Distinguishing Possible and Probable in Linguistic Theory

James Pustejovsky
Brandeis University

PRELIM, 2014
Charles University, Prague, Czech Republic
Faculty of Mathematics and Physics
July 7, 2014
Starved of adequate data, linguistics languished. . . . It became fashionable to look inwards to the mind rather than outwards to society.

*Sinclair (1991)*
Polysemy is all around
Coercion in Contextual Interpretation
Linguistic modulations reflect conceptual shifts in thought
Inherent tension between corpus data and theory
Probabilistic judgments for Compositional Operations
Questions

- How do words combine to make meanings?
- How do word meanings change in composition?
- How do we explain creative word use?
- How can linguistic models account for variability in language use?
Starting Assumptions

- Language meaning is *compositional*.
- **Compositionality** is a desirable property of a semantic model.
- Many linguistic phenomena appear **non-compositional**.
- **Generative Lexicon** exploits richer representations and rules to enhance compositional mechanisms.
- But semantics of words seems to encode **probabilistic conditions** on type selection.
- **Richer compositional models** are needed to accommodate such observed behavior.
- So, type theory needs to address probabilistic notions inherently, **GL included**.
van Eijck and Lappin (2013)

- Chomsky’s thesis: Natural languages can be described as formal systems.
- Montague’s thesis: Natural languages can be described as interpreted formal systems.
- The Harris-Jelinek thesis: Natural languages can be described as information theoretic systems, using stochastic models that express the distributional properties of their elements.
The Harris-Jelinek Thesis

- The Harris-Jelinek thesis implies the The Language Model Hypothesis (LMH) for syntax, which holds that grammatical knowledge is represented as a stochastic language model.
- On this hypothesis, a speaker acquires a probability distribution over the strings constituting the sentences of a language.
- This distribution is generated by a probabilistic automaton or a probabilistic grammar, which assigns a structure to a string with a probability that is the product of the rules applied in the derivation of that string.
- The probability of the string itself is the sum of the parses that the grammar generates for it.
- This probability represents the likelihood of a sentence’s occurrence in a corpus.
- Lexically-derived relations like synonymy, antinomy, polysemy, and hyponymy are prone to clustering and overlap effects.
Data in Linguistics

Pustejovsky and Hanks (2014)

- Theory driven Naturally Elicited Data (NED)
- Naturally Occurring Data (NOD) Contradict Theory
- Revisions to Theory accounting for NOD

- Post-Bloomfield Structuralism:
  Harris, Bar Hillel, Chomsky, Hockett,

- Transformational Grammars:
  Harris, Bar Hillel, Chomsky
Procedures

- **Discovery Procedure**: the theory must provide a practical and mechanical method for actually constructing the grammar given a corpus of utterances. Chomsky 1957

- **Decision Procedure**: the theory must provide a practical and mechanical method for determining whether or not a grammar proposed for a given corpus is in fact the best grammar. Chomsky 1957

- **Evaluation Procedure**: given a corpus and two grammars, G1 and G2, the theory must tell us which is the better grammar of the language from which the corpus is drawn. Chomsky 1957
1. What elements can select?
2. What is an argument?
3. What does it mean for a predicate to select an argument?
4. How does selection relate to composition and lexical decomposition?
(1) a. **Verb**: V How do we decompose the meaning?  
   b. **Arguments**: x, y, z, ...

(2) a. **Body**: the predicate, with bound variables.  
   b. **Arguments**: the parameter list.

\[
\begin{align*}
\text{Args} & \quad \text{Body} \\
\lambda x_i & \quad [\Phi]
\end{align*}
\]
Decomposition Strategies

1. atomic predication: do nothing, $P(x_1)$
2. add arguments: $P(x_1) \implies P(x_1, x_2)$
3. split the predicate: $P \implies P_1, P_2$
4. add and split: $P(x_1) \implies P(x_1, x_2), P_2(x_2)$
Richer typing for arguments:

1. Identifies specific predicates in the body of the expression that are characteristic functions of an argument;

2. pulls this subset of predicates out of the body, and creates a pretest to the expression as a restricted quantification over a domain of sorts, denoted by that set of predicates.
Types from Predicative Content

\[ \lambda x_2 \lambda x_1 [\Phi_1, \ldots, \Phi_{x_1}, \ldots, \Phi_{x_2}, \ldots, \Phi_k] \]

\[ \lambda x_2 : \sigma \lambda x_1 : \tau [\Phi_1, \ldots, \Phi_k - \{\Phi_{x_1}, \Phi_{x_2}\}] \]

\(\sigma\) and \(\tau\) have now become reified as types on the arguments.
Arguments can be viewed as encoding **pretests** for performing the action in the predicate.

If the **argument condition (i.e., its type)** is not satisfied, the predicate either:

- **fails** to be interpreted (strong selection);
- **coerces** its argument according to a given set of strategies.
Arguments can be viewed as encoding probability distributions of pretests for performing the action in the predicate.
(1) a. **formal**: the basic category of which distinguishes the meaning of a word within a larger domain;  
b. **constitutive**: the relation between an object and its constituent parts;  
c. **telic**: the purpose or function of the object, if there is one;  
d. **agentive**: the factors involved in the object’s origins or “coming into being”.

1. $e$ the general type of entities; $t$ the type of truth values.
   ( $\sigma$, $\tau$ range over all simple types, and subtypes of $e$.)
2. If $\sigma$ and $\tau$ are types, then so is $\sigma \rightarrow \tau$.
3. If $\sigma$ and $\tau$ are types, then so is $\sigma \otimes R \tau$; $R$ ranges over $A$ or $T$.
4. If $\sigma$ and $\tau$ are types, then so is $\sigma \bullet \tau$. 
(2) a. **Natural types:**

- **Simple:** Natural kind concepts consisting of reference only to Formal or Constitutive qualia roles;
- **Functional:** Additional reference to Telic (purpose or function)

b. **Artifactual types:** Concepts making reference to Agentive (origin) for a specific Telic (purpose or function);

c. **Complex types:** Concepts integrating reference to a logical coherence relation between types from the other two levels.
Entities formed from the application of the FORMAL and/or CONST qualia roles:

1. For the predicates below, $e_N$ is structured as a join semi-lattice, $\langle e_N, \sqsubseteq \rangle$;
2. physical, human, stick, lion, pebble
3. water, sky, rock
Predicates formed with \textbf{Natural Entities} as arguments:

1. \textit{fall}: $e_N \rightarrow t$
2. \textit{touch}: $e_N \rightarrow (e_N \rightarrow t)$
3. \textit{be under}: $e_N \rightarrow (e_N \rightarrow t)$

a. $\lambda x : e_N[\text{fall}(x)]$

b. $\lambda y : e_N \lambda x : e_N[\text{touch}(x,y)]$

c. $\lambda y : e_N \lambda x : e_N[\text{be-under}(x,y)]$
Artifactual Entity Types

Entities formed from the Naturals by adding the AGENTIVE or TELIC qualia roles:

1. Artifact Entity: $x : e_N \otimes_a \sigma$
   $x$ exists because of event $\sigma$

2. Functional Entity: $x : e_N \otimes_t \tau$
   the purpose of $x$ is $\tau$

3. Functional Artifactual Entity: $x : (e_N \otimes_a \sigma) \otimes_t \tau$
   $x$ exists because of event $\sigma$ for the purpose $\tau$

   a. beer: $(\text{liquid} \otimes_a \text{brew}) \otimes_t \text{drink}$
   b. knife: $(\text{phys} \otimes_a \text{make}) \otimes_t \text{cut}$
   c. house: $(\text{phys} \otimes_a \text{build}) \otimes_t \text{live\_in}$
Artifactual Predicate Types

Predicates formed with Artifactual Entities as arguments:

1. \textit{spoil}: \( e_N \otimes t \tau \rightarrow t \)
2. \textit{fix}: \( e_N \otimes t \tau \rightarrow (e_N \rightarrow t) \)

a. \( \lambda x : e_A[spoil(x)] \)
b. \( \lambda y : e_A \lambda x : e_N[fix(x,y)] \)

- The beer spoiled.
- Mary fixed the watch.
Complex Entity Types

Entities formed from the **Naturals** and **Artifactuals** by a **product type** between the entities, i.e., the dot, •.

1. a. Mary doesn’t believe the book.
2. a. The exam started at noon.
   b. The students could not understand the exam.
Motivating Dot Objects

When a single word or phrase has the ability to appear in selected contexts that are contradictory in type specification.

If a lexical expression, \( \alpha \), where \( \sigma \sqcap \tau = \bot \):

1. \([\_\_]_\sigma X\)
2. \([\_\_]_\tau Y\)

are both well-formed predications, then \( \alpha \) is a dot object (complex type).
1. **Act**\hspace{2pt}**Proposition**: promise, allegation, lie
   - I doubt *John’s promise of marriage*.
   - *John’s promise of marriage* happened while we were in Prague.

2. **Attribute**\hspace{2pt}**Value**: temperature, weight, height, tension, strength
   - The temperature is rising.
   - The temperature is 23.

1. **Event**\hspace{2pt}**Information**: lecture, play, seminar, exam, quiz, test
   a. My lecture lasted an hour.
   b. Nobody understood *my lecture*.

2. **Event**\hspace{2pt}**Music**: sonata, symphony, song, performance, concert
   a. Mary couldn’t hear *the concert*.
   b. The rain started during *the concert*.
1. **Event**\(\bullet\)**Physical**: lunch, breakfast, dinner, tea
   a. **My lunch** lasted too long today.
   b. I pack **my lunch** on Thursdays.

2. **Information**\(\bullet\)**Physical**: book, cd, dvd, dictionary, diary, mail, email, mail, letter
   a. Mary burned **my book on Darwin**.
   b. Mary believes **all of Chomsky’s books**.

1. **Organization**\(\bullet\)**(Information**\bullet\)**Physical**: magazine, newspaper, journal
   a. **The magazine** fired its editor.
   b. The cup is on top of **the magazine**.
   c. I disagreed with **the magazine**.

2. **Process**\(\bullet\)**Result**: construction, depiction, imitation, portrayal, reference
   a. Linnaeus’s **classification of the species** took 25 years.
   b. Linnaeus’s **classification** contains 12,100 species.

2. a. Mary answered every question in the class.
   b. Mary repeated every question in the class.
Predicates formed with a **Complex Entity Type** as an argument:

1. \( \text{read}: \text{phys} \bullet \text{info} \rightarrow (\text{e}_N \rightarrow t) \)

2. **Expressed as typed arguments in a \( \lambda \)-expression:**

   \[ \lambda y : \text{phys} \bullet \text{info} \lambda x : \text{e}_N[\text{read}(x,y)] \]

3. Mary read the book.
(3) a. **PURE SELECTION** (Type Matching): the type a function requires is directly satisfied by the argument;
b. **ACCOMMODATION**: the type a function requires is inherited by the argument;
c. **TYPE COERCION**: the type a function requires is imposed on the argument type. This is accomplished by either:
   i. **Exploitation**: taking a part of the argument’s type to satisfy the function;
   ii. **Introduction**: wrapping the argument with the type required by the function.
Direct Argument Selection

- The spokesman denied the statement (PROPOSITION).
- The child threw the ball (PHYSICAL OBJECT).
- The audience didn’t believe the rumor (PROPOSITION).
1. The rock fell.
1. The beer spoiled.

\[
\begin{align*}
S & \\
NP & \sigma \otimes T \tau \\
liquid \otimes_T drink & : e_A \\
the \ beer & \\
spoiled & \\
\lambda x : e_A[\text{spoil}(x)] & 
\end{align*}
\]
1. John read the book.

\[
\lambda y : p \bullet i \lambda x : e_N[read(x,y)] \\
\text{the} \\
\text{book}
\]
Coercion of Arguments

- The president denied the attack.
  \[ \text{EVENT} \rightarrow \text{PROPOSITION} \]

- The White House denied this statement.
  \[ \text{LOCATION} \rightarrow \text{HUMAN} \]

- This book explains the theory of relativity.
  \[ \text{PHYS} \bullet \text{INFO} \rightarrow \text{human} \]

- d. The Boston office called with an update.
  \[ \text{EVENT} \rightarrow \text{INFO} \]
1. The water spoiled.

\[ \lambda x : e_A[\text{spoil}(x)] \]
John read the rumor.

\[
\begin{align*}
\lambda y : p \cdot i \lambda x : e_N[\text{read}(x,y)] & \quad \text{Det} \quad \text{the} \\
\text{VP} & \quad \text{NP:info} \\
\text{V} & \quad \text{phys \bullet info}
\end{align*}
\]
1. Mary enjoyed her coffee.
1. Mary enjoyed her coffee.
Type Coercion: Dot Exploitation

1. The police burned the book.
2. Mary believes the book.

\[ \lambda y : \text{phys} \lambda x : e_N[burn(x,y)] \]

Diagram:

```
           VP
              |
              V     phys
             /     |
            burn   Det
              |
             N
```

\text{NP: phys • info}

James Pustejovsky  Brandeis University  Possible/Probable in Linguistics
<table>
<thead>
<tr>
<th>Argument Type</th>
<th>Natural</th>
<th>Artifactual</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>Sel/Acc</td>
<td>Tensor (Qualia) Intro</td>
<td>Dot Intro</td>
</tr>
<tr>
<td>Artifactual</td>
<td>Tensor Exploit Left (Acc)</td>
<td>Sel/Acc</td>
<td>Dot Intro</td>
</tr>
<tr>
<td>Complex</td>
<td>Dot Exploit</td>
<td>Dot Exploit</td>
<td>Sel/Acc</td>
</tr>
</tbody>
</table>
Assuming our theory has a type structure, $\mathcal{T}$:

- and compositional operations of coercion mentioned above:

  - What coercions occur in real corpus data?
  - What are the distributions of the different compositional mechanisms?
(4) book (*phys* • *info*)

Object

a. *phys*: close, open, shut, throw away, steal, keep, burn, put away, bind, design, store, grab, drop, destroy, dust, hold, shelve, pile, store

b. *info*: ban, consult, edit, find interesting, study, translate, review, love, judge, revise, examine, like, describe, discuss

’Jess almost dropped the book, then hastily replaced it on the shelf’

’The author will be discussing her new book’
Dot exploitation

\( (5) \) \textit{house (phys \bullet loc)}

Object

\begin{itemize}
\item \textit{phys}: built, buy, sell, rent, own, demolish, renovate, burn down, erect, destroy, paint, inherit, repair
\item \textit{loc}: leave, enter, occupy, visit, inhabit, reach, approach, evacuate, inspect, abandon
\end{itemize}

’they built these houses onto the back of the park’
’the bus has passed him as he left the house’
(6) *speech* (*event* • *info*)

Object

a. *event*: deliver, make, give, finish, interrupt, conclude, end, begin, start, complete, cut (short), open

b. *info*: analyse, interpret, understand, quote, applaud, criticize, condemn, revise, translate, oppose, appreciate

’He was forced to interrupt his speech while order was restored’

’US officials condemned the speech’
(7) *exit* (*event* • *loc*)

Object

- *event*: make, facilitate, follow, force, hasten, register
- *loc*: block, bar, take, find, mark, indicate, reach, choose, locate

’I very swiftly made my exit through the door’

’She was blocking the exit of a big supermarket’

Examples (4-7) show that the single aspects (senses) of a dot object are often picked up separately. Many lexical items which are typed as dots tend to show up in text in just one of their aspects instead of both.
As noted by Jezek and Lenci (2007) with respect to the object position of the complex type \textit{phys • info} (i.e. letter, article, book, novel etc.): It. \textit{articolo} ‘article’ combines more frequently with info-selectors rather than with phys-selectors:

(8) \textit{articolo (phys • info)}

Object

a. \textit{phys}: spostare ‘move’, ritagliare ‘cut out’

‘ritaglia tutti gli articoli che lo riguardano’
he cuts out all the articles about him
‘condivido interamente il suo articolo’
I agree entirely with his article
Jezek and Lenci (2007) also note that lexical items realizing the same dot type exhibit interesting variations as far as their asymmetry goes: for example in object position *romanzo* 'novel' avoids the *phys* sense more than *libro* 'book' does.

(9) *romanzo* 'novel' (*phys* • *info*)
   Object
   a. *phys*: collocare 'place', portare 'carry'

(10) *libro* 'book' (*phys* • *info*)
   Object
   a. *phys*: bruciare 'burn', portare 'carry', distruiggere 'destroy', rubare 'steal', conservare 'keep', custodire 'keep', buttare 'throw away'
Asymmetry of use can be a property of some dots, regardless of what argument they occupy. Both *door* and *gate* (*phys* • *aperture*) show preference for the *phys* interpretation in all arguments:

(11) **door** (*phys* • *aperture*)

Object

a. *phys*: open, shut, close, slam, push, pull, bolt, bang, kick, knock, smash, hold, open, paint, lock, fasten, secure, hit, remove, damage, replace, decorate

b. *aperture*: pass, enter, block

Subject

a. *phys*: open, slam, close, swing, shut, bang, burst open, click open, fly open, slide open, click shut, hang, face, shake

b. *aperture*: lead, go, give access, connect

’somewhere in the house a door slammed’

‘the main door went into a small lobby’
Interview (event • info) shows a distinct preference for the event interpretation in both subject and object position:

(12) interview (event • info)

Object

a. event: conduct, give, arrange, attend, carry out, terminate, conclude, close, complete, end, hold, cancel, undertake, extend, control, continue, begin

b. info: structure, discuss, analyze, describe

Subject

a. event: last, go well, take place, follow, end, progress, begin, become tedious, precede, start, happen

b. info: covers, centre on, concern, focus on

’Officials will be conducting interviews over the next few days’
’Let’s discuss the interview’

Asymmetries of corpus use may be seen as an additional diagnostic in addition to co-predication for identifying dot objects.
Artifactual (or Tensor) Exploitation

(13) \textit{finish} (Body: 'bring to an end'; Arg: \textit{event})

Object

a. \textit{event}: journey, tour, treatment, survey, race, game, training, ironing, shopping

b. E-I, Q-E of $\text{phys} \otimes_{\text{telic}} \tau$: penicillin, sandwich, cigarette, cake, dessert, food

c. E-I, Q-E of $\text{liquid} \otimes_{\text{telic}} \tau$: drink, wine, beer, whisky, coke

'when they finished the wine, he stood up'

'just finish the penicillin first'
Naturals tend not to show up as object arguments of *finish*. This confirms the predictions of our model. Naturals are simple types with no Tensor attached: as such, they do not lend themselves to compositional operation of Qualia Exploitation, as artifactuals do. This is not a characteristic of aspectual verbs in general: some aspectual verbs just don’t coerce their arguments or they do it to a lesser extent. *Last* exhibits a few artifacts as subjects, and they are all re-interpreted as the interval of time for which their function holds:

(14) *last* (Body: ’occur over a certain time span’; Arg: *event*)

Subject

a. *event*: marriage, trial, siege, honeymoon, war, journey, strike, storm, rainfall

b. E-I, Q-E of *phys* ⊗ _telic_ τ: battery, cartridge

’the battery lasts 24 hours’

’the cartridge lasted three weeks’
Weak Coercive Verbs

Many non-aspectual event selectors (such as attend, avoid, prevent, cancel, delay, schedule, skip etc.) are 'weak' coercive verbs (i.e. the vast majority of their arguments are events: in principle, those which are not, are coerced - but see section 5.1.2 for further discussion):

(15) attend (Body: 'be present at'; Arg: event)

Object:

a. event: meeting, wedding, funeral, mass, game, ball, event, service, premiere

b. E-I, Q-E of $loc \otimes_{telic} \tau$: clinic, hospital, school, church, chapel

'about thirty-five close friends and relatives attended the wedding' 'for this investigation the patient must attend the clinic in the early morning' 'he no longer attends the church'
(16) avoid (Body: 'keep away from, stop oneself from'; Arg: event)
Object:

a. event: collision, contamination, clash, damage, accident, pregnancy, injury, question, arrest, starvation, war
b. E-I-Q-I of \( \text{phys} \otimes_{\text{telic}} \tau \): food
c. E-I-Q-I of \( \text{abstr} \otimes_{\text{telic}} \tau \): tax
d. E-I-Q-I of \( \text{loc} \otimes_{\text{telic}} \tau \): prison

'try to avoid fried food'
'you can’t avoid the inheritance tax in those circumstances'
'his wife avoided prison because she is five months pregnant'
Distributions Inform Theory: Two Case Studies

Pustejovský and Rumshisky (2008)

- Theory driven Naturally Elicited Data (NED)
- Naturally Occurring Data (NOD) Contradict Theory
- Revisions to Theory accounting for NOD
Thesis: Natural types are not selected by artifactual predicates without coercion.

(17) a. Natural Predicates: touch, sleep, smile
    b. Artifactual Predicates: repair, break, mend, spoil

These classes are defined by the type assigned to the arguments. For example, the type structure for the Natural predicate touch is shown in (18):

(18) \[
\begin{array}{c}
touch \\
\text{ARGSTR} = \begin{cases}
\text{ARG1} = x : \text{phys} \\
\text{ARG2} = y : \text{phys}
\end{cases}
\end{array}
\]
An Artifactual predicate such as the verb *repair* would be typed as shown in (19).

\[
\begin{align*}
\text{repair} \\
\text{ARGSTR} = \begin{bmatrix}
\text{ARG1} = x : \text{human} \\
\text{ARG2} = y : \text{phys} \otimes ^{\text{Telic}} \alpha
\end{bmatrix}
\end{align*}
\]

Given these theoretical assumptions, what we expect to encounter as the direct object of artifactual predicates such as *repair*, *fix*, and so forth, are entities that are themselves artifacts.

(20) a. Mary repaired the roof.
    b. John fixed the computer.
    c. The plumber fixed the sink.
    d. The man mended the fence.
Natural typed NPs should not appear as objects of artifactual predicates:

- Except under coercion interpretations
Case Study 2: Verbs Selecting for Propositions


(21) a. Mary believes [that the earth is flat].
    b. John knows [that the earth is round].
    c. John told Mary [that she is an idiot].
    d. Mary realizes [that she is mistaken].

\[
\begin{align*}
\text{believe} \\
\text{ARGSTR} = \\
\begin{bmatrix}
\text{ARG1} = x : \text{human} \\
\text{ARG2} = y : \text{info}
\end{bmatrix}
\end{align*}
\]

(22)

(23) a. Mary believed the book.
    b. John told me a lie.
    c. The man realized the truth.
(24) John memorized then burned the book.

The composition involved in a sentence like (23a) is illustrated below, where the informational component of the type structure for *book* is “exploited” to satisfy the type from the predicate.

(25)
Case Study 1: Results

<table>
<thead>
<tr>
<th>repair.v</th>
<th>fix.v</th>
<th>mend.v</th>
</tr>
</thead>
<tbody>
<tr>
<td>damage 107</td>
<td>pipe 9</td>
<td>fence 23</td>
</tr>
<tr>
<td>roof 16</td>
<td>gutter 4</td>
<td>shoe 10</td>
</tr>
<tr>
<td>fence 10</td>
<td>heating 5</td>
<td>puncture 4</td>
</tr>
<tr>
<td>gutter 5</td>
<td>car 19</td>
<td>clothes 11</td>
</tr>
<tr>
<td>ravages 4</td>
<td>alarm 5</td>
<td>net 8</td>
</tr>
<tr>
<td>hernia 4</td>
<td>bike 5</td>
<td>roof 8</td>
</tr>
<tr>
<td>car 23</td>
<td>problem 23</td>
<td>car 14</td>
</tr>
<tr>
<td>shoe 10</td>
<td>leak 3</td>
<td>way 20</td>
</tr>
<tr>
<td>leak 5</td>
<td>light 12</td>
<td>air-conditioning 2</td>
</tr>
<tr>
<td>building 17</td>
<td>boiler 3</td>
<td>damage 6</td>
</tr>
<tr>
<td>crack 6</td>
<td>roof 5</td>
<td>hole 5</td>
</tr>
<tr>
<td>wall 14</td>
<td>motorbike 2</td>
<td>bridge 4</td>
</tr>
<tr>
<td>fault 7</td>
<td>fault 4</td>
<td>heart 5</td>
</tr>
<tr>
<td>puncture 3</td>
<td>jeep 2</td>
<td>clock 3</td>
</tr>
<tr>
<td>pipe 7</td>
<td>door 11</td>
<td>chair 4</td>
</tr>
<tr>
<td>bridge 8</td>
<td>chain 4</td>
<td>wall 5</td>
</tr>
<tr>
<td>road 13</td>
<td>bulb 2</td>
<td>chain 3</td>
</tr>
</tbody>
</table>

Table 1: Direct object complements for the repair-verbs
The first observation from analyzing organic data associated with the selectional behavior of verbs like *fix*, *repair* and *mend* is that there are, in fact, two major selectional clusters, not one.

(26) \( \textit{fix.v} \)
    \textbf{object}
    a. \textsc{artifactual}: pipe, car, alarm, bike, roof, boiler, lock, engine; heart; light, door, bulb
    b. \textsc{negative state} (condition on the artifact): leak, drip
    c. \textsc{negative state} (general situation): problem, fault

(27) \( \textit{repair.v} \)
    \textbf{object}
    a. \textsc{artifactual}: roof, fence, gutter, car, shoe, fencing, building, wall, pipe, bridge, road; hernia, ligament
    b. \textsc{negative state} (condition on the artifact): damage, ravages, leak, crack, puncture, defect, fracture, pothole, injury
    c. \textsc{negative state} (general situation): rift, problem, fault

(28) \( \textit{mend.v} \)
    \textbf{object}
    a. \textsc{artifactual}: fence, shoe, clothes, roof, car, air-conditioning, bridge clock, chair, wall, stocking, chain, boat, road, pipe
    b. \textsc{artifactual} (extended or metaphorical uses): matter, situation
Modifying the Theoretical Assumptions

(29) a. GENERAL NEGATIVE SITUATION: “fix the problem”
   b. CONDITIONS OF THE ARTIFACT: “hole in the wall”, “dent in the car”.

What do these clusters have in common? Does the verb select for either a negative situation or an artifact? The answer is: basically, the verbs select for a negative state of an artifactual. When the negative relational state is realized, it can either take an artifactual as its object, or leave it implicitly assumed:

(30) a. repair the puncture / leak
   b. repair the puncture in the hose / leak in the faucet

When the artifactual is realized, the negative state is left implicit by default.

(31) a. repair the hose / faucet
   b. repair the (puncture in) the hose / (leak in) the faucet
Modifying the Theoretical Assumptions

(32)

\[
\text{repair}
\]

\[
\text{ARGSTR} = \begin{bmatrix}
\text{ARG1} = x: \text{human} \\
\text{ARG2} = y: \text{neg_state}(z) \\
\text{D-ARG1} = z: \text{phys} \otimes \text{Telic } \alpha
\end{bmatrix}
\]
## Case Study 2: Results

<table>
<thead>
<tr>
<th>believe.v</th>
<th>know.v</th>
<th>realize.v</th>
</tr>
</thead>
<tbody>
<tr>
<td>luck</td>
<td>73</td>
<td>33.14</td>
</tr>
<tr>
<td>ear</td>
<td>48</td>
<td>22.5</td>
</tr>
<tr>
<td>story</td>
<td>72</td>
<td>20.58</td>
</tr>
<tr>
<td>word</td>
<td>95</td>
<td>19.02</td>
</tr>
<tr>
<td>whereabouts</td>
<td>37</td>
<td>24.64</td>
</tr>
<tr>
<td>eye</td>
<td>74</td>
<td>15.19</td>
</tr>
<tr>
<td>detail</td>
<td>142</td>
<td>17.77</td>
</tr>
<tr>
<td>myth</td>
<td>6</td>
<td>14.17</td>
</tr>
<tr>
<td>truth</td>
<td>19</td>
<td>13.31</td>
</tr>
<tr>
<td>lie</td>
<td>10</td>
<td>12.63</td>
</tr>
<tr>
<td>tale</td>
<td>13</td>
<td>12.61</td>
</tr>
<tr>
<td>opposite</td>
<td>7</td>
<td>12.15</td>
</tr>
<tr>
<td>tarot</td>
<td>3</td>
<td>12.0</td>
</tr>
<tr>
<td>nonsense</td>
<td>7</td>
<td>11.6</td>
</tr>
<tr>
<td>propaganda</td>
<td>7</td>
<td>11.12</td>
</tr>
<tr>
<td>thing</td>
<td>47</td>
<td>9.12</td>
</tr>
<tr>
<td>woman</td>
<td>41</td>
<td>9.06</td>
</tr>
<tr>
<td>fortune</td>
<td>8</td>
<td>8.82</td>
</tr>
<tr>
<td>stupidity</td>
<td>3</td>
<td>8.57</td>
</tr>
<tr>
<td>rubbish</td>
<td>5</td>
<td>8.01</td>
</tr>
<tr>
<td>rumour</td>
<td>5</td>
<td>7.96</td>
</tr>
<tr>
<td>evidence</td>
<td>19</td>
<td>7.81</td>
</tr>
<tr>
<td>promise</td>
<td>7</td>
<td>7.78</td>
</tr>
<tr>
<td>figures</td>
<td>21</td>
<td>7.78</td>
</tr>
<tr>
<td>forecast</td>
<td>5</td>
<td>7.49</td>
</tr>
<tr>
<td>poll</td>
<td>7</td>
<td>7.48</td>
</tr>
<tr>
<td>gospel</td>
<td>4</td>
<td>7.45</td>
</tr>
<tr>
<td>assurance</td>
<td>6</td>
<td>7.44</td>
</tr>
<tr>
<td>success</td>
<td>14</td>
<td>7.35</td>
</tr>
</tbody>
</table>

Table 2: Direct object complements for the PROPOSITION/INFO-verbs
## Propositional Speech Act Selectors

<table>
<thead>
<tr>
<th>tell.v/direct object</th>
<th>tell.v/ditransitive obj2</th>
</tr>
</thead>
<tbody>
<tr>
<td>story</td>
<td>secret</td>
</tr>
<tr>
<td>truth</td>
<td>name</td>
</tr>
<tr>
<td>lie</td>
<td>detail</td>
</tr>
<tr>
<td>tale</td>
<td>reason</td>
</tr>
<tr>
<td>fib</td>
<td>gossip</td>
</tr>
<tr>
<td>joke</td>
<td>ordeal</td>
</tr>
<tr>
<td>untruth</td>
<td>gist</td>
</tr>
<tr>
<td>anecdote</td>
<td>fact</td>
</tr>
<tr>
<td>difference</td>
<td>whereabouts</td>
</tr>
<tr>
<td>parable</td>
<td>trouble</td>
</tr>
<tr>
<td>fortune</td>
<td>plan</td>
</tr>
<tr>
<td>news</td>
<td>date</td>
</tr>
<tr>
<td></td>
<td>destination</td>
</tr>
</tbody>
</table>

**Table:** Direct object and ditransitive obj2 complements for *tell*. 
In order to understand this behavior better, let us examine the non-coerced complementation patterns of these verbs in corpora. Several subclasses of clausal complements are attested in the BNC for each of these verbs. Namely, we identify the following three complement types:

(33) a. FACTIVE: know, realize
    b. PROPOSITION: believe, tell
    c. INDIRECT QUESTION: know, tell

(34) a. John realized [that he made a mistake].
    b. Mary knows [that she won].

The class of “Indirect questions” includes verbs selecting a *wh*-construction that looks like a question, but in fact denotes a value. For example, the verb *know* allows this construction, as does *tell*:

(35) a. Mary knows [what time it is].
    b. John knows [how old she is].
Factive Results

(37) **believe**(ARG1: human, ARG2: prop)

(38) a. **tell**(ARG1: human, ARG2: info)
    b. **tell**(ARG1: human, ARG2: Ind_Question)

(39) a. **know**(ARG1: human, ARG2: factive)
    b. **know**(ARG1: human, ARG2: Ind_Question)

(40) **realize**(ARG1: human, ARG2: factive)
(41)  *tell.v*

object

a. PROPOSITION: story, truth, lie, tale, joke, anecdote, parable, news, suspicion, secret, tale, details, gossip, fact, legend; dream, thoughts

b. INDIRECT QUESTION: name, whereabouts, destination, age, direction, answer, identity, reason, position, plan, symptoms; outcome, trouble

(42)  *know.v*

object

a. FACTIVE: truth, secret, details, story, meaning, fact, reason, outcome, saying

b. INDIRECT QUESTION: answer, score, whereabouts, address, username, password, name; feeling, difference

With the verb *realize*, the data show that NPs complements can also assume a factive interpretation:

(43)  John realized his mistake.
But what is interesting is that the majority of the nominals are abstract relational nouns, such as importance, significance, futility, and so forth, as illustrated below.

(44) \textit{realize.v}  
\text{object}  
\text{FACTIVE: importance, significance, extent, implication, futility, value, error, predicament}

For the verb believe, all nominals are coerced to an interpretation of a proposition, but through different strategies. Those nominals in (45a) either directly denote propositions (e.g., lie, nonsense) or are complex types that have an information component which can interpreted propositionally (e.g., bible, polls). The sources in (45b) are construed as denoting a proposition produced by (e.g., woman), or coming through (e.g., ear) the named source. Finally, the last set is licensed by negative polarity context, and is a state or event; e.g., ”He couldn’t believe his luck.”).
(45)  

\textit{believe.}\textit{v}  

\textbf{object}  

a. \textbf{Proposition:} lie, tale, nonsense, myth, opposite, truth, propaganda, gospel  
b. \textbf{Source:} woman, government, bible, polls, military; ear, eye  
c. \textbf{Event/State:} luck, stupidity, hype, success
Inherent tension between corpus data and theory
Polysemy is a linguistic phenomenon
Coercion is contextually modulated and licensed
Distributions of readings point to what is required of models for compositionality
Probabilistic judgments for Compositional Operations
References Cited

- Pustejovsky, James and Patrick Hanks. 2014. ”On Data and Methodology in Linguistics”, manuscript.
- Nicholas Asher and James Pustejovsky. 2006. ”A Type Composition Logic for Generative Lexicon.” *Journal of Cognitive Science*. 