke out
three pounds of bananas please
we could have a banana souffle
Those bananas are going a bit mm
well I like them when they go soft

who eats bananas?
who do we have in our zoo that eats bananas?
monkeys eat bananas
shall we give the monkey a banana?
there’s your banana
someone went berserk with a <hammer>, that’s been known
we had <hammer> drilled the blunt bit
I was tapping
it with a <hammer> wasn’t I

she had a <hammer> and she was banging
on the wall
the <hammer> is the most useful tool, Lara. Whenever the telly
goes wrong you
just hit it
with your <hammer>
The Human Experience

- We can gain useful insights about human conceptual acquisition applying fancy learning techniques to whatever corpora we have available.
- But we suspect we will not make a major breakthrough until we learn from the same data humans learn from:
  - Corpora cue how adults transfer knowledge to other adults.
  - Knowledge in corpora is not organized incrementally.
  - Large corpora (still) lack multimodal information.
- Corpora in CHILDES are too small, sparse, often record speech in special occasions.
  - Lara, one of densest CHILDES corpora, contains transcripts of about 120 hours.
We can gain useful insights about human conceptual acquisition applying fancy learning techniques to whatever corpora we have available. But we suspect we will not make a major breakthrough until we learn from the same data humans learn from: corpora cue how adults transfer knowledge to other adults, knowledge in corpora is not organized incrementally, and large corpora (still) lack multimodal information. Corpora in CHILDES are too small, sparse, often record speech in special occasions. Lara, one of densest CHILDES corpora, contains transcripts of about 120 hours. The Human Experience: record full verbal and visual experience of multiple children uninterruptedly for first 3 years of life.
Some references


L. Wu & L. Barsalou (Submitted). *Grounding concepts in perceptual simulation: I. Evidence from property generation.*