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

The View from Beyond Building 20

Starved of adequate data, linguistics languished. . . . It became fashionable to look inwards to the mind rather than outwards to society.

Sinclair (1991)

Talk Outline

- 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 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

Selection in a Compositional Theory

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?

Verb Meaning

(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.

$$\overbrace{\lambda x_i}^{Args}\ \overbrace{[\Phi]}^{Body}$$

Decomposition Strategies

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

Argument Typing as Abstracting from the Predicate

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

$$\lambda x_2 \lambda x_1 [\Phi_1, \ldots \overbrace{\Phi_{x_1}}^{\tau}, \ldots \overbrace{\Phi_{x_2}}^{\sigma}, \ldots, \Phi_k]$$

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

σ and τ have now become reified as types on the arguments.

A Flexible Strategy of Selection

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.

A Probabilistic Strategy of Selection

Arguments can be viewed as encoding probability distributions of pretests for performing the action in the predicate.

Qualia Structure in GL

(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".

Type Composition Logic for GL (Asher and Pustejovsky, 2006)

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

Type Structures Pustejovsky (1995)

(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.

Natural Types

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*

Natural Predicate Types

Predicates formed with Natural Entities as arguments:

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

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

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

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

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

a. *beer*: $(liquid \otimes_a brew) \otimes_t drink$
b. *knife*: $(phys \otimes_a make) \otimes_t cut$
c. *house*: $(phys \otimes_a build) \otimes_t live\_in$

Artifactual Predicate Types

Predicates formed with Artifactual Entities as arguments:

1. *spoil*: $e_N \otimes_t \tau \rightarrow t$
2. *fix*: $e_N \otimes_t \tau \rightarrow (e_N \rightarrow t)$

a. $\lambda x : e_A[\mathit{spoil}(x)]$

b. $\lambda y : e_A \lambda x : e_N[\mathit{fix}(x,y)]$

- 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.
   b. John sold his book to Mary.
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, α, where $\sigma \sqcap \tau = \bot$:

1. [____]$_\sigma$ X
2. [____]$_\tau$ Y

are both well-formed predications, then α is a dot object (complex type).

Dot Objects 1/2

1. Act•Proposition: promise, allegation, lie
   - I doubt John's promise of marriage.
   - John's promise of marriage happened while we were in Prague.
2. Attribute•Value: temperature, weight, height, tension, strength
   - The temperature is rising.
   - The temperature is 23.

1. Event•Information: lecture, play, seminar, exam, quiz, test
   a. My lecture lasted an hour.
   b. Nobody understood my lecture.
2. Event•Music: sonata, symphony, song, performance, concert
   a. Mary couldn't hear the concert.
   b. The rain started during the concert.

Dot Objects 2/2

1. Event•Physical: lunch, breakfast, dinner, tea
   a. My lunch lasted too long today.
   b. I pack my lunch on Thursdays.
2. Information•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•(Information•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•Result: construction, depiction, imitation, portrayal, reference
   a. Linnaeus's classification of the species took 25 years.
   b. Linnaeus's classification contains 12,100 species.

Distinct Principles of Individuation in Dot Objects

1. a. John read every book in the library.
   b. John stole every book in the library.
2. a. Mary answered every question in the class.
   b. Mary repeated every question in the class.

Complex Predicate Types

Predicates formed with a Complex Entity Type as an argument:

1. *read*: $phys \bullet info \rightarrow (e_N \rightarrow t)$
2. Expressed as typed arguments in a λ-expression:
   $\lambda y : phys \bullet info\ \lambda x : e_N[read(x,y)]$
3. Mary read the book.

Modes of Composition in GL Type Theory

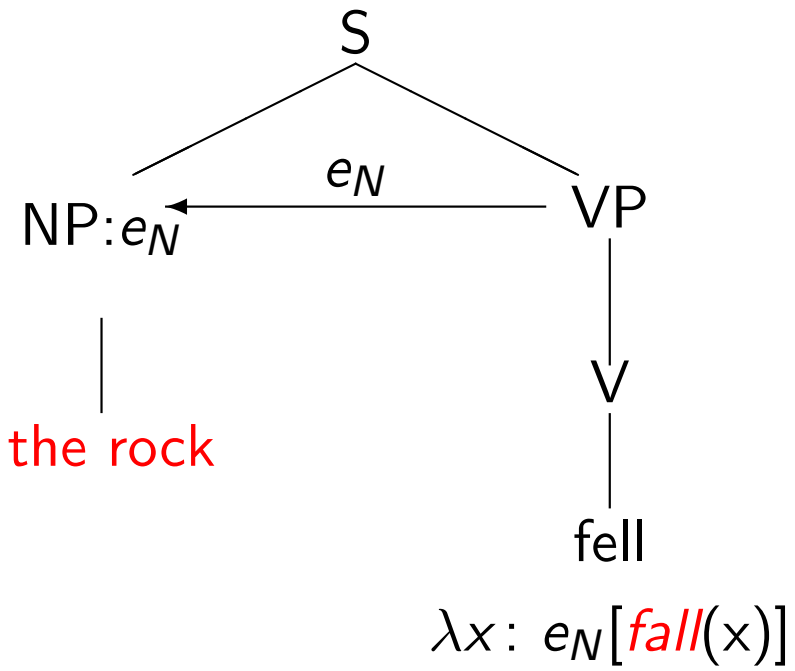
(3) a. PURE SELECTION (Type Matching): the type a function requires is directly satisfied by the argument;
b. ACCOMMODATION: the type a function requires is inherited by the argument;
c. TYPE COERCION: the type a function requires is imposed on the argument type. This is accomplished by either:
    i. *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).

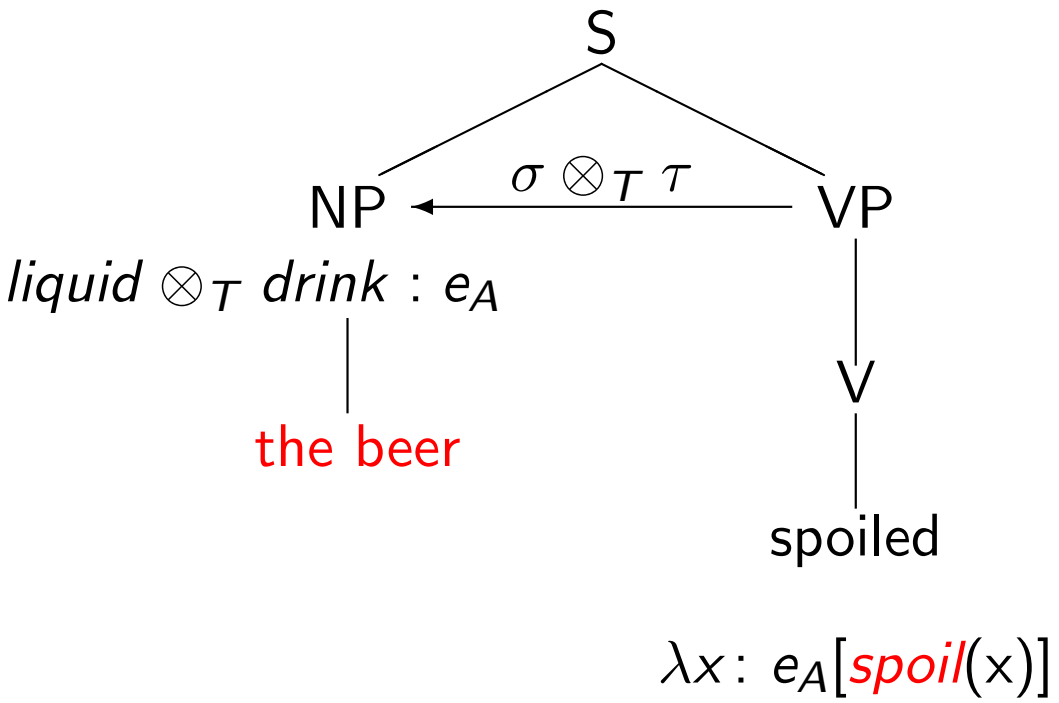
Natural Selection

1. The rock fell.



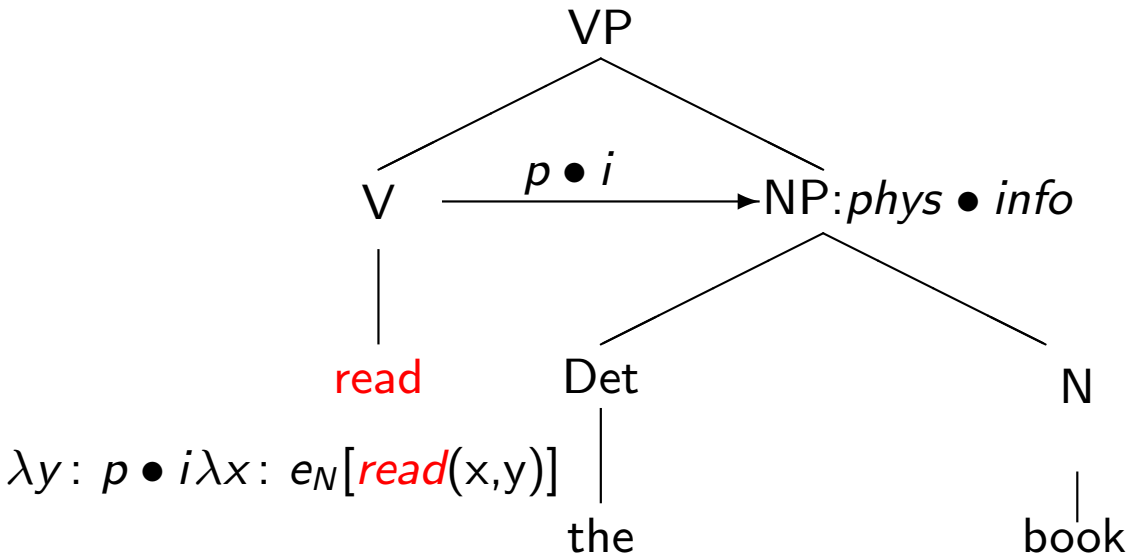
Pure Selection: Artifactual Type

1. The beer spoiled.



Pure Selection: Complex Type

1. John read the book.

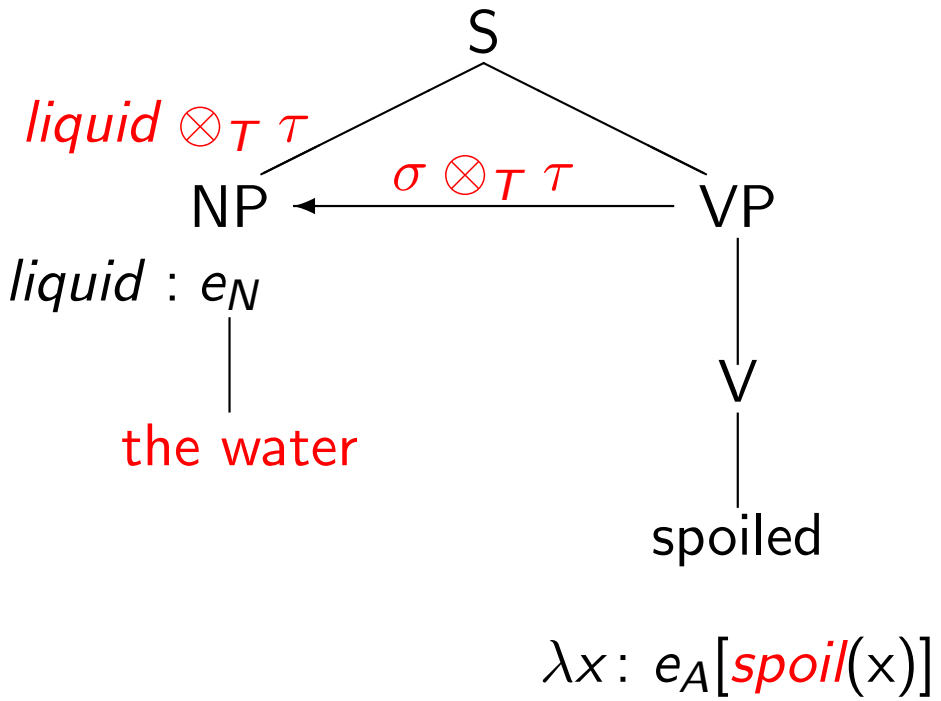


Coercion of Arguments

- The president denied the attack.
  EVENT → PROPOSITION
- The White House denied this statement.
  LOCATION → HUMAN
- This book explains the theory of relativity.
  PHYS • INFO → human
- d. The Boston office called with an update.
  EVENT → INFO

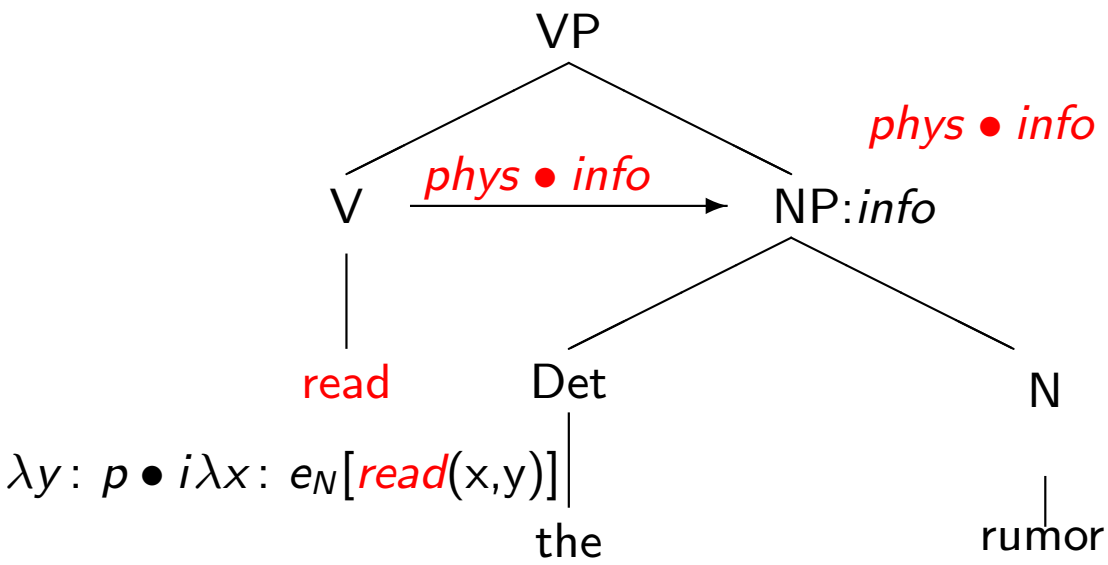
Type Coercion: Qualia-Introduction

1. The water spoiled.



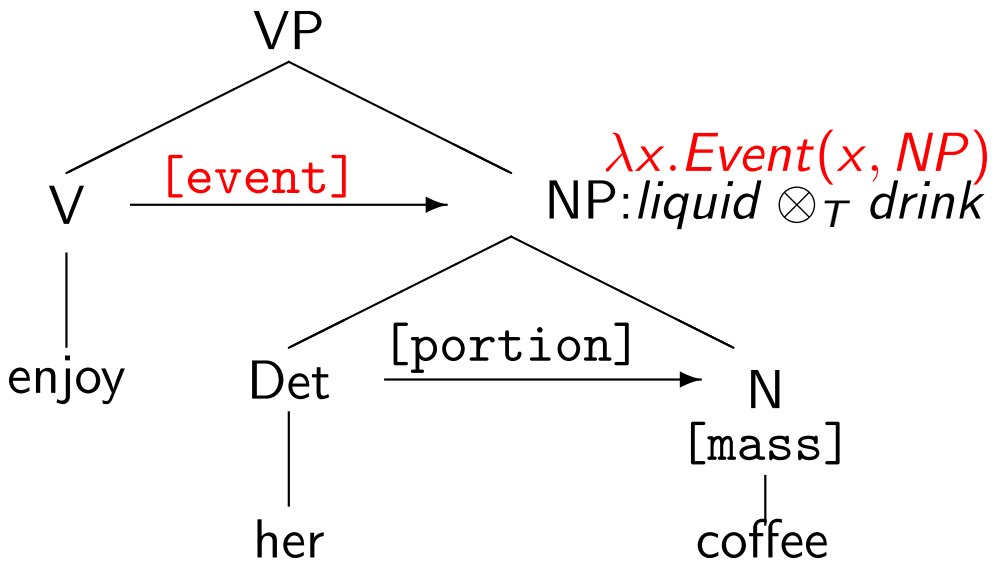
Type Coercion: Natural to Complex Introduction

John read the rumor.



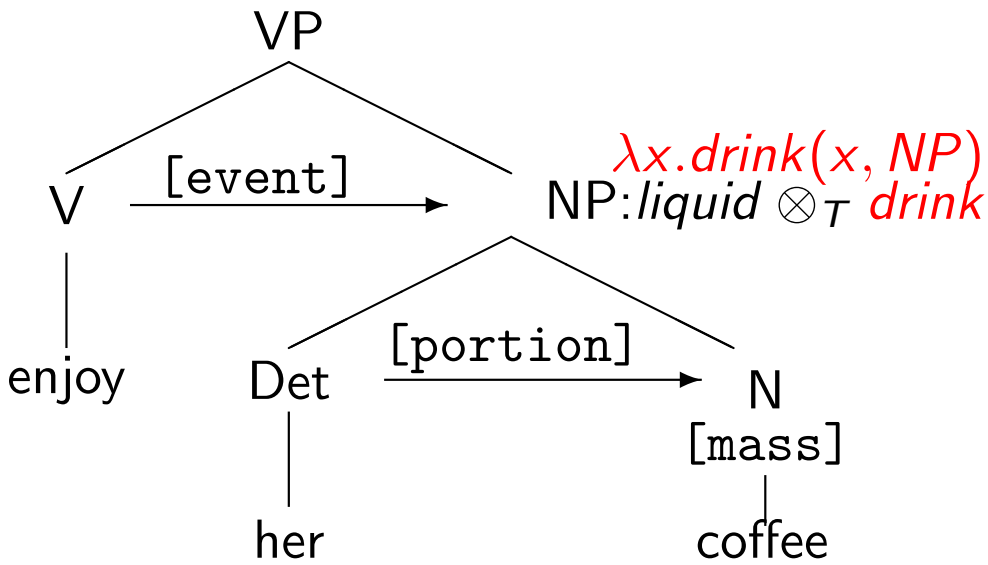
Type Coercion: Event Introduction

1. Mary enjoyed her coffee.



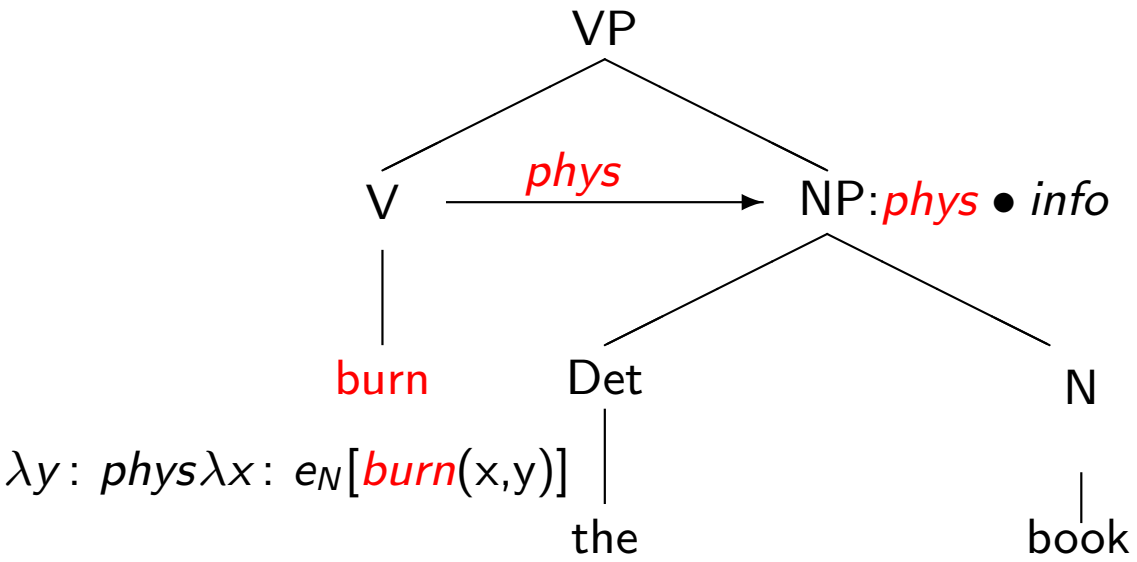
Type Coercion: Qualia Exploitation

1. Mary enjoyed her coffee.



Type Coercion: Dot Exploitation

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



Verb-Argument Composition Table

	Type Selected		
Argument Type	**Natural**	**Artifactual**	**Complex**
Natural	Sel/Acc	Tensor (Qualia) Intro	Dot Intro
Artifactual	Tensor Exploit Left (Acc)	Sel/Acc	Dot Intro
Complex	Dot Exploit	Dot Exploit	Sel/Acc

Corpus Distributions and Behavior Pustejovsky and Jezek (2008)

- 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?

Dot exploitation

(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 in on the shelf'
'The author will be discussing her new book'

Dot exploitation

(5) *house* (*phys* • *loc*)
Object

a. *phys*: built, buy, sell, rent, own, demolish, renovate, burn down, erect, destroy, paint, inherit, repair
b. *loc*: leave, enter, occupy, visit, inhabit, reach, approach, evacuate, inspect, abandon

'they built these houses onto the back of the park'
'the bus has passed him as he left the house'

Dot exploitation

(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'

Dot exploitation

(7) *exit* (*event* • *loc*)
Object

a. *event*: make, facilitate, follow, force, hasten, register
b. *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.

Dot Object Selection Asymmetries

As noted by Jezek and Lenci (2007) with respect to the object position of the complex type *phys* • *info* (i.e. *letter*, *article*, *book*, *novel* etc.): It. *articolo* 'article' combines more frequently with info-selectors rather than with phys-selectors:

(8) *articolo* (*phys* • *info*)
Object

a. *phys*: spostare 'move', ritagliare 'cut out'
b. *info*: approvare 'approve', bocciare 'reject', citare 'quote', correggere 'correct', ignorare 'ignore', commentare 'comment', conoscere 'know', condividere 'share'

'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', distruggere 'destroy', rubare 'steal', conservare 'keep', custodire 'keep', buttare 'throw away'

Dot Object Asymmetries

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'

Dot Object Asymmetries of Use

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) *finish* (Body: 'bring to an end'; Arg: *event*)
Object

a. *event*: journey, tour, treatment, survey, race, game, training, ironing, shopping
b. E-I, Q-E of $phys \otimes_{telic} \tau$: penicillin, sandwich, cigarette, cake, dessert, food
c. E-I, Q-E of $liquid \otimes_{telic} \tau$: drink, wine, beer, whisky, coke

'when they finished the wine, he stood up'
'just finish the penicillin first'

Strong Coercive Verbs

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 $phys \otimes_{telic} \tau$: 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 *event*s: 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'

Inducing Coerced Argument Types from Data

(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 $phys \otimes_{telic} \tau$: food
c. E-I-Q-I of $abstr \otimes_{telic} \tau$: tax
d. E-I-Q-I of $loc \otimes_{telic} \tau$: 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

Pustejovsky and Rumshisky (2008)

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

Case Study 1: Verbs Selecting for Artifactual Entities

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)
$$\left[\begin{array}{l} \textbf{touch} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x : \textit{phys} \\ \text{ARG2} = y : \textit{phys} \end{array}\right] \end{array}\right]$$

Artifactual Selection

An Artifactual predicate such as the verb *repair* would be typed as shown in (19).

(19)
$$\left[\begin{array}{l}\textbf{repair} \\ \text{ARGSTR} = \left[\begin{array}{l}\text{ARG1} = x : \textit{human} \\ \text{ARG2} = y : \textit{phys} \otimes_{\textit{Telic}} \alpha\end{array}\right]\end{array}\right]$$

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.

Case Study 1 Predictions

- Natural typed NPs should not appear as objects of artifactual predicates:
- Except under coercion interpretations

Case Study 2: Verbs Selecting for Propositions

Thesis: Coercion allows dot objects to appear in propositional argument positions.

(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].

(22)
$$\left[\begin{array}{l} \textbf{believe} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = x : \textit{human} \\ \text{ARG2} = y : \textit{info} \end{array}\right] \end{array}\right]$$

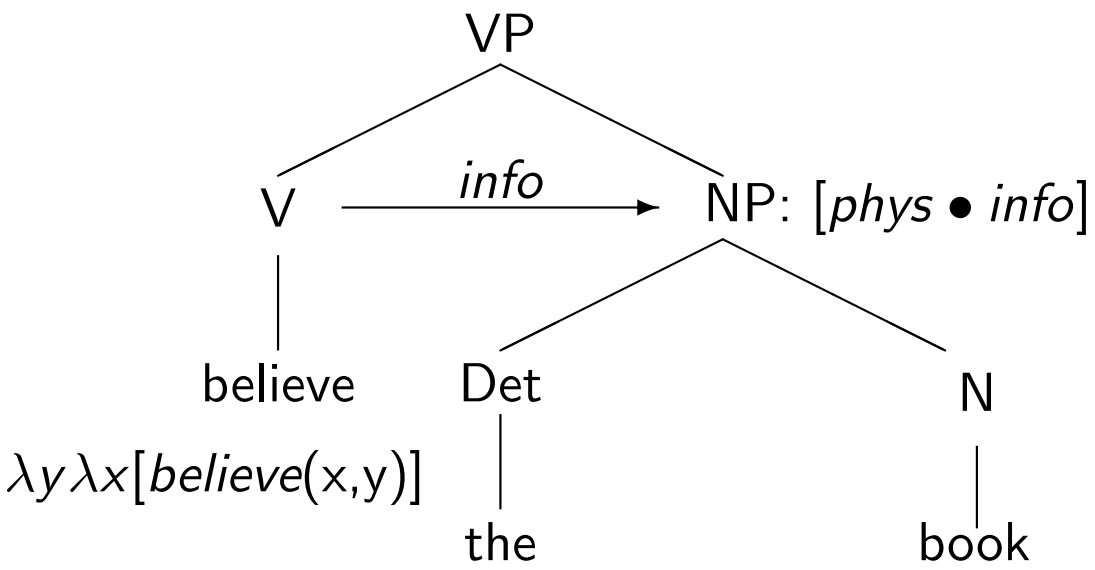
(23) a. Mary believed the book.
b. John told me a lie.
c. The man realized the truth.

Exploiting Complex Type Structure

(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

repair.v			fix.v			mend.v		
damage	107	42.66	pipe	9	11.83	fence	23	32.78
roof	16	20.27	gutter	4	11.45	shoe	10	19.01
fence	10	18.07	heating	5	9.66	puncture	4	18.91
gutter	5	15.87	car	19	9.43	clothes	11	18.68
ravages	4	15.76	alarm	5	9.13	net	8	18.01
hernia	4	15.61	bike	5	9.11	roof	8	16.99
car	23	15.39	problem	23	8.77	car	14	15.45
shoe	10	15.22	leak	3	8.58	way	20	14.26
leak	5	14.96	light	12	8.49	air-conditioning	2	12.71
building	17	14.02	boiler	3	7.96	damage	6	12.71
crack	6	13.99	roof	5	7.27	hole	5	11.38
wall	14	13.77	motorbike	2	7.19	bridge	4	9.68
fault	7	13.56	fault	4	6.91	heart	5	9.6
puncture	3	13.53	jeep	2	6.79	clock	3	9.45
pipe	7	12.89	door	11	6.65	chair	4	9.36
bridge	8	12.19	chain	4	5.48	wall	5	9.27
road	13	12.19	bulb	2	5.15	chain	3	8.3

Table 1: Direct object complements for the *repair*-verbs

Case Study 1 (Cont)

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) *fix.v*

object

a. ARTIFACTUAL: pipe, car, alarm, bike, roof, boiler, lock, engine; heart; light, door, bulb

b. NEGATIVE STATE (condition on the artifact): leak, drip

c. NEGATIVE STATE (general situation): problem, fault

(27) *repair.v*

object

a. ARTIFACTUAL: roof, fence, gutter, car, shoe, fencing, building, wall, pipe, bridge, road; hernia, ligament

b. NEGATIVE STATE (condition on the artifact): damage, ravages, leak, crack, puncture, defect, fracture, pothole, injury

c. NEGATIVE STATE (general situation): rift, problem, fault

(28) *mend.v*

object

a. ARTIFACTUAL: fence, shoe, clothes, roof, car, air-conditioning, bridge clock, chair, wall, stocking, chain, boat, road, pipe

b. ARTIFACTUAL (extended or metaphoric 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)
$$\begin{bmatrix} \textbf{repair} \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG1} = x : \textit{human} \\ \text{ARG2} = y : \textit{neg\_state}(z) \\ \text{D-ARG1} = z : \textit{phys} \otimes_{\textit{Telic}} \alpha \end{bmatrix} \end{bmatrix}$$

Case Study 2: Results

believe.v			know.v			realize.v		
luck	73	33.14	answer	389	35.17	mistake	15	20.02
ear	48	22.5	truth	219	30.92	extent	18	19.0
story	72	20.58	name	548	29.03	truth	15	18.7
word	95	19.02	whereabouts	37	24.64	importance	15	16.42
eye	74	15.19	secret	73	22.0	significance	11	16.11
hype	6	14.17	detail	142	17.77	implication	11	15.6
myth	12	14.07	story	141	17.48	futility	3	13.78
truth	19	13.31	meaning	78	16.58	value	17	13.28
lie	10	12.63	fact	159	16.28	danger	7	12.01
tale	13	12.61	reason	137	15.89	error	7	11.87
opposite	7	12.15	score	47	14.83	possibility	8	11.78
tarot	3	12.0	outcome	45	14.53	predicament	3	11.56
nonsense	7	11.6	saying	14	14.29	folly	3	10.09
propaganda	7	11.12	God	77	14.23	limitations	4	9.7
thing	47	9.12	username	7	14.02	strength	4	6.77
woman[7]	41	9.06	difference	105	13.98	need	6	6.07
fortune	8	8.82	feeling	79	13.75	threat	3	5.7
stupidity	3	8.57	word	162	13.74	benefit	4	5.31
rubbish	5	8.01	basics	10	13.53	problem	7	5.17
rumour	5	7.96	rules	99	13.03	advantage	3	5.04
evidence	19	7.81	address	42	12.74	difficulties	3	4.79
promise	7	7.78	password	10	12.4	effects	5	4.68
figures	21	7.78	identity	37	12.38	risk	3	4.68
forecast	5	7.49	joy	23	12.23	power	5	4.21
poll	7	7.48	trick	20	12.18	nature	3	3.7
gospel	4	7.45	place	171	11.88	fact	3	3.27
assurance	6	7.44	date	67	11.26	cost	3	2.94
success	14	7.35	extent	46	11.26			

Table 2: Direct object complements for the PROPOSITION/INFO-verbs

Propositional Speech Act Selectors

tell.v/direct object			tell.v/ditransitive obj2					
story	1286	52.0	secret	36	22.42	suspicion	4	5.62
truth	600	49.48	name	122	22.21	history	13	5.34
lie	254	45.67	detail	32	12.67	answer	9	5.33
tale	274	42.04	reason	37	11.06	direction	9	5.3
fib	18	30.84	gossip	6	10.4	dream	6	5.17
joke	94	28.85	ordeal	5	9.9	thought	10	5.08
untruth	8	19.08	gist	3	9.61	legend	3	4.92
anecdote	15	17.08	fact	34	9.5	age	13	4.7
difference	108	16.82	whereabouts	4	9.09	outcome	5	4.6
parable	8	12.75	trouble	9	6.98	symptom	4	4.32
fortune	24	12.57	plan	19	6.9	position	14	4.15
news	53	12.13	date	13	6.71	fate	3	4.08
			destination	4	6.54	identity	4	3.91

Table : Direct object and ditransitive obj2 complements for *tell*.

Case Study 2 (cont)

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*)

Results

(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.

Results

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) *realize.v*
object
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.").

Believe

(45) *believe.v*

object

a. PROPOSITION: lie, tale, nonsense, myth, opposite, truth, propaganda, gospel

b. SOURCE: woman, government, bible, polls, military; ear, eye

c. EVENT/STATE: luck, stupidity, hype, success

Closing Remarks

- 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.
- Pustejovsky, James and Elisabetta Jezek. 2008. "Semantic coercion in language: Beyond distributional analysis". *Italian Journal of Linguistics / Rivista di Linguistica*.
- Pustejovsky, James and Anna Rumshisky. 2008. "Between Chaos and Structure: Interpreting Lexical Data through a Theoretical Lens." *International Journal of Lexicography*.
- Nicholas Asher and James Pustejovsky. 2006. "A Type Composition Logic for Generative Lexicon." *Journal of Cognitive Science*.