The geometry of the verbal semantic space

Jean-Pierre Koenig
University at Buffalo

‘This report is a record of issues in the semantics of natural languages thathave concerned me in the past few years, some of the things I have had to say about them, and someof the things others have had to say about them. There is nothing new in these pages, and there ismuch that is borrowed. I use numbered paragraphs mostly to create favorable associations-- but alsoto make it obvious that I do not expect the reader to perceive any structure beyond that of sheersequence.’(Charles Fillmore, 1972).

Collaborators

- Breton Bienvenue
- Kathy Conklin
- Anthony Davis
- Gail Mauner
- Eun-Kyung Yi

The verbal semantic space

- English native speakers know about 4,000 verbs;
- The English verbal polysemy factor is between 3 (Cobuild dictionary) and 4 (Wordnet);
- What kinds of meanings do these 12,000 to 16,000 verb meanings express?
  - What constraints on possible verb meanings are there?
  - How can we relate verb meanings?

The organization of the verbal semantic space

- Vendler and Dowty have argued that verb meanings/event-types fall into a few categories:
  - States (*believe*) = *DEAD(y)*
  - Processes *(work)* = *WORK(x)*
  - (internal) Changes (*blossom, die*) = BECOME(*DEAD(y)*)
  - Induced Changes (*kill*) = CAUSE(*x,* BECOME(*DEAD(y)))*
- These categories:
  - define various quadrants of semantic space;
  - Lexical decomposition can be used to represent relations between these meanings;

Where do idiosyncrasies of meaning reside?

- Verbs can be quite idiosyncratic in the constraints they impose on base predicates and participants;

- … but the *structure* of word meanings is built from the same parts;

- *Chambrer*: 'to keep a bottle of wine in a warm room so that it can slowly be brought to room temperature'
  - CAUSE (X, BECOME(*AT-ROOM-TEMPERATURE("wine")))*

A “generative” view of lexical decomposition

- Sometimes lexical decomposition is interpreted as if:
  - It is a generative device:
    - A set of base predicates (DEAD, BROKEN…)+ a limited set of operators (CAUSE, BECOME, NOT) ⇒ set of possible meanings;
  - It exhausts the linguistically relevant ways in which to categorize types of verbal meanings;
- **Goal:** Describe some ways in which the organization of the verbal semantic space is richer than suggested by that view;

Major claims

1. There are systematic cross-linguistic differences in "molecular" lexical meanings (not just base predicate or operators);
2. Lexical semantics is not recursive (CAUSE(CAUSE…)), but it can be more complex than previously thought;
3. The operator CAUSE is actually a *family* of *meta-intensional* operators;
4. There is a difference between semantic internal verbal semantic space and syntax-relevant verbal semantic space.
5. Idiosyncratic aspects of verb meaning is result-biased

1. The Incompleteness Effect

- In some East Asian (Thai, Mandarin) and possibly some South Asian (Hindi, Tamil) languages, many sentences *implicate* a (definite) change of state (kill → dead) when their English translation *entails* it = Incompleteness Effect
- If you think lexical meanings can vary only in their base predicates and constraints on their arguments (cf. *chambrer*), one has a hard time modeling the Incompleteness effect;

Examples

- Xu Mei he Sun Mazi ba Lao Lo sha le mei sha-si (attested)

‘Xu Mei and Sun Mazi killed Lao Lo but didn't make him die. (lit.)‘

- wo (…) gai le xin fangzi, fangzi hai mei gai-wan

‘I build a new house, but it is not finished.'

- Tuoersitai-de Zhanzheng yu Heping wo bu xihuan, du le ji ci dou mei du-wan (attested)

‘I don't like Tolstoy's War and Peace, I read it several times, but never finished reading it.'

Incomplete and Complete stems

- *Incomplete stems:* jiǎn ‘to cut with scissors’, xiū ‘to repair’, quàn ‘to persuade’, shā ‘to kill’, guān ‘to close’, niàn ‘to read’, chī ‘to eat‘, ‘hōng ‘to dry (clothes)’, xǐ ‘to wash', zhǔ ‘to cook', xiě ‘write’, bèi ‘to recite (memorize’', gài ‘to build‘, zhì ‘to cure’, …
- *Complete stems:* zhuǎn ‘to turn’, zhà ‘to deep fry’, yān ‘to pickle’, fù ‘to pay’…

The Source of the Incompleteness Effect

- The source of the effect lies in the meaning of incomplete stems, not in the meaning of Mandarin or Thai proto-patient DPs or viewpoint aspect (see Koenig and Chief, 2008);
- **Hypothesis:** Incomplete stems denote *induced* changes of state whose result state is a property that can be *gradable* (a matter of degree);

Examples

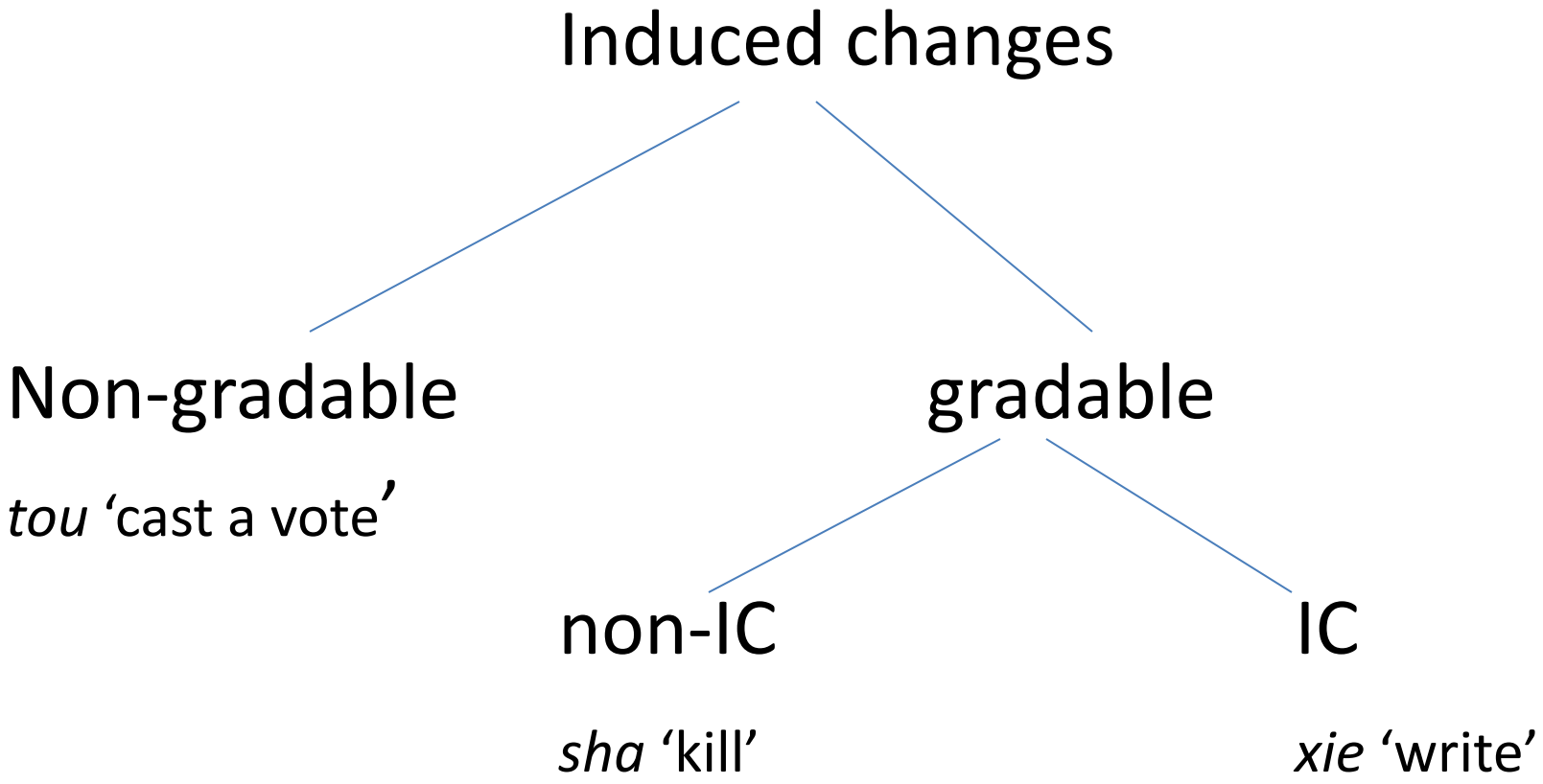
- Killing involves seriously lowering the degree of somebody's health to degree -10, if we follow Schank's (1973) Conceptual Dependency);
- Cooking involves changing more or less the chemical structure of vegetables or meat so that it conforms to some cultural norm *and* the more you cook the food, the less raw/more cooked it is (Hay et al.; Kennedy and McNally; Beavers; …)

How can *shā ‘ kill’* be gradable?

- Killing involves changing a gradable property (health; vital forces…) in somebody else = induced normative gradable change;
  - Can *categorize* killing this way, but it is not required

- A normative gradable property is a pair consisting of a gradable property and a designated degree;
- *Induced normative gradable changes* (i) include an activity that leads to a result (ii) the result is “equivalent” to a gradable normative property, (iii) the proto-patient argument's degree on the relevant scale has increased during the event.

A three-way distinction



Incremental vs. non-incremental normative gradable changes

- We need to distinguish between gradable changes where the change is necessarily progressive/incremental (*write*) and gradable changes where the change is not (*kill*);
- A change is incremental if there is a correspondence between the event's parts and the ordered set of degrees on the scale.
  - The more you write, the more is written

Mandarin vs. English

- Incompleteness Hypothesis:
  - *In languages like Mandarin, sentences with main verbs that describe induced gradable changes entail that a change occurred with degree* $d_0 < d \leq d_N$.
  - *In languages like English, telic sentences with corresponding main verbs entail that a change occurred with degree* $d = d_N$

Conclusion 1

- The Incompleteness Hypothesis is hard to explain in a "generative" interpretation of a Vendler-Dowty-style decomposition:
  - The difference between Mandarin and English is *not* in their base predicates or the meaning of CAUSE or BECOME, but in "molecular" meanings built out of these;
  - There are interesting systematic "molecular" differences in lexical meaning across languages

2. Is lexical semantics recursive?

- Can CAUSE and BECOME be used recursively in lexical semantic representations, as a "generative" interpretation of lexical decomposition might suggest?
  - Are there words that mean X CAUSE Y TO CAUSE Z…

The one-cause constraint

- No more than one cause-effect per monomorphemic root = **one-cause constraint** (Carter; Levin and Rappaport, 1998)

  - Cause(x, Become (Pred(y_n)))
  - *Cause (t, Cause(x, Become (Pred(y_n))))

- Some verbs seem to challenge the one-cause constraint.

An impossible verb?

‘Carles esteit es vals de Moriane,
Quant Deus del cel li mandat par sun a[n]gle,
Qu'il te dunast a un cunte cataignie’ *Chanson de Roland*, CLXXII

‘Charles was in the Maurienne valey when God from the sky **had** his angel **tell** him that he would give you to one of his captain counts’

Dictionnaire de l'Académie française, 1st Edition (1694)

MANDER. v. act. Envoyer dire, faire sçavoir, ou par lettres, ou par Messager. (…) je luy ay mandé par un tel que ....

‘send to say, cause to know, by letters or messenger’

On dit prov. pour faire entendre qu'on n'a point craint de dire en face à quelqu'un une chose fascheuse, Je ne luy ay point mandé, je luy ay dit.

‘To imply that one didn’t shy away from being the bearer of bad news’

Trésor de la Langue Française

1. **a)** Faire savoir, faire connaître quelque chose à quelqu'un par lettre ou par message.

   (have somebody know something via a letter or a message)

mander': X cause Y to tell Z that P *or*

X cause Y to cause Z to know that P

A systematic class of “exceptions”

- *Mander* is only one of a list of verbs that appear to challenge the one-cause constraint: Verbs that denote situations that require an instrument;
  - We need to add a “use tools to induce a change of state” to the maximum complexity of verbs;
  - The complexity of lexical meaning is more than previously thought *but* there is still no need to recursively use operators like CAUSE

Obligatory/Optional instruments

A meaning of 12% of the English verbs require and instrument and that of 35% merely allow an instrument.

Semantically Obligatory Instrument Verb

The barbarian hacked someone with a sword during the attack

Semantic Optional Instrument Verb

The barbarian injured someone with a sword during the attack

Methodology

- We classified the list of verbs that semantically require (≈500) or allow an instrument (≈1,300) in terms of:
  - Subsituations: *s1* (Agent and possibly instrument); *s2 (Instrument and possibly patient); s3 (patient and possibly instrument) (crucially, s2 was not necessarily present);*
  - *s_1 precedes s_2; s_2 precedes s_3.*

Examples

CUT.

cause(s1, s2) ∧ act(s1, A, I) ∧ contact(s2, I, P) ∧ cause(s2, s3) ∧ incised(s3, P)

1. Incise : *carve (a piece of wood), notch, plow, scratch, etch ;*
2. Pierce : *puncture, harpoon, knife, prick, lance ;*
3. Sever : *amputate, bone, core, eviscerate, castrate, gore, hack, prune, mow ;*
4. Shred : *shred, it includes grind, dice, cube, scallop, and mince ;*

DRUG: *drug, gas, anesthesize, immunize, vaccinate, dope ; flavor season.*

cause(*s1, s2)* ∧ *act(s1, A, I)* ∧ *in (s2, I, P)* ∧ *cause(s2, s3)* ∧ *change-of-state(s3, P)*

FILL.

cause(*s1, s2)* ∧ *act(s1, A, I)* ∧ *in(s2, I, P)* ∧ *cause(s2, s3)* ∧ *change-of-configuration(s3, P)*

(1) Jim loaded the truck with boxes with a forklift

SKI. *Canoe, bicycle, skate, drive, toboggan.*

cause(*s1, s2)* ∧ *act(s1, A, I)* ∧ *pred2(s2, I, A)* ∧ *and* ∧ part-cause+(*s2, s3)* ∧ $move_{manner}$*(s3, A)*

SCOOP. *Spoon, pump, milk, sponge, ladle, shovel, siphon.*

cause(*s1, s2)* ∧ *act(s1, A, I)* ∧ *in(s2, P, I)* ∧ *enable(s2, s3)* ∧ *go-to(s3, P, Z)*

(1) The plug's coming loose let the water flow from the tank.

Are instrument verbs exceptions to the one-cause constraint?

*Entailments of **cut** (with animate subject)*

(1) there is a situation s1 which is an act performed by an agent A on an instrument I; (2) there is a contact (with force) relation s2 between I and a patient P; (3) there is a (final) state of incision for P; (4) s1 caused s2, which in turn caused s3.

- Are the two causes part of the semantic *structure* of *cut*?

Iterative adverbs and multiple instruments

Iterative adverbs (s3 or s3+s2)

(1) a. Bill fastened Marc's seat belt again.
b. A KGB agent poisoned Marc again.

Multiple *with*-PPs (s2)

(2) a. The nurse anesthetized him (with a mask) (with more than ether).
b. Marc bakes pizza (with a wood oven) (with yeast).
c. Jim filled the truck (with boxes) (with a forklift).

(3) a. The nurse used a mask to anesthetize him with ether.
b. ??The nurse used ether to anesthetize him with a mask.

Patient-subject inchoatives (s1) (*pace* Koontz-Garboden)

(1) a. Bill cut the bread with the knife.

b. The knife cut the bread.

(2) a. cause(s1, s2) ∧ act(s1, A, I) ∧ contact(s2, I, P) ∧ cause(s2, s3) ∧ incised(s3, P)

b. contact(s2, I, P) ∧ cause(s2, s3) ∧ incised(s3, P)

Conclusion 2: The one-and-half cause constraint

- Expansion of maximum bound on *structural* semantic complexity is needed, but limited: **use of tools/middle-men**;
  - A more precise semantic representation of the **CUT** class is:

  use (s1, A, I, s'2) ∧ s2 < s'2 ∧ s3 < s'2 ∧ contact(s2, I, P) ∧ cause(s2, s3) ∧ incised(s3, P)
- Why cannot semantic structure be more complex?

3. How many *CAUSE* operators are there?

a. John watches birds all day with his binoculars.

b. Bill cooks his steaks with butter.

c. Floyd baked the cake with yeast.

d. Bill entered Joan's room with a duplicate key.

e. Joe scooped the ice-cream with a wooden spoon.

f. Connie skied down the slope with her new skis.

g. Alisa walks her cat with a leash.

A family of CAUSEs

- 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 better: *cook with butter*, *lecture with slides*, *watch with binoculars*

A meta-intensional analysis of "helping"

- e1 helps e2 if it causes e2 to be "better" than it otherwise would have been.
  - The steak's taste is "better" with butter than without butter.
- **Definition 1** 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.

Meta-intensional causality

- Our analysis of helping seems to make causality dependent on how events are *described* (e.g., cooked *better*);

(1) Jeri's new shoes made her run *fast* (#made her run);

(2) Marc's numbness made him drive *above the speed limit* (#made him drive);

(3) Roberto's painkillers made him paint *less realistically* (#made him paint);

- But isn't it the same event that can be categorized as running or running fast, or…?

Two solutions

- Meta-intensionalize events: events are categorized processes (Link);
- Meta-intensionalize causal statements: causal statements do not relate two events vis-à-vis possible worlds, but two events as instances of two categories vis-à-vis possible worlds;

Conclusion 3

- CAUSE is a cover term for a family of operators:
  - The meaning of these operators requires an meta-intensional definition that refers to the event's category/description

4. Semantic space is not (always) a democracy

- Within each quadrant of semantic space, each verb meaning is the equal of any other verb meaning:
  - Semantic similarity (distance) is a symmetric relation;
  - True when describing the meaning space from a theoretical semantic/philosophical perspective;
- Some verbs are more important than others *grammatically* and their meanings are also more important than others w.r.t. some grammatical processes.
- Some meaning distinctions are syntactically irrelevant:
  - Lexical submodality (Koenig and Davis)

Semantic Attractors

- There is a semantic "basis" for the set of syntactic valences/subcategorization a verb has (Pinker, Levin);

(1) Mary *gave* John a book (ditransitive valence: NP V NP NP)
(2) Mary *gave* John a book (PO valence: NP V NP PP)

- Some verbs describing "basic scenes" (e.g., *give*, *put*) might serve as attractors for other semantically similar verbs when it comes to valence possibilities:
  - This might help learning of argument-structure constructions (Goldberg);
  - $\Rightarrow$ Semantic similarity might be *asymmetric*!

Semantic Similarity Hypothesis (Yi and Koenig)

- The more semantically similar verbs are, the more their valence profiles will be similar;
  - A verb's valence profile: The set of possible valences of a verb and their associated frequencies.
- The more semantically similar to an "attractor" verbs are, the more their valence profile will be similar to that of the attractor;

Testing the similarity hypothesis on the ditransitive construction

- *Give* is the prototypical ditransitive verb:
  - Most frequent verb in the construction
  - Learnt first by kids;
- Is it an attractor for other verbs w.r.t. the ditransitive?
  - If yes, we expect verbs that are more semantically similar to *give* to match better its valence profile (= proportion of ditransitive uses)

“Give” as a ditransitive attractor

	NP NP	**GIVE NP PP**	
Roland’s corpus	19,112	11,795	61.8%

	NP NP	**GIVE NP PP_{to}**	
Switchboard	303	0	100.0%
BNC (Charniak)	15,021	8,199	64.7%

- Give = 53% of ditransitive uses of *all* verbs!

How we tested the attractiveness of *giving*

- For each verb cited in Levin 1993 as occurring in the ditransitive construction and which occurred in the BNC:
  - We estimated their semantic similarity to *give* using Latent Semantic Analysis (Landauer et al.);
    - LSA measures semantic similarity by calculating co-occurrences of words in contexts;
    - Semantically similar words occur in the same "texts" or with a similar set of words;
  - We estimate their (ditransitive) valence profile by computing the proportion of V NP NP and V NP PP uses in a parsed BNC corpus;

Predicting valence profiles

- **Hypothesis:** There is a correlation between how semantically similar to *give* a verb is and the proportion of times it is used in a ditransitive frame;

Correlation (Pearson)	***Partial Correlation***
$r = 0.379$, $p < 0.01$	$\mathbf{r = 0.314,\ p < 0.01}$

- Many factors influence the choice of a ditransitive valence over a PO valence (Bresnan et al.): One of them is how semantically similar to *give* a verb is!

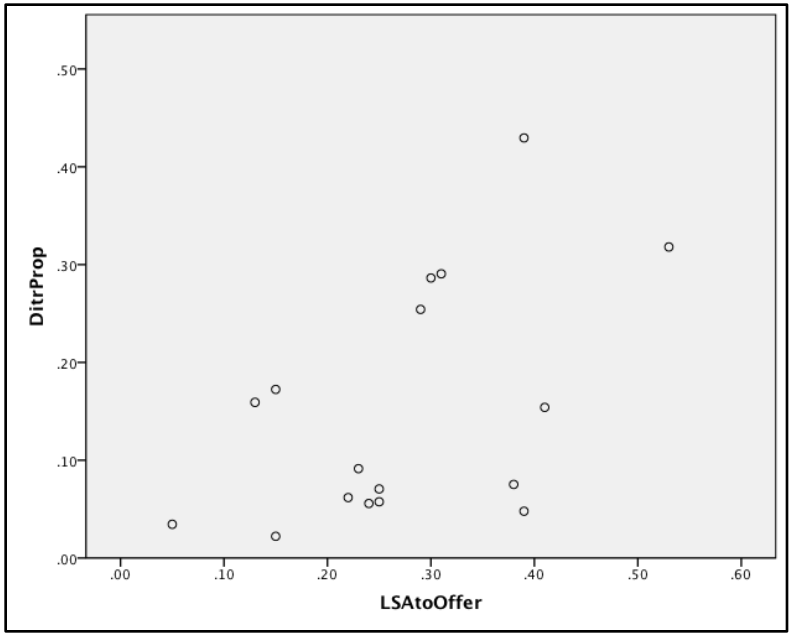
Narrow class attractors

- If *give* is an attractor, maybe each narrow ditransitive class (*give* class, future having class…) are better attractors than *give*?

Verb class	**Attractor**	**Verb class**	**Attractor**
Future having	*offer*	*Send* class	*send*
Transfer message	*tell*	Instrument of communication	*sign*
Give class	*pay*	*Slide* class	*slide*
Carry class	*carry*	*Bring* & *take*	*bring*
Drive class	*drive*	*Throw* class	*throw*

Narrow class attractors: *Offer*

Correlation (Pearson)	***Partial Correlation***
$r = 0.505$, $p < 0.05$	$\mathbf{r = 0.508,\ p < 0.05}$



Transfer of possession… or not

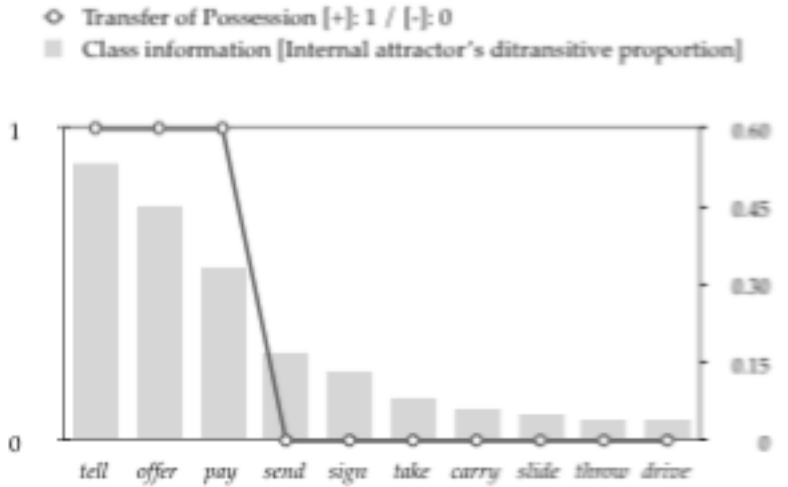
- The correlation between semantic similarity to *give*/valence profile hides differences between three distinct cases:

1. Transfer of possession classes (Rappaport Hovav and Levin);
2. The *send* class;
3. Other classes

Two regression models

[A]+[B]	Predictors	R	R^2	Sig. F change
Model 1	[ii] Internal attractor' s ditransitive proportion	0.469	0.220	0.000
Model 2	[i] Semantic similarity to *give* [ii] Internal attractor' s ditransitive proportion	0.515	0.265	0.011

	Predictor [i]	Predictor [ii]	R	R^2	Sig. F change
[A] only	✓	×	0.423	0.179	0.005
[B] only	×	✓	0.418	0.175	0.001



What causes the correlation?

- Meaning of construction = meaning of *give*
  - Semantic similarity to *give* = degree of "fit" between meaning of verb and meaning of construction;
  - What about correlation that pertains to narrow class internal attractors?
- Structural "priming" (Bock, Pickering and Garrod, Snyder):
  - Semantically similar verbs occur in the same context and structurally prime each other

Conclusion 4.1

- The meaning of verbs can gravitate around the meaning of other verbs (directly or indirectly) for syntactic purposes;
  - How far in semantic space a verb is from an attractor is one predictor of its syntactic distribution (its valence profile, e.g., ditransitive vs. PO valences);

Some semantic differences do not matter for syntax: sublexical modality

(1)Burns sent/offered/owed/promised/charged/denied Smithers $10 for the dinner.

(2) Sue perceived/noticed/overlooked/missed him.

(3) Bill had/received/lost/lacked/needed many books

(4) Bill managed/tried/failed/neglected to read the books.

Event core and sublexical modality

	Neutral	Negative	Deontic	Energetic	Inchoative	Inchoative -negative
Possession	have	lack	need		receive	lose
Perception	perceive	miss	overlook		notice	
Causing possession	give	deny	Promise, owe, charge	Send, offer		
Carrying out an action	manage	fail	neglect	try		
Causing another to act	Force, persuade	forbid	require	Urge, defy		

Conclusion 4.2

- Subject and object assignment ignores systematically some lexical semantic information:
  - Verbs that share a situation core form an equivalence class w.r.t. subject and object assignment;
  - Lexical entailments relevant for subject and object assignment can be relativized to a subset of possible worlds.

5. What about idiosyncratic semantic information?

- What parts of event structure is specified by individual verbs?
- *Idiosyncratic* information specifies more instrument activity and change of state in patient than agent activity*;*
  - Obligatory instrument verbs specify more s2 and s3 than optional instrument verbs

Obligatory class	s_1	s_2	s_3	Causality	**I**
Cut	act	contact	incised	cause	
Whip$_1$	act		contact	N/A	
Whip$_2$	act	contact	change-of-state	cause	
Whip$_3$	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$_{\text{manner}}$	part-cause$^+$	
Cover$_1$	act		cover	N/A	
Cover$_2$	act	at	obscure	cause	
Garnish	act	at	more-ornate	cause	
Scoop	act	in	go-to	enable	
Doodle	act	mark	represent	cause	
Notify	produce$^+$		change-of-state	cause$^+$	expression
Roast	act	contact	cooked/hot	cause	hot
Phone	act		pred$_1$	cause$^+$	

Table 1 Semantic category of s_1, s_2 and s_3, the kind of causality involved and the semantic category of **I** for obligatory instrument classes

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	$pred_2$	$move_{manner}$	$part\text{-}cause^+$	
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

Conclusion 5

- Idiosyncratic aspects of verb meaning target more result states than causal activities.
  - Another example of goal bias (voir Lakusta and Landau (2005));
  - Another example of the lexical reification of discourse distribution (Slobin (2004));