

Why do we have the causal predicates we have?

(The delirium of interpretation of a causal fraud)

Jean-Pierre Koenig

(together with Breton Bienvenue, LianCheng Chief, Kathy Conklin,
Jeruen Dery, Hongmo Kang, Aron Marvel, Gail Mauner, Doug
Roland, Honggak Yun)

Causal predicates are everywhere

- English speakers have a lot of causal predicates in their mental lexicon (Koenig et al. 2003):
 - About 30% of English verbs have at least **one sense** that *entails* the presence of a participant playing a causal role in the event description (\approx 1,200 verbs, up to 3 times as many verb senses)
 - About 12% of English verbs have at least **one sense** that *entails* the presence of a participant playing an intermediary causal role in the event description (\approx 500 verbs) and 35% of English verbs allow an instrument (\approx 1,400 verbs)
- 1. The student dissected the cadaver with kitchen knives

Causal descriptions are what matters

- When talking of the “causal lexicon,” we are talking about speakers’ judgments on causal *descriptions*
1. The new shoes made Jill run faster/Jill run faster with her new shoes ≠ The new shoes made Jill run
 2. Roberto’s medication made him paint the landscape less realistically ≠ Roberto’s medication made him paint the landscape

Research questions

1. What semantic variation is there in our 4,000 or so causal predicates? We focus on 500 obligatory and 1,400 optional instrument verbs
2. How much information about causal predications is used in sentence processing (and sentence production)?
3. What does it tell us about our causal lexicon? Do naive speakers analyze causal predicates below the level of verb sense?

Dimensions of variation

- Causal descriptions involve a causal situation and a result situation with a possible intermediary causal situation

*Causal template 1: **cause (s1, s3) & P1(s1, a....) & P3(s3, p....)***

*Causal template 2: **cause (s1, s2) & P1(s1, a....) & cause (s2, s3) & P2(s2, i....) & P3(s3, p....)***

- What can causal predicates constrain?

- Participants in the various situations (*chambrier* 'let **wine** out so tha it is at room temperature, *limoger* 'remove a **high level government official** from her responsibilities')
- Categories of the situations: (-*tek*-, 'reluctant to go out (of harbour etc.) because of **the weather**')
- Presence of an intermediary causal relation (*mander* 'have somebody **tell** somebody to **do** something')
- **Nature of the causal relation (CAUSE, LET, HELP)**
- **Amount of scalar change** (cross-linguistic variation)

A survey of mediated causal relations

1. Two raters determined which English verbs *require* (*dissect*) or *allow* (*eat*) instruments
2. We analyzed each verb in terms of the second causal template (Koenig et al. 2008)

	Obligatory class	s_1	s_2	s_3	Causality	I
Cut	act	act	contact	incised	cause	
Whip₁	act		contact	contact	N/A	
Whip₂	act		contact	change-of-state	cause	
Whip₃	act		contact	compressed	cause	
Drug	act		in	change-of-state	cause	
Fill	act		in	change-of-configuration	cause	
Putt	act		forceful-contact	move	cause	
Fasten	act		connected	attached	cause	
Paint	act		on	change-of-state	cause	
Ski	act		pred ₂	move _{manner}	part-cause ⁺	
Cover₁	act			cover	N/A	
Cover₂	act		at	obscure	cause	
Garnish	act		at	more-ornate	cause	
Scoop	act		in	go-to	enable	

Optional class	s_1	s_2	s_3	Causality	I
Charm/break	act		pred ₁	cause ⁺	
Convert	act		pred ₁	cause ⁺	machine
Eat	pred ₂		pred ₁	help ⁺	
Enter	act		change-of-state	precondition ⁺	
Attach	act	connected	attached	cause	
Trap	act		not(able(move))	cause ⁺	
Administer	act		in	cause ⁺	medicine
Heat	act	contact	cooked/hot/cold	cause	hot
Liberate	act	act	change-of-state	cause	human
Move_{tr}	act	forceful-contact	move	cause	
Commute	act	pred ₂	move _{manner}	part-cause ⁺	
Document	act	mark	represent	cause	

Table 2 Semantic category of s_1 , s_2 and s_3 , the kind of causality involved and the semantic category of **I** for optional instrument classes

Linguistic results of survey

1. More than one cause relation can occur in the meaning of mono-morphemic verbs, but only in limited cases: use of tools;
2. Idiosyncratic information specifies more instrument activity and change of state in patient than agent activity (a possible example of goal bias and lexical reification of discourse distribution)
3. **Intermediary causes involve a wide variety of causal relations**
4. Obligatory instrument verbs constrain more instrument properties than optional verbs do (cf. *behead* vs. *kill*);

A causal menagerie

1. Mary cut her steak *with a butcher knife* (= cause)
2. John watches birds all day *with his binoculars* (= help)
3. Bill cooks his steaks *with butter* (= help)
4. Bill entered Joan's room *with a duplicate key* (= precondition)
5. Joe scooped the ice-cream *with a wooden spoon* (= enable)
6. Connie skied down the slope *with her new skis* (= part cause)

Defining the menagerie

- s2 can be the true cause of the final change of state s3: *cut*
- s2 can be the cause of a precondition of the change of state s3: *open*
- s2 can be one of a joint set of causes of the change of state s3: *ski*
- s2 can enable a change of location s3: *scoop*
- s2 can cause the event/action to lead to a better resulting state or to be performed: *cook with butter*

An eventuality *e1 helps* the occurrence of token *e2* of the event category *C* iff (i) there is an ordering of tokens of *C* along a pragmatically defined scale (ease of performance, how good the resulting state is, fewer unwelcome 'side effects'); (ii) *e1* caused the token *e2* of *C* to be higher on that ordering than it would otherwise have been.

Is that enough: *use* ≠ *with*?

- We focused on *volitional causal instruments* by using the *with* and *use* tests conjunctively
- What if we had distinguished between *use* and *with* (Riessman and Rawlins 2017)?
 1. (where Chloe is wearing scuba gear in order to breathe underwater)
 - A. Chloe used scuba gear to clean the floor of the tank
 - B. #Chloe cleaned the floor of the tank with scuba gear

Resources vs. Intermediary Causes

- *Use* requires volitionality of the Agent, *with* requires a more “core” causal role of the Instrument
 1. Sam polished the floor with his pants
 2. #Sam used his pants to polish the floor
 - But we need to somehow explain the role the Instrument of *use* plays
- 3. #John used the ladder to paint the ceiling (if the ladder is not used to help with the painting) (John acted on the ladder and had the intention of paint the ceiling)
 - Two distinct, semantically related notions of instrumentality
 - *Use*: Instrument = Resource (weaker causal role) for goal (volitionality of Agent)
 - *With*: Instrument = Intermediary cause (stronger causal role), no goal required

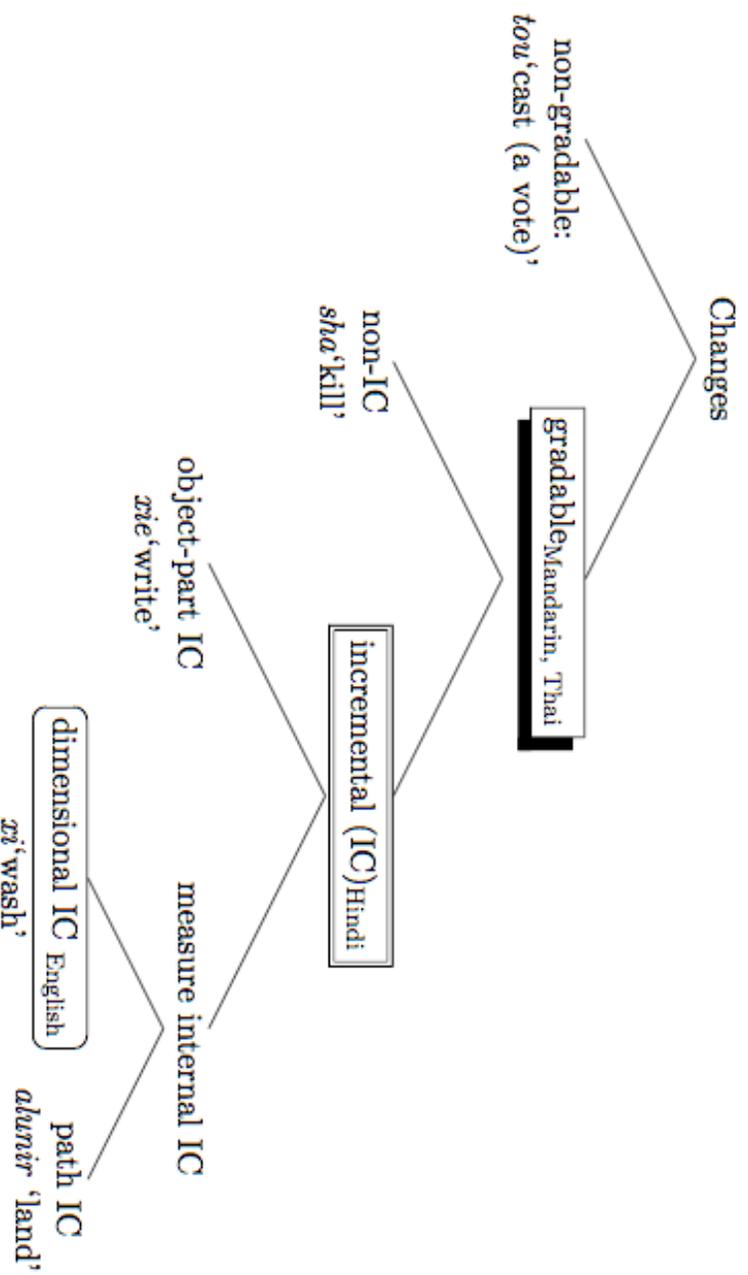
Variation in what is caused: Non-
culminating accomplishments (Koenig
and Chief 2008)

1. wo chi le liang chuan dakao, dan mei chi-wan
eat PERF two CL kabob, but not eat-finish
'I ate two kabobs, but didn't finish eating. (lit.)'
2. Xu Mei he Sun Mazi ba Lao Lo sha le mei sha-s
Xu Mei and Sun Mazi BA Lao Lo kill le not kill-die
'Xu Mei and Sun Mazi killed Lao Lo but didn't make
him die. (lit.)'

A scalar model for non-culminating accomplishments

Hypothesis (NC scalar hypothesis). NC arises when the main verb entails that a normative scalar change occurred with degree $d_0 < d \leq d_N$ rather $d = d_N$

Cross-linguistic variation in event realization



A semanticist view of semantics vs. a psycholinguist view of semantics

- Lots of information associated with verbs is not included in our semantic skeletons
- In talking of causal predicates, we have assumed:
 - A. “verb sense” is the right kind of unit when discussing word meaning
 - B. Thematic role is the right kind of level of analysis when talking about causes and instruments included in verb meaning
- Lexical information is richer and more specific than traditional lexical semantics templates (Elman 2009);
 - *cut the grass vs. cut the cake*
- Our parameters of lexical variation in causal predicates are a small portion of the semantic information associated with verbs and it is this richer that is used in processing

Language processing ain't lexical semantics

TABLE 1
Role Features and Production Frequencies for FRIGHTEN

<i>Agent Feature</i>	<i>Production Frequency</i>	<i>Patient Feature</i>	<i>Production Frequency</i>
is mean	10	is scared	10
is scary	10	is small	7
is ugly	8	is weak	7
is big	7	is helpless	4
is sadistic	6	is jumpy	4
has problems	4	is nervous	4
is insensitive	4	is not knowledgeable	4
is heartless	3	is insecure	3
is unfriendly	3	is in trouble	3
		is shuddering	3

- When asking to list features of causes, participants can list features semanticists might think are irrelevant (McRae et al. 1998)

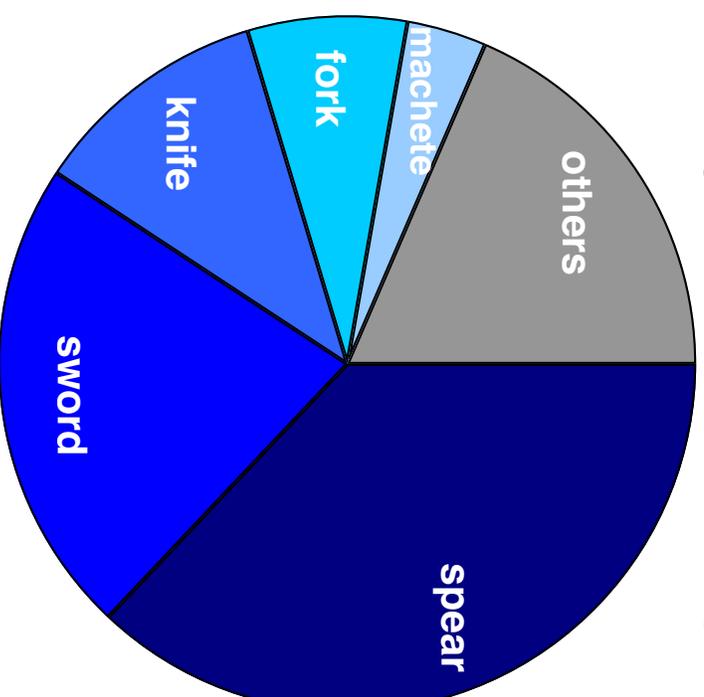
Semantically irrelevant information matters in processing

- The distribution in semantic space of filler-category features has an effect on sentence processing
- **The shared semantic similarity hypothesis:** The more similar the likely instruments of the event described by the verb, the easier it is to process an instrument-denoting NP (Roland et al. 2012)

The aboriginal man *jabbed/attacked* the angry lion with a *spear*

Filler-category predictability affects processing

The aboriginal man **jabbed** the angry lion with a/n _____.

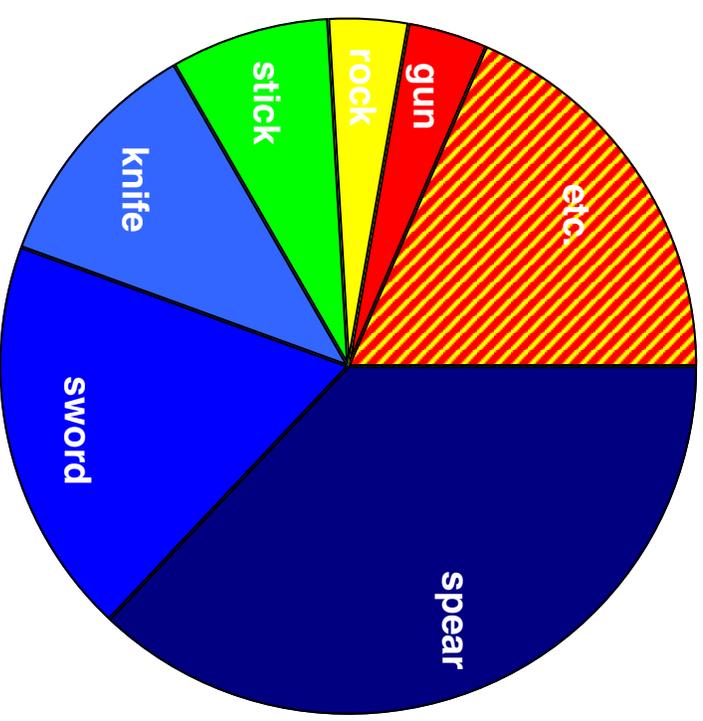
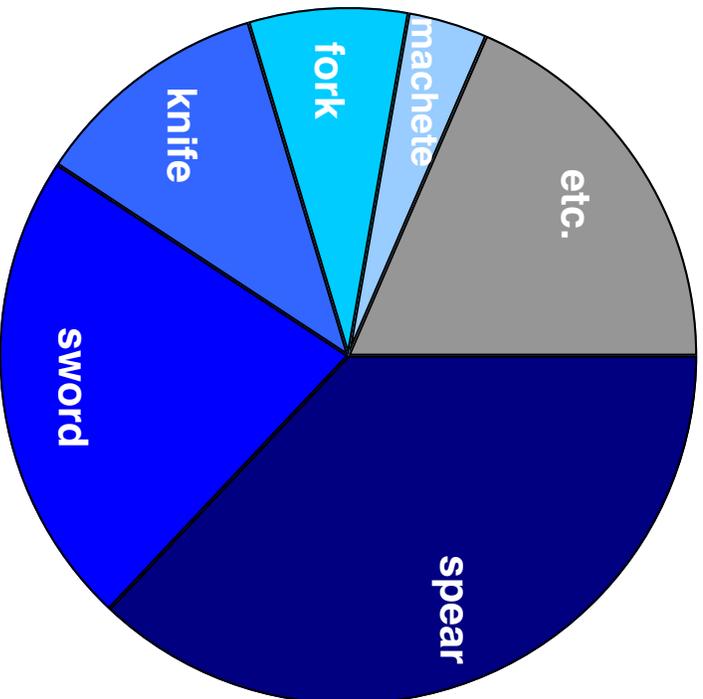


Spear is more likely than *fork*, so it will be easier to process (e.g., Altmann & Kamide, 1999; 2003; McRae et al., 1997; Speer & Clifton, 1999).

Imagine a second verb has slightly different expectations...

attacked

The aboriginal man ~~jabbed~~ the angry lion with a/n _____.



Filler predictability *and* semantic similarity affect processing

Question 1:

If *spear* is equally likely for both verbs, is it easier to process after *jab*, due to the greater degree of shared similarities between the possible instruments?

YES

Question 2:

Is *machete* easier to process after *jab* than *rock* is after *attack*, even though they are equally likely, due to the greater degree of shared similarities between *machete* and the other possible instruments of *jab*?

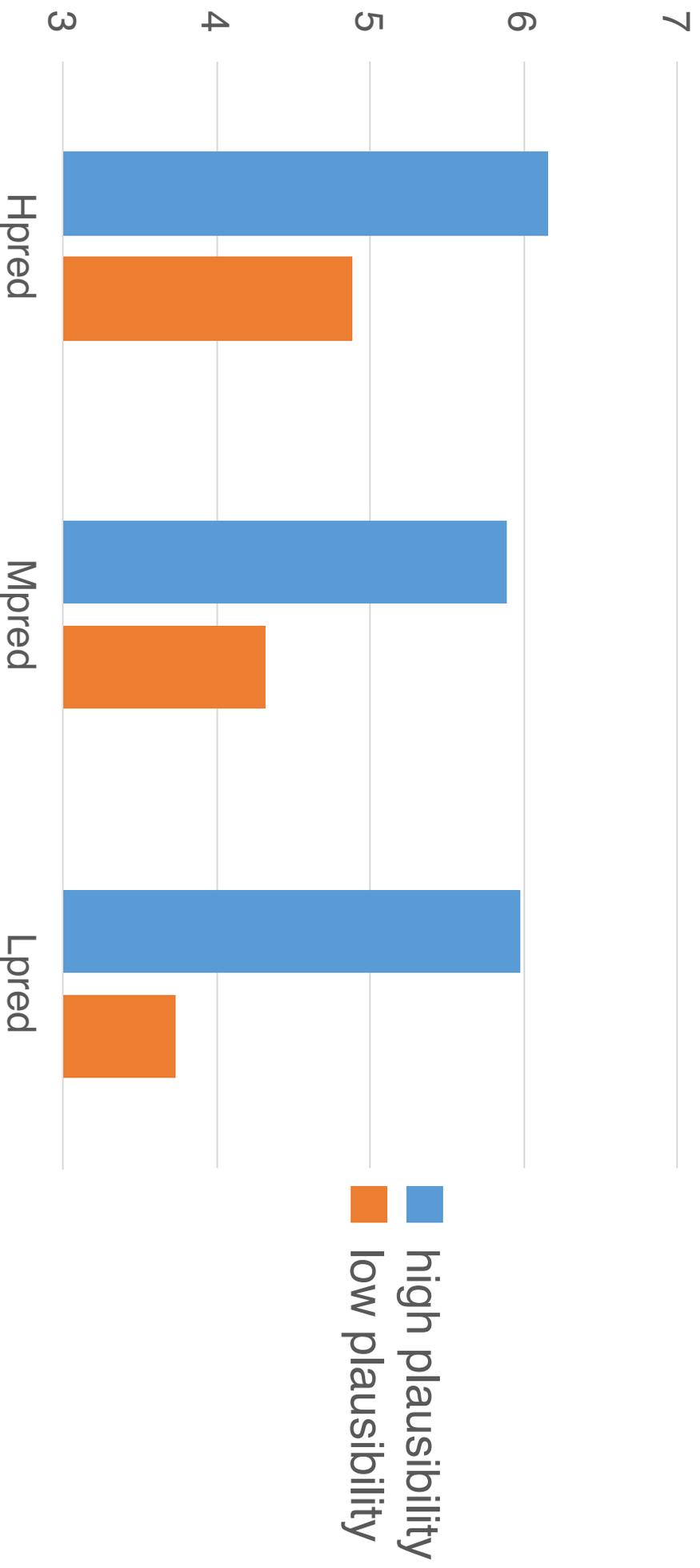
YES

Predictability, similarity ... and plausibility

- Processing is sensitive not just to predictability or similarity of fillers, but also their plausibility (given what precedes)
- Kang (forthcoming) carefully distinguishes predictability (estimated via weighted sum of mentions in a completion task) and plausibility (in a rating study)

Condition	Example	Predictability	Plausibility
High predictability High plausibility	The architect showed the sponsor his design	0.0447	6.8
High predictability Low plausibility	The architect showed the sponsor the paper	0.0139	5.5
Mid predictability High plausibility	The architect showed the sponsor his portfolio	0.0046	6.5
Mid predictability Low plausibility	The architect showed the sponsor wood	0.0048	4.85
Low predictability High plausibility	The architect showed the sponsor a replica	0.0006	6.3
Low predictability Low plausibility	The architect showed the sponsor sapphires	0.0006	3.85

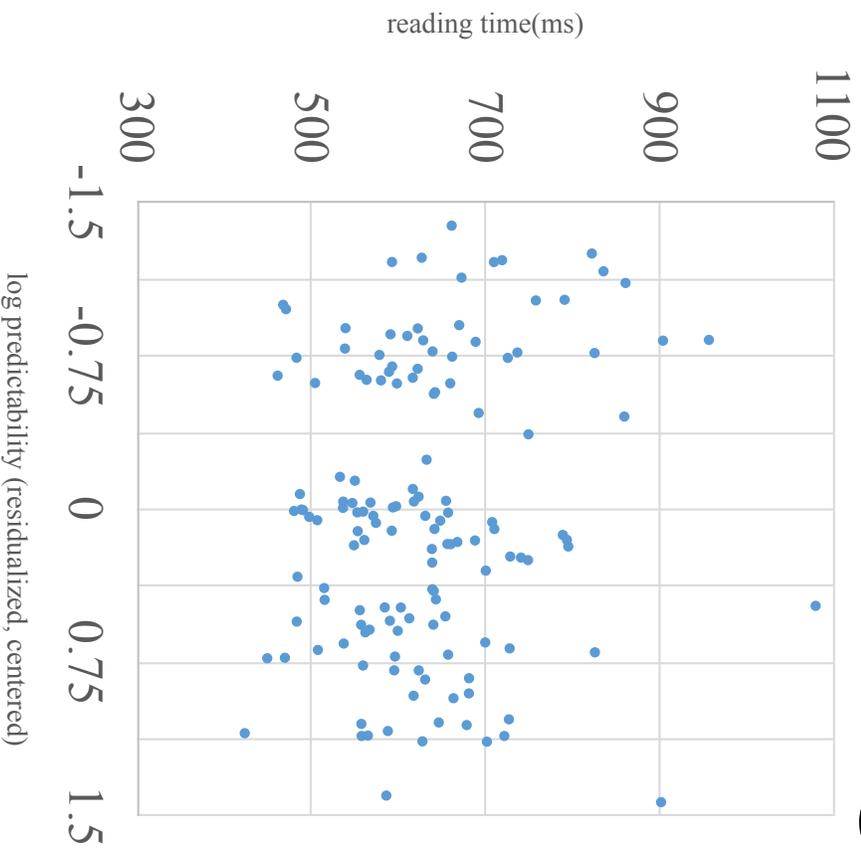
An interaction between plausibility and predictability in off-line measures



Reading time study

- Region-by-region self paced reading task
 - The architect /showed /the sponsor /his portfolio / yesterday /afternoon./

An effect of predictability in target region



Significant effect of predictability:

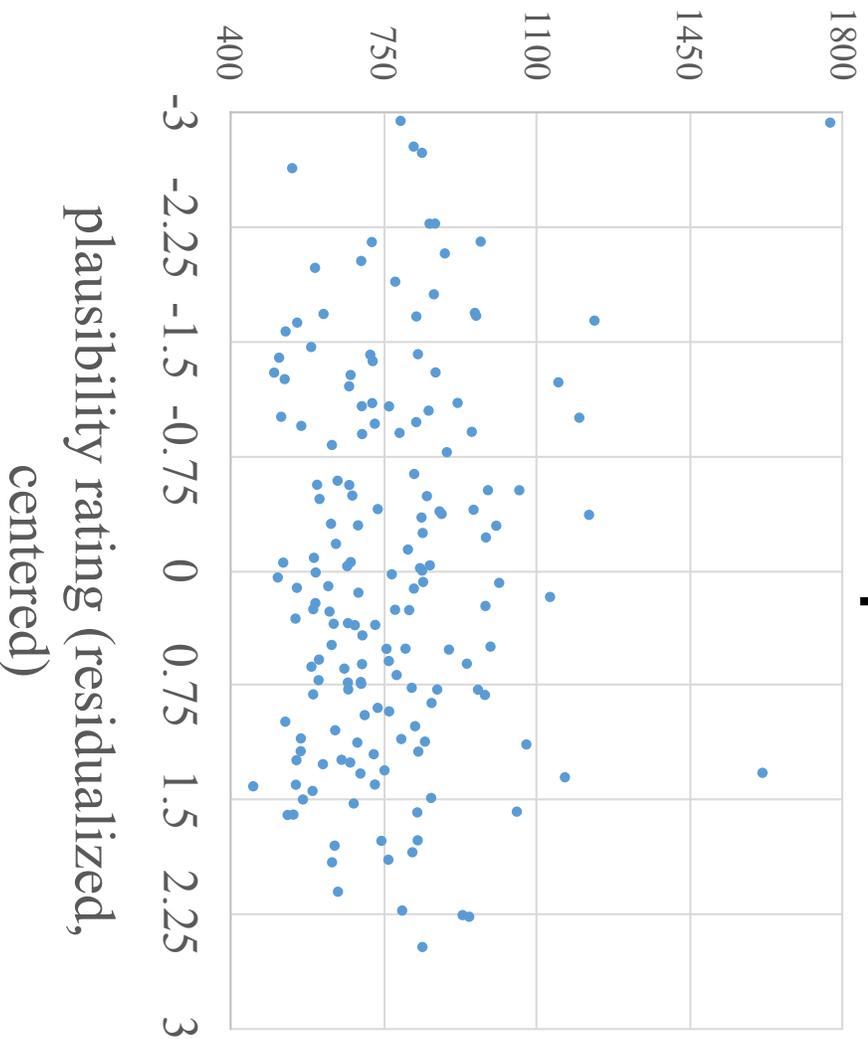
coef=-17.71, SE: 7.32, $t=-2.42$

No effect of plausibility:

coef=-4.83, SE=6.17, $t=-0.78$

Condition	Reading time(ms)
High predictability(>0.5)	617.2388
Medium predictability (-0.5~0.5)	625.0121
Low predictability(<0.5)	651.2404

An effect of plausibility in spillover region



Significant effect of plausibility:
coef=-14.78, SE=4.82, t=-3.07

Condition	Reading time(ms)
High plausibility(>0)	745.00
Low plausibility(<0)	791.10

Event complexity matters to discourse processing

- How complex an event is affects discourse expectations about time interval between events (*contra* a narrow interpretation of simulation semantics) (Dery and Koenig 2015)
1. Mary poured water in a glass. After a few seconds_{short}/ weeks_{long}, she drank it.
 2. The hospital collected DNA sample from AIDS patients. Several minutes_{short}/months_{long} later, they tested them and analyzed the data.

Predictability, plausibility, filler similarity, event complexity rarely matters to grammar

- Lexical meaning, according to semanticists, abstracts away from all of this specific information ...
- So does grammar, for the most part:
 - Although it does affect a little direct object omission (*eat/ eat something*; Resnick 1993, Heider 2015)
 - And modification rate of expressed instruments (less constraining a verb is w.r.t. its instruments the more likely they are to be modified) (Heider 2015)
- 1. The court in Belfast ruled that Christie was more responsible for her actions when she killed Penny McAllister *with a sharpened butcher's knife* than was originally thought.

Constructing a sentence/representation ≠ categorizing a situation

- The difference between the kind of information relevant for lexical semantics of causal predicates and production/comprehension of causal predicates is not surprising
 - Predictability/Plausibility/Semantic similarity are ways of helping **construct** a sentence/representation
 - Lexical meanings correspond to ways of **classifying** events
- Is a richer semantic information part of our classification of causal events descriptions (is there life below lexicographers' verb senses)?

Rethinking our notion of verb sense

- When we discuss the causal lexicon, we assume an unreconstructed view of lexical semantics in which determining verb senses can be operationalized
- But there is no such operationalization
- In fact, there is no theoretically viable notion of verb sense, it's more and more specific categories all the way down (and that is OK)

Non-lexical semantic similarities

- Speakers can classify events based on similarities below the lexicographer's notion of verb sense (Marvel and Koenig 2015)
1. The senator raised a glass in celebration.
 2. The crane raised the car out of the water.

Six degrees of similarity beyond verb senses

1. Event complexity (individual selling vs. corporation selling)
2. Time scale (buying a bank vs. buying two cheese rolls)
3. Agent/Patient type (a fleet rescuing the city vs. archaeologists rescuing information)
4. Sociocultural salience (raising a glass vs. raising a car)
5. Inferences (covering bruises vs. covering your eyes)
6. Distinct motion sequences (pulling the ball vs. pulling the barricade)

These six parameters allow raters to categorize event categories below verb senses

Verb	AHD senses	Categories
<i>bake</i>	2	10
<i>borrow</i>	2	18
<i>buy</i>	3	18
<i>cover</i>	8	30
<i>deliver</i>	7	17
<i>frighten</i>	2	14
<i>immerse</i>	3	8
<i>pull</i>	6	24
<i>rescue</i>	1	13
<i>sell</i>	4	13
Average	3.8	16.5

Table 1. Comparison of AHD senses to event categories discovered by application of the parameters discussed in Section 3.

Participant similarities can serve as cues to event similarities

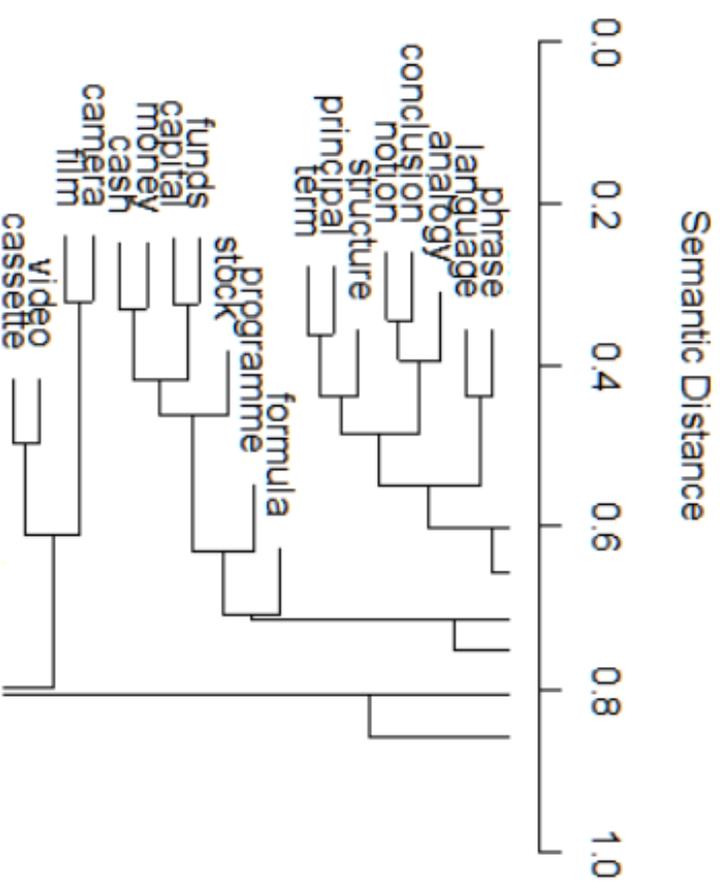


Figure 1. A section of the dendrogram created by using LSA semantic distance values to group direct objects of the verb *borrow*.

Participant similarities match better raters' categories than random categorization

List	P LSA	r LSA	F LSA	F rand	Ratio
Subj	40%	80%	.53	.39	1.38
DO	35%	66%	.46	.32	1.46
Overall	38%	73%	.50	.35	1.42

Table 2. *F*-scores for LSA categories, compared to *F*-scores for randomized categories. Ratios represent how much better than chance LSA categorization performed.

The LSA automated categorization resulted in an average of 42% more accurate categorization than that obtained by random categorization.

- “Subject” properties matter more than “object properties” except for *frigten*

Do naive speakers agree with raters categorization?

- 96 (edited BNC) sentence pairs for 8 verbs balanced across 3 groups (same sense-same category; different sense-different category, same sense-different category)
- Participants were asked to judge the similarity of described situations on a 1-7 scale
- Judgments binarized into “above participant’s median score” and “below participant’s median score”
- Scores agreed with categorization if above (below) median and in the same (different) rater category

Yes, they do

- 78% of judgments were below the participant's median when the two sentences were in different rater categories
- There was a significant relationship between similarity judgments and category assignment with a medium to large effect size ($X^2=218.64$, $N=1129$, $p<.001$, $V=.44$)

A more refined experiment is under way

- 24 verbs over a large section of semantic space

verbs of feeling: love, hate, fear, enjoy

verbs of physical action: dance, play, work, push

verbs of perception/mental attitude: hear, watch,
value, want

verbs of movement: send, follow, place, cross

verbs of possession/transfer of possession: sell,
receive, own, keep

verbs of change of state: clean, damage, improve,
ruin

Extracting similarity features

- 20 curated ps-random sentences from BNC for each verb
- Participants:
 1. sort sentences on the basis of the perceived similarity of described events into as many group as possible
 2. Briefly indicate what made them group the sentences together

A partial corroboration of raters' dimensions of similarity

- 20 curated BNC sentences per verb
- Sorting task + justification of grouping task
- Half of the data: agent/patient type matters (group, animacy); temporal size/complexity matters
- We need to automate grouping of features
- Further experiments to establish relative importance of features in sorting

Lexical semantics is an abstraction over richer meanings

- The meaning associated with verbs is much richer than the abstract templates we started with and involves event knowledge that is more specific than our decomposed semantic representations assume
- That richer meaning matters to what we use language for: produce and comprehend sentence and discourses
- Categorization of events associated with verbs goes below the verb sense (is sensitive to at least some of this information) and there is no reason to believe it's not tortoise all the way down

Why do we group micro-senses the way we do?

- Your average undergraduate speaker of English knows 1,200 distinct causal stems not 12,000 or 120,000. Why don't we have many more names of event categories?
- We group events in categories in many ways, but we group these micro-categories within a single stem in only restricted ways when it comes to *distinct names* of those categories
 - Many different categories of *selling* but still all instances of a larger category of selling events with a single name for this overarching category
- Some “salient” states and processes can serve as primitives of event categorization
- Opposition structure and focus on change leads to additional predicates (Pustejovsky)
- Causal predicates are those whose processes that lead to change are agentive

Further limitations

- Quasi-manner/result complementarity further reduces the range of predicates
- Limits on lexical resources: A trade-off between unique names for distinct categories and an unwieldy lexicon (for both learning and retrieval)
 1. pulling the ball vs. pulling the barricade
- This won't give you instrument verbs!