The Great (Penn Treebank) Robbery: When Statistics is Not Enough or The Incredible Lightness of Being a Corpus

Robert C. Berwick, Michael Coen, Sandiway Fong & Partha Niyogi
A cognitive checklist

- Does it attain ‘knowledge of human language’?
  - Grammatical/ungrammatical
  - More important: the right structures
- Does it not attain “non-knowledge” of human language (eg, Fortran, permutation language; cf Epun, Smith)
- Cognitively plausible in terms of # of input examples, kind of input data, robustness to example variation?
- Are we Fox News? (Fair and balanced) Well, consider the advertising:
Knowledge of language

• Checklist
Pre Syntactic Structures

• Aux system
• Just memorizes sequences - Big Blue analogy with a very large opening book
• Berkeley parser as Pachinko machine
PTB

- The Penn Treebank (PTB) project selected 2,499 stories from a three year Wall Street Journal (WSJ) collection of 98,732 stories for syntactic annotation.
- Picture: of PTB, stuffed inside kid’s head.
- Problem: license fee to LDC? Nonmembers $2500
PTB: rules of the game

Stat model ([Bikel, 2004])

Estimate

Parse

Evaluate

49,208 total sentences

40,007 training sentences

2,416 test sentences

Cogsci 2007 Workshop: Psychocomputational Models
Cogsci 2007 Workshop:
Psychocomputational Models

Input Data = $2500/baby

???
PTB: the Discrete Charm of the
Basic results and outline

- Don’t cry over spilled milk: Excessively fragile
- Mirror input data
The Penn Treebank (PTB)
What we don’t have time to talk about here

- Conceptual issues
- Engineering/Methodological issues
  - Overtraining: no x-validation
  - Evaluation metric
  - Other G’s: CCGs.
Remember our picture of modularity

- Syntax + lexicon
- We want to avoid duplicating information
- Lexicon already has ‘semantic’ type information in it
- How does this information enter a parse?
- Do the statistical parsers enter all that’s needed?
- What information does/must the lexicon contain?
But how well does this work?
Does it do better? Here is the std recall & precision used

Gold standard: the guy saw the person with the cookies

NP 1 2
NP 4 5
NP 4 8
PP 6 8
NP 7 8
VP 3 8
S 1 8
The guy saw the person with the cookies

Gold std
NP 1 2
NP 4 5
NP 4 8
PP 6 8
NP 7 8
VP 3 8
S 1 8

Actual
NP 1 2
NP 4 5
NP 4 8
PP 6 8
NP 7 8
VP 3 8
S 1 8

G = # phrases in gold std = 7
P = # in parse output = 6
C = # correct = 6

Recall = C/G = 6/7 = 86%
Precision = C/P = 6/6 = 100%

Cogsci 2007 Workshop:
Psychocomputational Models
Some actual results on PTB, train on 40,000 sentences, test on 2,000

<table>
<thead>
<tr>
<th>Approach</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCFG</td>
<td>70.6%</td>
<td>74.8%</td>
</tr>
<tr>
<td>Lexical head dep.</td>
<td>85.3%</td>
<td>85.7%</td>
</tr>
<tr>
<td>Generative lexical (what we sketched)</td>
<td>86.7%</td>
<td>86.6%</td>
</tr>
</tbody>
</table>

“Latest” models nearly 88-89% on both P & R…
What are the remaining issues?
Does this really work?
You will find out in the next R&R, and Lab 3a…!
Adding Heads

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCFGs (Charniak 97)</td>
<td>70.6%</td>
<td>74.8%</td>
</tr>
<tr>
<td>Conditional Models – Decision Trees (Magerman 95)</td>
<td>84.0%</td>
<td>84.3%</td>
</tr>
<tr>
<td>Lexical Dependencies (Collins 96)</td>
<td>85.3%</td>
<td>85.7%</td>
</tr>
<tr>
<td>Conditional Models – Logistic (Ratnaparkhi 97)</td>
<td>86.3%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Generative Lexicalized Model (Charniak 97)</td>
<td>86.7%</td>
<td>86.6%</td>
</tr>
<tr>
<td>Model 1 (no subcategorization)</td>
<td>87.5%</td>
<td>87.7%</td>
</tr>
<tr>
<td>Model 2 (subcategorization)</td>
<td>88.1%</td>
<td>88.3%</td>
</tr>
</tbody>
</table>
Some problems

- Conjunctions - only 50% precision & recall
- Why?
Does this fix this problem?

'Stelescope' should force this pref, above 'early' attachment
No “Unified Theory of Semantics”*

Different goals ≠ different semantic theories:
Syntactician: why do different words appear in different constructions?
Semanticist: what is an adequate meaning representation of a vocabulary item?
Lexicographer: what are all the things we know about a word’s meaning?
IR Engineers: what is the meaning abstraction of a piece of text?
Roboticist: how can the robot appear to understand me?
Child Dev Psych (Vocab + Grammar)
Historical linguist

(Is there a unified theory for chemistry? physics?)
Uncertainty in terms*

- Grammar: How much semantics should be in it?
- Grammaticality: Is a semantically anomalous sentence ungrammatical?
  - He gave the book to John.
  - He thought the book to John.
- Grammatical category: What are their essences?
- Word Meaning: What is a meaning representation?
- Concepts: How are they related to words?
  - How is what we know about TIGER related to /tiger/?

*No one knows the answer! Wait: When did science know the definition of an atom, electron, proton, …?
Is Meaning About Truth?

(1) John met a unicorn.
   Is (1) false because unicorns don’t exist?

(2) John met a unicorn in my dream.
   How does “in my dream” change things?
Some examples to worry about…

John thought the book to Mary
John's growth of tomatoes
Sue walked in an hour
Bob shelved the windowsill with the book
Bob buttered the margarine onto the bread

Where to put the information about this???
What is the information about this???

Cogsci 2007 Workshop:
Psychocomputational Models
Words appear in a very wide variety of constructions

He sirened her down.
The car sirened its way to NY.
She sirened Bill the message.

Fantasy:
VP [\text{\textsf{\textregistered}}] V142 PP
V142 [\text{\textsf{\textregistered}}] siren

A more flexible approach needed!
Does lexicon hold Subcategorization information?  
(Chomsky 1965)

- Verbs have classes:
  - John ate a tomato.
  - A tomato was eaten.
  - John resembled a tomato.
  - ? A tomato was resembled.

  You have seen this in your labs!

- Use features:
  - +animate, -passive, +male, +human, ...
  - If we allow +human, then do we allow +good-to-eat-with-chicken?
  - Wait: where are the restrictions on these features?
  - Major problem: Blank check on features
Does lexicon hold ‘thematic roles’?

• “Who does what to whom”
• Link syntactic positions to thematic roles
• Thematic roles: Agent, Affected Object, Beneficiary, Goal, Theme (where does this list come from?)
• Example: eat
  Agent(x) & Theme(y) & Eat(e, x, y)

Problem: ‘rules’ linking thematic roles to structure recovered by parser can vary enormously (which is why we said before we couldn’t eliminate it from the lexicon)

Examples: John climbed the mountain; Who married Martha; John ate an apple, etc.
If so...

- Then lexicon holds the subcategorization information, thematic roles and ‘linking rules’

- What else?
Does Lexicon hold semantically grounded classes? (lexical-semantic structure)

- +motion +contact –effect
- Hit, touch, break, cut classes

Any notion that rules apply blindly without paying attention to “semantics” is pure wishful thinking. The question is how much attention.
Levin classes (3100 verbs)

• 47 top level classes, 150 second and third level

• Based on pairs of syntactic frames.
  
  John broke the jar. / Jars break easily. / The jar broke.
  John cut the bread. / Bread cuts easily. / *The bread cut.
  John hit the wall. / *Walls hit easily. / *The wall hit.

• Reflect underlying semantic components
  contact, directed motion,
  exertion of force, change of state

• Synonyms, syntactic patterns, relations
Other alternation examples

- Causative/inchoative
  The window broke
  John broke the window
  The rabbit suddenly appeared
  *The magician appeared the rabbit

- Benefactive:
  Sue carved a toy out of wood for Hansel
  Sue carved Hansel a toy out of wood
  Sue carved some wood into a toy for Hansel
  *Sue carved Hansel some wood into a toy

- Middle formation:
  The whale frightens easily
  *The whale sees easily
Lexical-semantic structure

- Instead of:
  \[ \text{Agent}(x) \& \text{Theme}(y) \& \text{Eat}(e, x, y) \]
- We have:
  \[ \text{CAUSE}([\text{Thing } i], \text{GO}([\text{Thing } j], \text{IN-MOUTH-OF}([\text{Thing } i]))) \]
Lexical semantic Structure

- Each node contains:
  - **Primitive:**
    - CLOSED CLASS: GO, STAY, BE, ON, IN, AT…
    - OPEN CLASS: JOHN, RUN-INGLY, …
  - **Field:** Analogy to motion/position in Localist approach: LOCATIONAL, POSSESSITIONAL, TEMPORAL,…
  - **Type:** EVENT, STATE, PATH, POSITION, MANNER…
Event structure – a fuller picture

• /Bob put -ed the book on the shelf/
  (cause :agent (bob) :effect (go :theme (book)
    :path (path :oper (on) :terminal+ (shelf))):tense
     past))

• /What did Bob put on the shelf/
  (cause :agent (bob) :effect (go :theme (? (what))
    :path (path :oper (on) :terminal+ (shelf))):tense
     past))

• /What did Bob put the book on/
  (query :event (cause :agent (bob) :effect (go :theme (book)
    :path (path :oper (on) :terminal+ (? (what))))):tense
     past))

• /Where did Bob put the book/
  (query :event (cause :agent (bob) :effect (go :theme (book)
    :path (path :oper ():terminal+ (? (where))))):tense
     past))
Structural vs. Content Meaning

- Verbs in a class share structural component
- Verbs in a class are distinguished by content component
Structural vs. Content Meaning

- go-loc
  - thing2
  - from-loc
  - to-loc
    - thing2
    - at-loc
    - thing2
    - at-loc
      - thing2
      - thing4
      - thing2
      - thing6

- amble-ingly
- lope-ingly
- skitter-ingly
- zoom-ingly
Common objections

Definition = Structure + Plus X, for unknown X

• Consider paint, water, butter, …:
  – She painted a house, he watered a plant, he buttered bread
• Claim: Structure is “put N on X” (Hale & Keyser 2003)
• Plus X: (story about putting)

**Undefinable primitives:**

(1) Thematic Roles: Agent, Patient, Goal, …
Remedy: Define/derive them structurally (Hale & Keyser 2003)
(2) Lexical Semantic Primitives: CAUSE, GO, BE, HAVE, ..
Remedy: Decompose them even more (Jackendoff 1991, 1996)
Review: What does the lexicon look like?

• Examples:
  *Bob put. *Bob put butter.
  Bob put butter on the bread.
  Butter was put on the bread
  What was put on the bread?
  Where was the butter put?

• Traditional view - encode in rules with ‘vanilla’ nonterminals
What does the lexicon look like?
Ans 1 (traditional): use lots of rules, essentially exhaustive listing

- VP \[\wedge\] V9 NP PP\textsubscript{LOCATIVE} V9 \[\wedge\] put
- VP \[\wedge\] was VPass ; VPass \[\wedge\] V9 PP\textsubscript{LOCATIVE}
- VP/NP \[\wedge\] V9 NP/NP PP\textsubscript{LOCATIVE}
- VP/NP \[\wedge\] V9 NP PP\textsubscript{LOCATIVE}/NP
- PP\textsubscript{LOCATIVE} \[\wedge\] P\textsubscript{LOCATIVE} NP ; P\textsubscript{LOCATIVE} \[\wedge\] on | in |.. 
- PP\textsubscript{LOCATIVE}/NP \[\wedge\] P\textsubscript{LOCATIVE} NP/NP

Cogsci 2007 Workshop:
Psychocomputational Models
Lexical-semantics

- \[ \text{Put V NP}_j \text{ PP}_k \text{ CAUSE([[BOB]_i \text{, GO([[BUTTER]_j, TO([[BREAD]_k]]))})] \]

- Semantic templates mirrors alternation patterns, but are ad-hoc
- Syntax a bit simpler w/ the semantic types factored out
Hypothesis 2: **Lexicon** Contains Selection Criteria

/\texttt{shelf}/ has $p_{LOCATION}$ selection in lexicon ($=p_{LOCATION} =d(et) \ V$)

Also: /\texttt{shelf}/ is $n_{LOCATION}$

/\texttt{butter}/ has $p_{LOCATUM}$ selection in lexicon ($=p_{LOCATUM} =d(et) \ V$)

Also: /\texttt{butter}/ is $n_{LOCATUM}$

So then the **Lexicon** cannot derive:

* 1. Bob shelved the windowsill with the book.
* 2. Bob buttered the margarine onto the bread.

Information about butter and shelf – **where** is it located?
Hypothesis 3: Encyclopedia vs. Lexicon

Does Encyclopedia holds knowledge 'rejecting' the following “GRAMMATICAL” sentences? Or does the lexicon?

# John thought the book to Mary
# John's growth of tomatoes
# Sue walked in an hour
# Bob shelved the windowsill with the book.
# Bob buttered the margarine onto the bread.
2 Language Acquisition Problems: Lexicon vs Encyclopedia

<table>
<thead>
<tr>
<th>ROOT</th>
<th>LEXICON ENTRIES</th>
<th>ROOT</th>
<th>ENCYCLOPEDIA ENTRIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>/shelf/</td>
<td>n, =p =d v_{cause}</td>
<td>/shelf/</td>
<td>n_{LOCATION}, =p_{LOCATION} =d v</td>
</tr>
<tr>
<td>/butter/</td>
<td>n, = =d v_{cause}</td>
<td>/butter/</td>
<td>n_{LOCATUM}, =p_{LOCATUM} =d v</td>
</tr>
<tr>
<td>/into/</td>
<td>=d +k p</td>
<td>/into/</td>
<td>=d +k p_{LOCATION}</td>
</tr>
<tr>
<td>/with/</td>
<td>=d +k p</td>
<td>/with/</td>
<td>=d +k p_{LOCATUM}</td>
</tr>
</tbody>
</table>

LEXICON ACQUISITION:
How do LEXICAL roots get assigned to feature set?

ENCYCLOPEDIA ACQUISITION:
How do ENCYCLOPEDIA roots get assigned to feature set?
The question

• What generalizations has the parser made from being trained on the Penn Tree Banks?
• Are these the right generalizations?
Testing the parser

• Look at verbs in an alternation class (which is ‘semantically’ and syntactically coherent)
• Find the –logprobs for the alternations, including the ‘ungrammatical’ ones
• Do these match up with intuitions and expectations from frequencies in the Penn Tree Bank?
• **What would you expect the Collins parser to say about a set of alternations?**

  - $[\text{VP} \ V \ \text{NP} \ \text{PP}]$
  - $[\text{VP} \ V \ \text{PP} \ \text{NP}]$

  **Data sources:**
  - $f_1 [\text{VP} \ V \ \text{NP} \ \text{PP}]$
  - $f_2 [\text{VP} \ V \ \text{PP} \ \text{NP}]$

  **Collins parser**

  **Linguistic intuitions**

  logprob values for $[\text{VP} \ V \ \text{NP} \ \text{PP}]$ and $[\text{VP} \ V \ \text{PP} \ \text{NP}]$

  $[\text{VP} \ V \ \text{NP} \ \text{PP}] < [\text{VP} \ V \ \text{PP} \ \text{NP}]$
join belongs to section 22.1
Mix verbs

Syntactic frames:
- 84 NP
- 11 NP PP
- 1 NP-and (together)
- 7 PP
- 8 []
- ADJ PP-with
- ADJ (together)
22 Verbs of Combining and Attaching

References: Condoravdi and Sanfilippo (1990), Gentner (1978)

These verbs are all related to combining or attaching. Their hallmark is participation in the simple reciprocal alternations, the *together* reciprocal alternations, or both. Members of this class are never found in the *apart* reciprocal alternations. The various subclasses differ according to whether the meanings of their members involve a result or means component.

22.1 Mix Verbs

Class Members:
*with*: blend, combine, commingle, concatenate, connect, fuse, join, link, merge, mingle, mix, pool
*into*: blend, cream, mix
*to*: add, connect, join, link, network

Properties:

(313) Simple Reciprocal Alternation (transitive):
   a. Herman mixed the eggs with the cream. (prepositional variant)
   b. Herman mixed the eggs and the cream. (reciprocal variant)

(314) Simple Reciprocal Alternation (intransitive; most verbs):  
   a. The eggs mixed with the cream.
   b. The eggs and the cream mixed.

(315) *Together* Reciprocal Alternation (transitive):  
   a. Herman mixed the eggs with the cream.
   b. Herman mixed the eggs and the cream together.

(316) *Together* Reciprocal Alternation (intransitive; most verbs):  
   a. The eggs mixed with the cream.
   b. The eggs and the cream mixed together.

(317) Causative/Inchoative Alternation (most verbs):  
   a. I mixed the soap into the water.  
   The soap mixed into the water.
from mit.six863.parse.treebank import *
from string import join

def demog():
give = lambda t: t.node == 'VP' and len(t)> 2
    and t[1].node == 'NP'
    and (t[2].node == 'PP-DTC' or t[2].node == 'NP')
    and ('gave' in t[0].leaves() or 'give' in t[0].leaves())
for tree in g.parsed():
    for t in g.subtrees(give):
        print "%s [ %s: %s ] [ %s: %s ]" %
            (join(t[0].leaves()),
            t[1].node, join(t[1].leaves()),
            t[2].node, join(t[2].leaves()))
‘give NP to PP’, sentence 824
‘Give’ NP NP vs. NP PP-DTV

- 256 total give NP NP or NP PP-DTV in PTB
- 205 are NP NP  80%
- 51 are NP PP-DTV  20%

- Which frame is therefore going to be preferred?
The verb “join”

- Look for VP nodes that:
  - immediately dominates VB*, and
  - that VB* is the 1st child
Caveats

• Some verbs have multiple senses
• Not all instances of a category label hold the same ‘semantic’ or ‘thematic’ role
• To be completely accurate, we’d have to review each and every tree and label each node with a semantic role, very carefully
• But as a first approximation, let’s conflate category labels
Patterns (39)

- 1 [NP, PP-CLR, NP-TMP]
- 1 [PP-CLR, PP-LOC]
- 1 [NP-TMP, "S-ADV"]
- 1 [NP-LOC, S-PRP]
- 1 [PP-TMP, S-PRP]
- 1 [NP, ADVP-MNR, NP-TMP]
- 1 [SBAR-TMP]
- 1 [NP-CLR, PP-CLR]
- 1 [ADVP-CLR, S-PRP]
- 1 [NP, PP-CLR, ADVP-TMP]
- 1 [NP, SBAR-TMP]
- 1 [SBAR-NOM]
- 1 [PRT]
- 1 [NP-TMP]
- 3 [PP-CLR]
- 2 [PRT, PP-CLR]
- 4 [NP, ADVP-TMP]
- 1 [NP, ADVP-TMP, ADVP-PRP]
- 1 [PP-CLR, S-CLR]

- 1 [NP, PP, NP-TMP]
- 1 [NP, PP-CLR, PP-LOC]
- 1 [NP, ADVP]
- 1 [NP, "S-ADV"]
- 11 [NP, PP-TMP]
- 3 []
- 1 [PP]
- 2 [PP-CLR, PP]
- 1 [NP, PP, PP-TMP, "PP-TMP"]
- 2 [NP, PP-TMP, PP]
- 1 [PP-TMP, PP-CLR]
- 1 [NP, PP-CLR, PP-TMP, ADVP-TMP]
- 1 [NP, PP-TMP, S-PRP]
- 2 [NP, PP]
- 8 [NP, PP-LOC]
- 1 [PP-CLR, S-PRP]
- 16 [NP, PP]
- 2 [NP, PP, PP-TMP]
- 4 [NP, PP-CLR]
- 58 [NP]
Delete ‘TMP’ nodes

- Why?
- Temporal PPs, etc.
Patterns (27)

- 1 [PP-CLR, PP-LOC]
- 1 [, , S-ADV]
- 1 [PP-LOC, S-PRP]
- 1 [S-PRP]
- 1 [NP, ADVP-MNR]
- 1 [NP-CLR, PP-CLR]
- 1 [ADVP-CLR, S-PRP]
- 1 [SBAR-NOM]
- 1 [PRT]
- 2 [PRT, PP-CLR]
- 1 [NP, ADVP-PRP]
- 1 [PP-CLR, S-CLR]
- 1 [NP, PP-CLR, PP-LOC]
- 1 [NP, ADVP]

- 1 [NP, , S-ADV]
- 5 []
- 1 [PP]
- 2 [PP-CLR, PP]
- 1 [NP, PP, ,]
- 4 [PP-CLR]
- 1 [NP, S-PRP]
- 2 [NP, , , PP]
- 8 [NP, PP-LOC]
- 1 [PP-CLR, S-PRP]
- 21 [NP, PP]
- 7 [NP, PP-CLR]
- 74 [NP]
Delete the extra ‘commas’
Mr. Craven joined Morgan Grenfell as group chief executive in May 1987, a few months after the resignations of former Chief Executive Christopher Reeves and other top officials because of the merchant bank’s role in Guinness PLC’s controversial takeover of Distiller’s Co. in 1986.

\[
[\text{VP} \ joined \ NP \ PP \ PP-TMP \ , \ PP-TMP]
\]
Patterns (25)

- 1 [PP-CLR,PP-LOC]
- 1 [S-ADV]
- 1 [PP-LOC,S-PRP]
- 1 [S-PRP]
- 1 [NP,ADVP-MNR]
- 1 [NP-CLR,PP-CLR]
- 1 [ADVP-CLR,S-PRP]
- 1 [SBAR-NOM]
- 1 [PRT]
- 2 [PRT,PP-CLR]
- 1 [NP,ADVP-PRP]
- 1 [PP-CLR,S-CLR]
- 1 [NP,PP-CLR,PP-LOC]

- 1 [NP,ADVP]
- 1 [NP,S-ADV]
- 5 []
- 1 [PP]
- 2 [PP-CLR,PP]
- 4 [PP-CLR]
- 1 [NP,S-PRP]
- 8 [NP,PP-LOC]
- 1 [PP-CLR,S-PRP]
- 24 [NP,PP]
- 7 [NP,PP-CLR]
- 74 [NP]
PP-LOC

- Cases 2401, 3983 and 43738
ADVP and -MNR

- Cases 11601 (ADVP-MNR) and 30526 (ADVP)
-ADV

- adverbial
- Cases 5710 (S-ADV) and 35906 (S-ADV)

Psychocomputational Models
-PRP

- purpose
- Case 46400 (S-PRP)
Patterns

- Delete ADV(P), -LOC, -MNR, _PRP
  - 1 [NP-CLR, PP-CLR]
  - 1 [SBAR-NOM]
  - 1 [PRT]
  - 2 [PRT, PP-CLR]
  - 1 [PP-CLR, S-CLR]
  - 9 []
  - 1 [PP]
  - 2 [PP-CLR, PP]
  - 6 [PP-CLR]
  - 24 [NP, PP]
  - 8 [NP, PP-CLR]
  - 87 [NP]

- Delete ADV(P) but not anything with -CLR, -LOC, -MNR, _PRP
  - 1 [NP-CLR, PP-CLR]
  - 1 [ADVP-CLR]
  - 1 [SBAR-NOM]
  - 1 [PRT]
  - 2 [PRT, PP-CLR]
  - 1 [PP-CLR, S-CLR]
  - 8 []
  - 1 [PP]
  - 2 [PP-CLR, PP]
  - 6 [PP-CLR]
  - 24 [NP, PP]
  - 8 [NP, PP-CLR]
  - 87 [NP]

can’t simply delete ADV everywhere
SBAR-NOM

- headless relative
- Case 16409
-CLR

- closely related ("middle ground between arguments and adjuncts")
- Cases 37840 and 21224 (PP-CLR PP)
PRT

- particle
- Cases 18521, 22356 and 5824

do we treat join up/in as different from join?

Cogsci 2007 Workshop:
Psychocomputational Models
Verbs of Combining and Attaching

References: Condoravdi and Sanfilippo (1990), Gentner (1978)

These verbs are all related to combining or attaching. Their hallmark is participation in the simple reciprocal alternations, the *together* reciprocal alternations, or both. Members of this class are never found in the *apart* reciprocal alternations. The various subclasses differ according to whether the meanings of their members involve a result or means component.

22.1 Mix Verbs

Class Members:

*with*: blend, combine, commingle, concatenate, connect, fuse, join, link, merge, mingle, mix, pool

*into*: blend, cream, mix

*to*: add, connect, join, link, network

Properties:

1. Simple Reciprocal Alternation (transitive):
   a. Herman mixed the eggs with the cream. (prepositional variant)
   b. Herman mixed the eggs and the cream. (reciprocal variant)

2. Simple Reciprocal Alternation (intransitive; most verbs):
   a. The eggs mixed with the cream.
   b. The eggs and the cream mixed.

   a. Herman mixed the eggs with the cream.
   b. Herman mixed the eggs and the cream *together*.

4. *Together* Reciprocal Alternation (intransitive; most verbs):
   a. The eggs mixed with the cream.
   b. The eggs and the cream mixed *together*.

5. Causative/Inchoative Alternation (most verbs):
   a. I mixed the soap into the water.
   b. I mixed the soap and the water.

6. Middle Alternation:
   a. I mixed the eggs with cream.
   b. I mixed the eggs and cream (*together*).

Join belongs to section 22.1

Mix verbs

Syntactic frames:

- 84 NP
- 11 NP PP
- 1 NP-and (together)
- 7 PP
- 8 []
- ADJ PP-with
- ADJ (together)
WSJ PTB vs. EVCA

- **WSJ PTB**
  - 1 [NP-CLR, PP-CLR]
  - 1 [ADVP-CLR]
  - 1 [PP-CLR, S-CLR]
  - 8 []
  - 1 [PP]
  - 2 [PP-CLR, PP]
  - 6 [PP-CLR]
  - 24 [NP, PP]
  - 8 [NP, PP-CLR]
  - 87 [NP]

- **EVCA**
  - NP PP-with
  - PP-with
  - []
  - NP-and (together)
  - ADJ PP-with
  - ADJ (together)

- **Note:** ADJ is JJ in PTB tagset
Further work on WSJ PTB

- WSJ PTB [ADVP-CLR]
  - Recently, some 60 environmental and outdoor groups representing such divergent points of view as the Sierra Club, the League of Women Voters and the National Rifle Association joined together to request a reassessment of the environmentally unsound Central Utah Project.

- EVCA
  - NP-and (together)

Cogsci 2007 Workshop:
Psychocomputational Models
Further work on WSJ PTB

- PP-CLR for *join*
  - always headed by *with*?
    - *in 4*
    - *with 11*
    - *as 5*

- PP for *join* headed by?
  - *for 1*
  - *upon 1*
  - *by 3*
  - *on 1*
  - *from 4*
  - *in 8*
  - *as 8*

*with* is always a PP-CLR for *join*
Further work on WSJ PTB

- Intransitive cases
- 37056, 24842, 20028, 11626, 11625, 11556, 8500, 5710

- other carriers to join
- others to join
- several other companies to join
- to join
- Britain would join SBAR-TMP
- Several of the New York Stock Exchange's own listed companies, led by giant Contel Corp., are joining
- most of the smaller makers joined [PP-LOC under the Microsoft Corp. umbrella]
- he joins [NP-TMP Nov. 13th]
• What would you expect the Collins parser to say about the verb alternations?

• Example:
  - Collins parser is trained on the PTB
  - we might expect logprob rankings to reflect frequencies of the various VP frames in the PTB
  - if they don’t, why not?

• Example:
  - do the logprob values reflect your personal intuitions or expectations about the sentences?

• Example:
  - are the differences between logprob values noise or significant?
  - e.g. -99.1/-99.7 vs. -99.1/-200.4
  - what is the general effect of sentence length?
  - e.g. what is the typical penalty for adding a word to a sentence?
  - e.g. are ungrammatical sentences assigned radically different logprobs?
  - e.g. can the Collins parser distinguish valid and invalid alternations for a given verb?

Cogsci 2007 Workshop:
Psychocomputational Models
• What would you expect the Collins parser to say about the alternations?

• Focus on discrepancies:
  – Collins vs. PTB
  – Collins vs. Linguistic Intuition
  – PTB vs. Linguistic Intuition

• Question:
  – are the discrepancies regular or random?
  – e.g. one verb patterns one way, another a different way
A case study: *join, merge*

- “Bristol-Meyers agreed to merge with Sun Microsystems”
- “Boeing and Sun Microsystems agreed to merge”
- Which would be more likely? Which is more likely? Which ‘should be’ more likely (according to linguistic accounts)
Some counts

- join - 49 VB
- mix - 1
- water 114 NN
- 24 milk NN
- 14 toys NN
- 207 computers NNS
Some sentences

- John NNP mixed VBD the DT water NN and CC the DT milk NN
- John NNP mixed VBD the DT milk NN and CC the DT water NN
- John NNP mixed VBD the DT water NN with IN the DT milk NN
- John NNP mixed VBD the DT milk NN with IN the DT water NN
- John NNP joined VBD the DT water NN and CC the DT milk NN
- John NNP joined VBD the DT milk NN and CC the DT water NN
- John NNP joined VBD the DT water NN with IN the DT milk NN
- John NNP joined VBD the DT milk NN with IN the DT water NN
- John NNP joined VBD the DT water NN and CC the DT water NN
- John NNP joined VBD the DT water NN with IN the DT water NN
- John NNP joined VBD the DT computers NNS and CC the DT computers NNS
- John NNP joined VBD the DT computer NNS with IN the DT computer NNS
The envelope please…

- log prob: (closer to 0 = more likely)
  -55.6292
  -54.307
  -54.3957
  -51.2094
  -48.4139
  -46.1579
  -46.1015
  -43.0599

- J. mixed the water and the milk
- J. mixed the milk and the water
- J. mixed the water with the milk
- J. mixed the milk with the water
- J. joined the water and the milk
- J. joined the water with the milk
- J. joined the milk with the water
- J. joined the water with the milk
And more

- John joined the computers and the computers  $-39.699$
- John joined the computers with the computers  $-43.054$
- John joined the milk and the milk  $-48.0987$
- John joined the milk with the milk  $-46.3324$
First of all…

- John mixed the water with the milk
Then
• John mixed the milk with the water

Hmm… what about ‘mixed’? Try ‘join’
‘Join’

- J. joined the water and the milk
‘Join’

- John joined the milk with the water
In fact...

- No matter what lexical item we choose, ‘milk’ (but not ‘water’ or ‘toys’ or ‘computer’) forces a low attachment like this – all the others, in all other combinations, force the high PP attachment...
Where do the numbers come from? A breakdown

- John joined the water and the milk
- John joined the milk and the water
Where do the numbers come from?

• J. joined the water with the milk
• J. joined the milk with the water
24 ‘milk’ sentences, only a few as a noun...

Cogsci 2007 Workshop:
Psychocomputational Models
#21273

Cogsci 2007 Workshop:
Psychocomputational Models
#21382, 21314

NP
- PP
  - IN range for
  - NP milk
  - CC or
  - NP grain

#23482, 23488

NP
- NP
  - , NN corn
  - CC and
  - NN milk
Cogsci 2007 Workshop:
Psychocomputational Models
Another example: ‘bark’

(a) The dog barked to Mary.
(b) The dog barked about Mary.
(c) The dog barked at Mary.

bark: 1 
the: 33164
dog: 12
: to: 20655
: Mary: 31
: about: 2508
: at: 5364
### Table 26: (26a) *The dog barked to Mary.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Probability of Sentence</th>
<th>Probability of VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-36.5968</td>
<td>-17.117</td>
</tr>
<tr>
<td>2</td>
<td>-36.4098</td>
<td>-17.0518</td>
</tr>
<tr>
<td>3</td>
<td>-36.4164</td>
<td>-17.0758</td>
</tr>
</tbody>
</table>

### Table 27: (26b) *The dog barked about Mary.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Probability of Sentence</th>
<th>Probability of VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-40.5624</td>
<td>-21.0827</td>
</tr>
<tr>
<td>2</td>
<td>-40.4081</td>
<td>-21.0501</td>
</tr>
<tr>
<td>3</td>
<td>-40.3831</td>
<td>-21.0425</td>
</tr>
</tbody>
</table>

### Table 28: (26c) *The dog barked at Mary.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Probability of Sentence</th>
<th>Probability of VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-37.4994</td>
<td>-18.0197</td>
</tr>
<tr>
<td>2</td>
<td>-37.2983</td>
<td>-17.9403</td>
</tr>
<tr>
<td>3</td>
<td>-37.3178</td>
<td>-17.9772</td>
</tr>
</tbody>
</table>
Entries in the lexicon are Root Lexical Conceptual Structures, RLCS

*-marked positions are unified with other compatible RLCSs during composition.
Lexical Conceptual Structure

• Each node contains:
  • **Primitive:**
    • CLOSED CLASS: GO, STAY, BE, ON, IN, AT…
    • OPEN CLASS: JOHN, RUN-INGLY, …
  • **Field:** Analogy to motion/position in Localist approach: LOCATIONAL, POSSESSSIONAL, TEMPORAL,…
  • **Type:** EVENT, STATE, PATH, POSITION, MANNER…
LCS Structure & Composition

Recursively compose the children, then assign the composed children to *-marked positions in the current RLCS. This yields a Composed LCS, CLCS.

The ball rolled towards Beth.

Roll: "Roll"

Towards: "towards"

Roll: "GO"

*(theme) *path ROLL

(theme) AT

(theme) (goal)
Language of Thought (LOT) 
(Fodor 1975)

• Children acquiring a language are mapping words onto internal language
• This internal language **cannot** be induced on the basis of language learning – Why?

• Are the lexical semantics primitives the LOT?
Structural vs. Content Meaning Component

- Verbs in a class share a structural component
- Verbs in a class distinguished by content component
What information is in the lexicon?

Answer 1: structural info encodes agent, patient, goal,… (Hale & Keyser)

Answer 2: nothing but vanilla syntactic categories (N, V, P, …)
### Hypothesis 1 Problem

Problem: How does **Lexicon** acquire the following:

<table>
<thead>
<tr>
<th>Word</th>
<th>Role</th>
<th>Property 1</th>
<th>Property 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>/shelf/</td>
<td>n_{LOCATION}</td>
<td>p_{LOCATION} = d</td>
<td>v</td>
</tr>
<tr>
<td>/butter/</td>
<td>n_{LOCATUM}</td>
<td>p_{LOCATUM} = d</td>
<td>v</td>
</tr>
<tr>
<td>/shovel/</td>
<td>n_{INST-MOT}</td>
<td>p_{INST-MOT} = p_{LOCATION} = d</td>
<td>v</td>
</tr>
<tr>
<td>/pencil/</td>
<td>n_{INST-IMP}</td>
<td>p_{INST-IMP} = p_{LOCATION} = d</td>
<td>v</td>
</tr>
<tr>
<td>/mop/</td>
<td>n_{INST-REMOVAL}</td>
<td>p_{INST-REMOVAL} = p_{SOURCE} = d</td>
<td>v</td>
</tr>
<tr>
<td>/email/</td>
<td>n_{INST-COMM}</td>
<td>p_{INST-COMM} = p_{HAVE} = d</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p_{INST-COMM} = p_{DEST} = d</td>
<td>v</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypo 1: /put/ vs. /shelf/ vs. /butter/

```
  vP
   \   /
   N v'  /put/
  /Bob/ VP
   \   //
      N V'  PP
        /book/  P'
           /put/  P
               //
      N LOCATION
        /on/
```

Bob put the book on the shelf.
Bob put the book with the others.
Bob put the book on the shelf.
Bob put the book with the others.
/put/ vs. /shelf/ vs. /butter/

/vP
  \N v'
/v
  \N v'
/vP
  \N v'
/book/
/v
  \N v'
PP
/shelf/
\P'
LOCATION
/on/
windowsill/

Bob shelved the book on the windowsill.
/put/ vs. /shelf/ vs. /butter/
Bob put the book on the shelf.
Bob put the book with the others.
He buttered the bread with margarine.

/impose PLOCATION on arguments
/impose PLOCATUM on arguments
Hypothesis 1: **Lexicon** Contains Selection Criteria

/shelf/ has \( p_{LOCATION} \) selection in lexicon \((= p_{LOCATION} = d(et) \ V)\)

Also: /shelf/ is \( n_{LOCATION} \)

/butter/ has \( p_{LOCATUM} \) selection in lexicon \((= p_{LOCATUM} = d(et) \ V)\)

Also: /butter/ is \( n_{LOCATUM} \)

So then the **Lexicon** cannot derive:

1. Bob shelved the windowsill with the book.
2. Bob buttered the margarine onto the bread.

Information about butter and shelf – *where* is it located?
What to do?

Solution 1: Solve the above problem

Solution 2: Push problem OUT of Lexicon and INTO Encyclopedia
Solution 2: Push problem OUT of Lexicon and INTO Encyclopedia

Encyclopedia, not lexicon, is source of 'Oddness' of:

# (1) Bob shelved the windowsill with the book
# (2) Bob buttered the margarine onto the bread

Lexicon is NOT:

/\text{shelf}/ =_{p\text{LOCATION}} =d(\text{et}) \ V /\text{butter}/ =d +k p_{\text{LOCATUM}}

/\text{into}/ =d +\text{case} \ p_{\text{LOCATION}} /\text{with}/ =d +\text{case} p_{\text{LOCATUM}}

But instead:

/\text{shelf}/ =p =d V /\text{butter}/ =d +\text{case} p

/\text{into}/ =d +\text{case} p /\text{with}/ =d +\text{case} p

Thus insofar as the lexicon is concerned,

(1) and (2) are GRAMMATICAL
WordNet
(Miller et al 1998)

- Widely used in computational linguistics
- Dictionary-like definitions organized by links:
  - Nouns: X is a kind-of/part-of Y
  - Verbs: X involves doing Y
    - Also with common syntactic frames
  - Other than the above, no conceptual structure, no meaning postulates
- Enumerates lists of senses, does not relate these senses
Senses

• How many senses per a word? WordNet examples:
  • bank – 10 noun senses, 8 verb senses
  • have – 1 noun sense, 19 verb senses
  • smoke – 8 noun sense, 2 verb senses
• Are these different senses? How are they structurally related?
  • relating them structurally requires conceptual metalanguage
Meaning isn’t (always) at the Word Level

pick up, throw up, turn on does NOT have picking, throwing, turning (at least not directly)

Antidisestablishmentarism
   (morphosemantics theory very poor)

And there is pragmatics (too large a topic)
Idioms / Constructions

- Are idioms to be stored in the lexicon?
  - Examples:
    - Kicked the bucket, Paint the town red
    - Spic-and-span, kit and kaboodle
    - What’s X doing Y? The X-er, The Y-er
  - H1: Yes
    - BUT then: how do you treat Tense, agreement, …
  - H2: No
    - BUT then: then where is “meaning” stored?
      - Answer: the encyclopedia
        - But that is a non-answer
Failure to Compose

- **Defeasability:**
  
  He climbed the mountain vs He climbed down the mountain

- **Red hair vs red Porsche**
  - Does this work? Red(x) & Hair(x)
  - Meaning of RED in context > outside context?

- **Former friend**
  - Does this work? Former(x) & Friend(x)

- **Good knife vs good book vs good life**
  - Does this work? Good(x) & Knife(x)
  - Good knives cut well, Good books …, Good lives …
Metaphoric Meanings

• “No silver bullet on 9/11”
• “My surgeon is a butcher” vs “My butcher is a surgeon”
• “Don’t get high on Montague grammar”
• Appears way more often than you think.
Meanings are highly private

Before they become adults, children think:

/uncle/ is a friendly middle-aged man
/island/ is a beachy area with palm trees
/two/ is some small number greater than one
and not anything like

Blind children’s meaning of LOOK

Cogsci 2007 Workshop:
Psychocomputational Models
What can one do?
Let’s …

Show why everyone is wrong (Fodor 1998)

Promise: negative control
Unpromise: Has atypical ideas on what it means to “have” a concept

Summarize corpora statistically

\[ P(V_{142}|D) = .011 \]
\[ P(V_{143}|D) = .004 \]
\[ P(V_{144}|D) = .0014 \]

Promise: Helps parsing.
Unpromise:
  (1) Why parse?
  (2) This is a mere redescription
  • I thought the book to Mary.
Let’s …

Build robots

Promise: Machine Learning used to get /apple/ associated to RED, …

Unpromise: only as good as your concept metalanguage, which is sensorimotor by nature. Reading minds is much harder.

Collect knowledge from people

Promise: If machines could understand what is collected, Plus-X goes away.

Unpromise: (1) IF (2) Data without a theory.

Cogsci 2007 Workshop:
Psychocomputational Models
Let’s

- Figure out how children learn
- John joined the eggs with the cream
- John joined the eggs and the cream

*How robust is the Collins Parser?*

*Would lexical selection be a factor in comparing logprobs?*

- John joined the cream with the eggs
- $[_{PP} P(with) [_{NP} DET cream]]$
- $[_{PP} P(with) [_{NP} DET eggs]]$
- how about noun complements of preposition with that actually occur in the WSJ PTB vs. those that don't?

*WSJ PTB frequencies:*
- 5 Herman
- 442 John
- 28 eggs
- 3 cream
File:
Herman joined the eggs with the cream.
Herman joined the eggs and the cream.
Herman joined the cream with the eggs.
Herman joined the cream and the eggs.
John joined the eggs with the cream.
John joined the eggs and the cream.
John joined the cream with the eggs.
John joined the cream and the eggs.

MXPOST:
Herman_NNP joined_VBD the_DT eggs_NNS with_IN the_DT cream_NN .
Herman_NNP joined_VBD the_DT eggs_NNS and_CC the_DT cream_NN .
Herman_NNP joined_VBD the_DT cream_NN with_IN the_DT eggs_NNS .
Herman_NNP joined_VBD the_DT cream_NN and_CC the_DT eggs_NNS .
John_NNP joined_VBD the_DT eggs_NNS with_IN the_DT cream_NN .
John_NNP joined_VBD the_DT eggs_NNS and_CC the_DT cream_NN .
John_NNP joined_VBD the_DT cream_NN with_IN the_DT eggs_NNS .
John_NNP joined_VBD the_DT cream_NN and_CC the_DT eggs_NNS .
Analyzing the probabilities

Collins input format:
8 Herman NNP joined VBD the DT eggs NNS with IN the DT cream NN ..
8 Herman NNP joined VBD the DT eggs NNS and CC the DT cream NN ..
8 Herman NNP joined VBD the DT cream NN with IN the DT eggs NNS ..
8 Herman NNP joined VBD the DT eggs NNS and CC the DT cream NN ..

8 John NNP joined VBD the DT eggs NNS with IN the DT cream NN ..
8 John NNP joined VBD the DT eggs NNS and CC the DT cream NN ..
8 John NNP joined VBD the DT cream NN with IN the DT eggs NNS ..
8 John NNP joined VBD the DT cream NN and CC the DT eggs NNS ..

Output (model 1):
PROB 762 -36.766 0
PROB 888 -39.1046 0
PROB 935 -35.825 0
PROB 873 -36.8356 0
PROB 746 -40.9796 0
PROB 863 -43.3183 0
PROB 909 -40.0386 0
PROB 841 -41.0492 0

Sample Analysis:
- Herman/John change doesn’t affect the logprob rankings (as expected)
- NP with NP < NP and NP
- cream then eggs

But too few Hermans!

Cogsci 2007 Workshop:
Psychocomputational Models
Lexicon does NOT hold real-world knowledge, only:

<table>
<thead>
<tr>
<th>ROOT</th>
<th>Lexicon</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>arrive</td>
<td>+v, +DP, -cause</td>
<td>John arrived. The arrival of John</td>
</tr>
<tr>
<td>big</td>
<td>-v, +DP</td>
<td>The big X.</td>
</tr>
<tr>
<td>open</td>
<td>±v, +DP, ±cause</td>
<td>John opened X. X opened.</td>
</tr>
<tr>
<td>destroy</td>
<td>+v, +DP, +cause</td>
<td>John destroyed X. John's destruction of X.</td>
</tr>
</tbody>
</table>