Why do verbs with similar meaning behave similarly syntactically?

Eunkyung Yi and Jean-Pierre Koenig
(jpkoenig@buffalo.edu)

Where we are

- Previous lecture: Representation of nouns with multiple senses (shared features) and how it affects processing
- This lecture: Verbs have more or less similar meanings
    - How that similarity of meaning relates to syntactic similarity
    - How that similarity of meaning affects sentence production

A menagerie of syntactic frames

- Verbs in most languages occur in many different syntactic frames
- In GKPS (1985) about 40 VP rules are listed
- In GGF there are 28 main syntactic frames for verbs
- A single verb can occur in several VP frames (valence alternations)

The meaning of a verb is “predictive” of its syntactic frames

- Null hypothesis: The relation between the meaning of verbs and the syntactic frames they can occur in is arbitrary

- Linguists have documented for a long time (50 years, starting with Fillmore) that the null hypothesis is false

- Is it false *just* for coarse semantic reasons (NP_NP is impossible for one-place predicates) or more subtle semantic/syntactic reasons?

    1. This room seats/sleeps/fits thirty people.

- How predictive of syntactic frame(s) is the meaning of verbs? Hard to tell!

What are valence alternations?

= Pairs of syntactic frames that sets of verbs can occur in and where the verbs have *roughly* the same meaning

1. Joann took the bow to Dennis

2. Joann took the bow from Dennis

3. The television series never took and was later canceled

Different classes of valence alternations

Information load	Object drop
Information structure	Passive
Event set-subset	Conative, Ditransitive
Event part	Inchoative

Not all types of valence alternations are created equal

- Some alternations are actually alternations in meaning (Ackerman *morpholexical operations,* causative-inchoative: Change in valence are predicted by change in meaning

- Some alternations are more morphosyntactic (active-passive): Change in valence is not entirely predicted by change in meaning, if any

- Important distinction as some languages may necessarily lack the latter (e.g., Oneida, Northern Iroquoian)

- How much semantic difference between the two alternate frames in a valence alternation is a subject of debate for many alternations

The ditransitive alternation

1. Mary gave a book to Sam
2. Mary gave Sam a book

- About 200 verbs participate in the valence alternation
- It is productive (in the extensibility sense of the term)

In search of an explanatory mechanism for the pairing of meaning and syntactic frames

- The null hypothesis could not have been right

- But what mechanism explains that the meaning of verbs “predicts” their context of occurrence?

- Our answer:

  - “Spreading activation”: Activating a meaning (concept) activate related meanings (concepts)

  - Association strength between verbs and syntactic frames

Production of a sentence

1. *Think of a message*

2. Think of “essential” lexical items

3. *Retrieve relevant constructions (including syntactic frames for verbs)*

4. Output

“Spreading activation”

- Old way of thinking about it: Localist view
    - Nodes representing concepts
    - Links between nodes
    - A certain amount of activation going around
    - Activation diminishes with number of nodes to spread to and distance

Distributed vs. localist representations

- Localist view is tied to a view of concepts as monads and relation between concepts handled through something akin to meaning postulates

- Distributed view: a concept is represented as a packet of nodes

- Relations between concepts = overlap in sets of activated nodes

- “Spreading activation" is actually activation of overlapping features

Dell production model

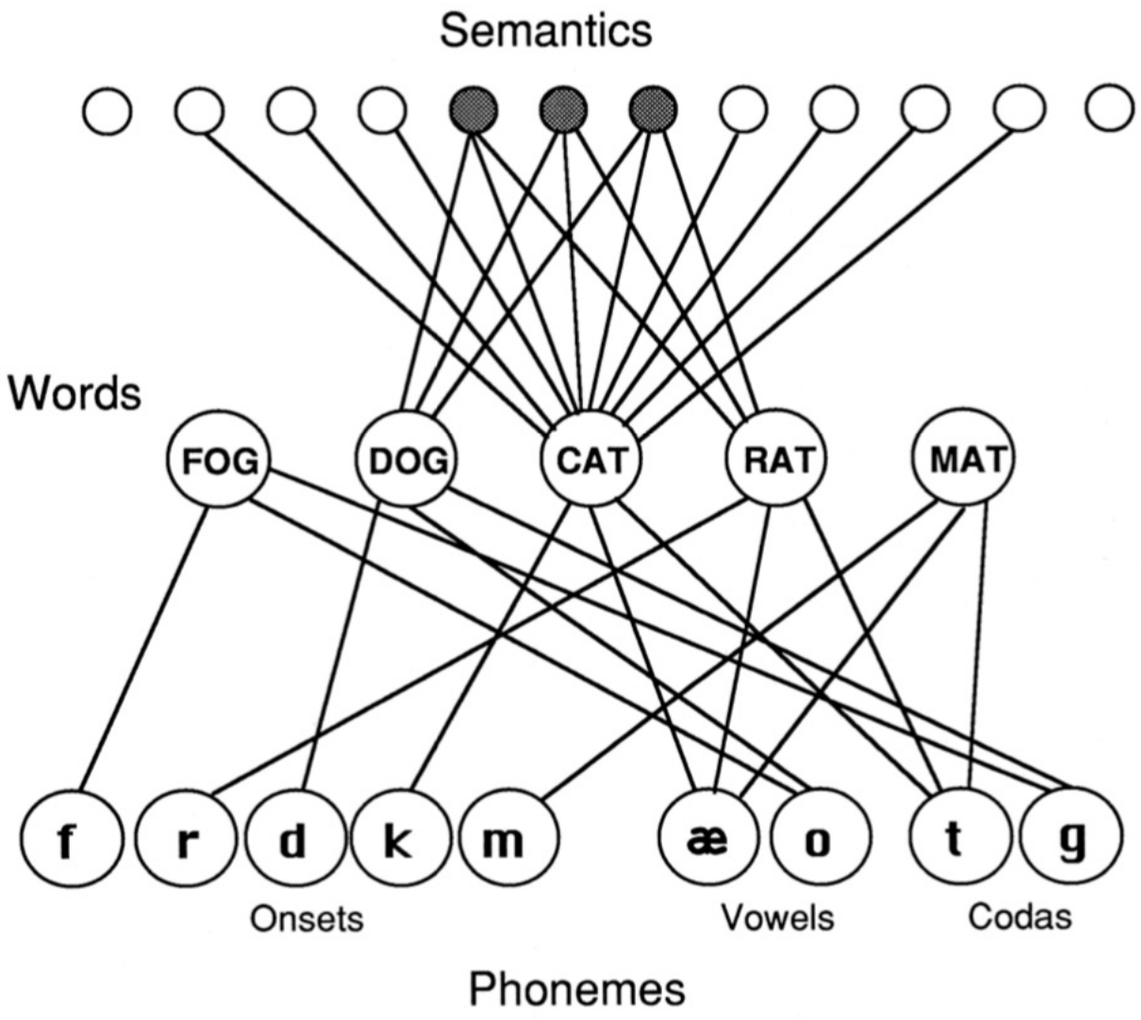


Figure 1. The aphasia model. Connections are excitatory and bidirectional.

First component of explanation: Activation of semantically similar meanings

- Verb meaning consists of sets of features and similarity *between* verbs consists of overlapping features

- When we think of a situation category we want to talk about, other categories that share features with it are activated

- Words lexicalizing the category we want to talk about *and* verbs lexicalizing categories that overlap with it are activated

- The activation of other verbs depends on how similar their meaning is to the verb we chose to express our message

Second component of explanation: Path-breaking verbs

- Some verbs occur very frequently in a syntactic pattern (*give* for ditransitive; *put* for caused motion construction)

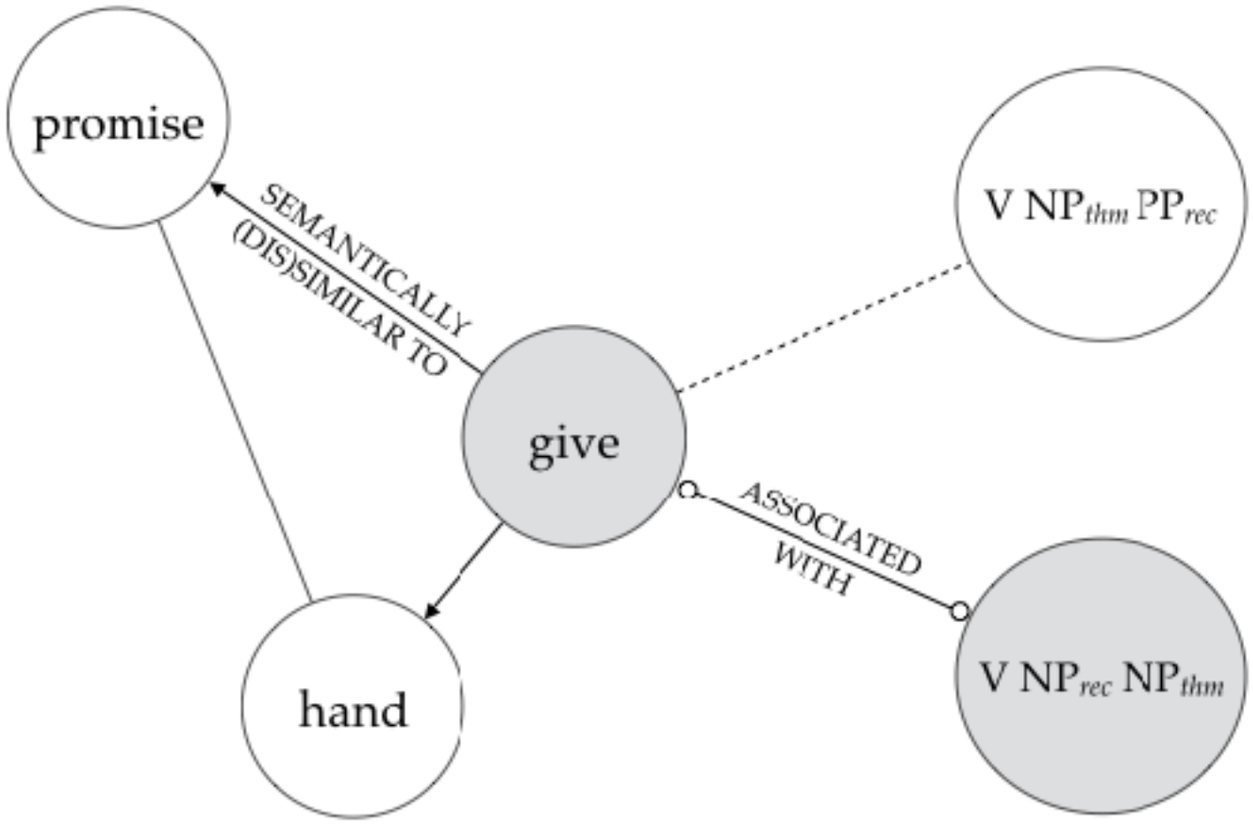
- These verbs have general meaning similar to the meaning of the construction, are best representative of construction

- These verbs may play an important role in learning the construction (Goldberg, Ellis): Uneven distribution helps learning

- These verbs are strongly associated with a particular syntactic frame

The Verb Anchor Hypothesis

- Verbs that are strongly associated with a syntactic frame may lead verbs semantically similar to them to occur in that syntactic frame



The mechanism behind the Verb Anchor Hypothesis

- When you think of an event category, event categories with overlapping features are activated

- Words associated with these other event categories are activated

- Syntactic frames (strongly) associated with those event categories are activated

- This activation increases likelihood that speakers choose one frame rather than the other in a valence alternation

What the VAH predicts

- The more similar a verb is to the syntactic frame anchor, the more likely it is to occur in the anchor's "favorite" frame

- Measure semantic similarity between verbs

- Estimate distribution of verbs in syntactic frames

A BNC count

- We started with Levin's (1993) list of alternating verbs (127 verbs)

- Only counted verbs that occurred in our corpus, had a transfer sense in our corpus, and did not have two senses in Levin's list

- We ended up with 105 verbs and 62,713 sentences

The ditransitive frame anchor: *give*

Verb	Tokens					Proportions
	Ditransitive (D)		Prepositional (P)		D+P	D:P
give	15,311	58 %	8,402	22%	23,713	65:35
other 104 verbs	10 732	42 %	28 268	78%	39 000	28:72
Total	26 043	100%	36 670	100%	62 713	42:58

How to measure association strength

- Various measures proposed to assess bias of verbs towards a particular frame (or association strength)

- These measures have *pro*s and *con*s; some are sensitive to absolute frequency of occurrence in frames; some sensitive to relative frequency of occurrence in frames

- For our purposes, makes no real difference

Collocation strength (Stefanowitsch and Gries)

Table 3. Crosstabulation of *accident* and the [N *waiting to happen*] construction

	accident	¬accident	Row totals
[N *waiting to happen*]	*14*	21	*35*
¬[N *waiting to happen*]	8,606	10,197,659	10,206,265
Column totals	*8,620*	10,197,680	*10,206,300*

- Fisher exact test to assess if a word favors one construction or another

- Cannot determine effect size

Delta-P (Ellis and Ferreira-Junior)

$$\Delta P = P(O|C) - P(O|{-}C)$$

- Probability of outcome given cue minus probability of outcome when there is no cue
- ΔP Construction → Word (ΔP Attraction) vs. ΔP Word → Construction (ΔP reliance)
- ΔP reliance overestimates importance of ditransitive biased *low* absolute frequency verbs (e.g., *email* 1 occurrence in ditransitive)

Hebbian rule

$\Delta w = \eta a_v a_F$

$w_{vF}(t) = \eta(t_1(v, +F) - t_2(v, -F))$

- Association between v(erb) and F(rame) is the learning rate times the number of times the verb is seen in the frame minus the number of times the verb is seen in the alternative frame
- Sensitive to the absolute frequency of occurrence of verbs (*oversensitive* for verbs with low association strength)

Very few verbs anchored to ditransitive frame

No	Verb (D:P)	Verb class	W_{verbD}	ΔP D->V	*p* of Fisher's	LSA cosine	LSA cosine *(residualized)*
1	give (15311:8402)	*Give*	1609	0,35879	0,00E+00	1,000	-0,311
2	tell (2702:339)	*TrsMsg*	2163	0,09431	0,00E+00	0,859	-0,311
3	ask (688:194)	*TrsMsg*	1094	0,02813	1,11E-109	0,857	0,389
4	teach (172:100)	*TrsMsg*	172	0,00530	0,02E-13	0,464	0,018
5	loan (12:11)	*Give*	1	0,00016	4,62E-04	0,190	-0,247
6	e-mail (1:0)	*Instr*	1	0,00004	1,25E-01	0,325	-0,111
7	promise (43:43)	*FutHav*	0	0,00048	4,74E-01	0,767	0,328
8	barge (0:1)	*Drive*	-1	-0,00003	3,00E-01	0,431	-0,005
100	pay (712:1363)	*Give*	-651	-0,00983	6,11E-145	0,555	0,043
101	leave (468:1390)	*FutHav*	-922	-0,01994	2,00E-50	0,857	0,353
102	sell (190:1288)	*Give*	-1098	-0,02783	4,49E-131	0,405	-0,085
103	send (658:3134)	*Send*	-2476	-0,06020	2,44E-236	0,765	0,189
104	take (2044:5620)	*Bring&take*	-3576	-0,07477	1,08E-182	0,946	0,227
105	bring (580:4927)	*Bring&take*	-4347	-0,11209	0,00E+00	0,891	0,252

How to measure semantic similarity

- Human judgment of similarity (c.f. first lecture)
- Wordnet:
  - Organized around a handful of semantic relations
  - Relations determined via human judgments
  - Requires choosing between senses of a verb

Distributional semantics

- Meaning of words, phrases, documents can be approximated (if not reduced to) by other words they co-occur with
- Latent Semantic Analysis (late 90's) first off the door
- Other measures available these days: word2vec, Glove
- Even if distributional semantics misses "true" meaning of texts (Glucksberg), it is a useful estimate of what we are after
- It does not depend on syntax or determining verb sense
- Correlation between LSA and human judgments:
  - 34 verbs; Pearson's $r = .534$, $p < .01$

Factors that influence ditransitive frame selection

- Bresnan et al. (2008): (mostly) verb-external factors:
    - Pronominality of theme/recipient, animacy, definiteness of theme/recipient, difference in length of theme and recipient phrases
    - Future having, transfer semantic classes (based on full sentence ratings)
- Verb-internal factors we focus on:
    - Semantic similarity to anchor (*give*)
    - Entailment of caused possession

Caused possession entailment (Rappaport and Levin 2008)

(1) A woman *gave/threw* her friend a ball.

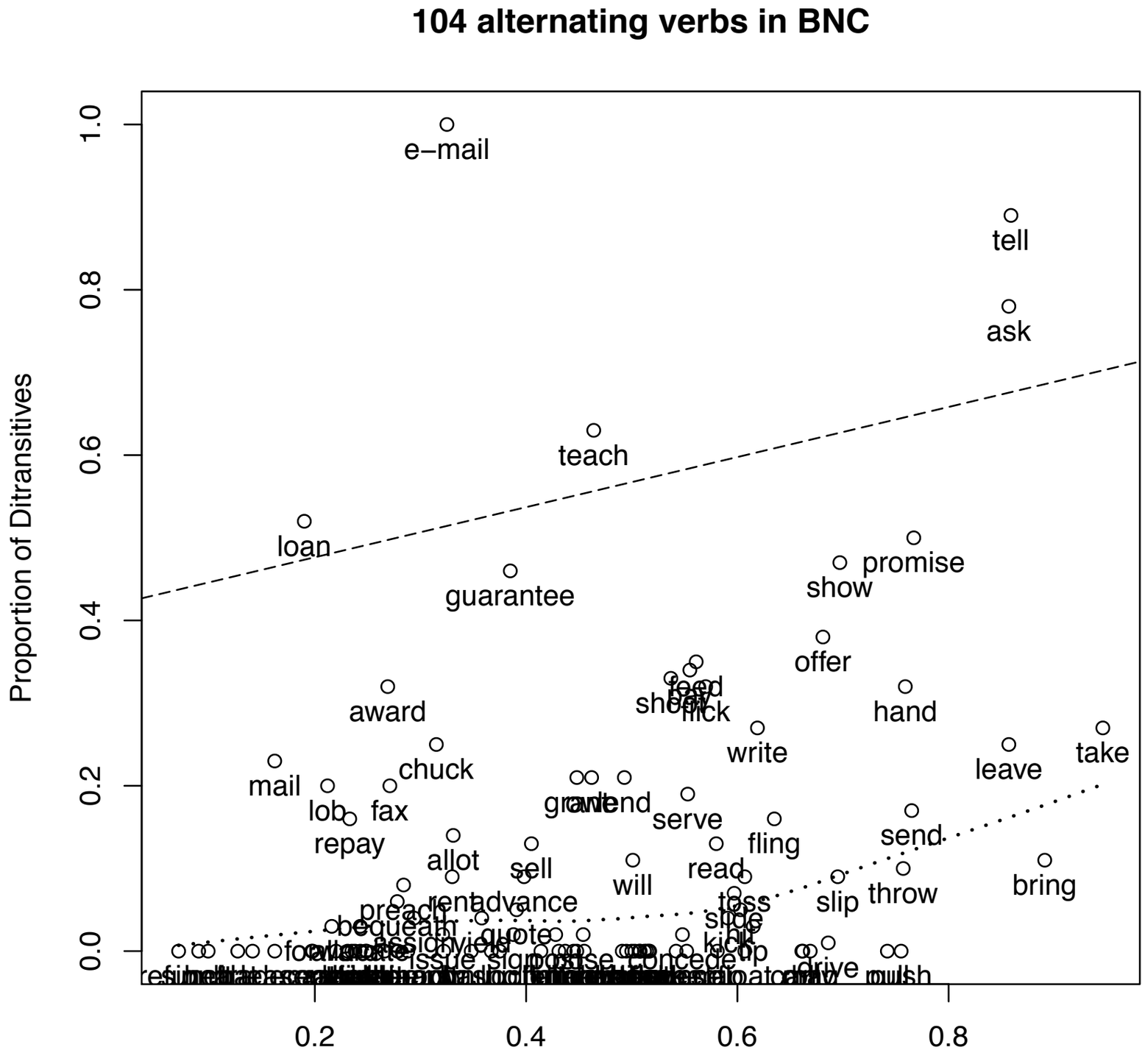
(2) A woman *gave/threw* a ball to her friend.

- Some verbs “entail” a transfer of possession whichever syntactic frame they occur in (e.g., *give*)
- Some verbs “entail” a transfer of possession only when occur in the ditransitive frame (e.g., *throw*)

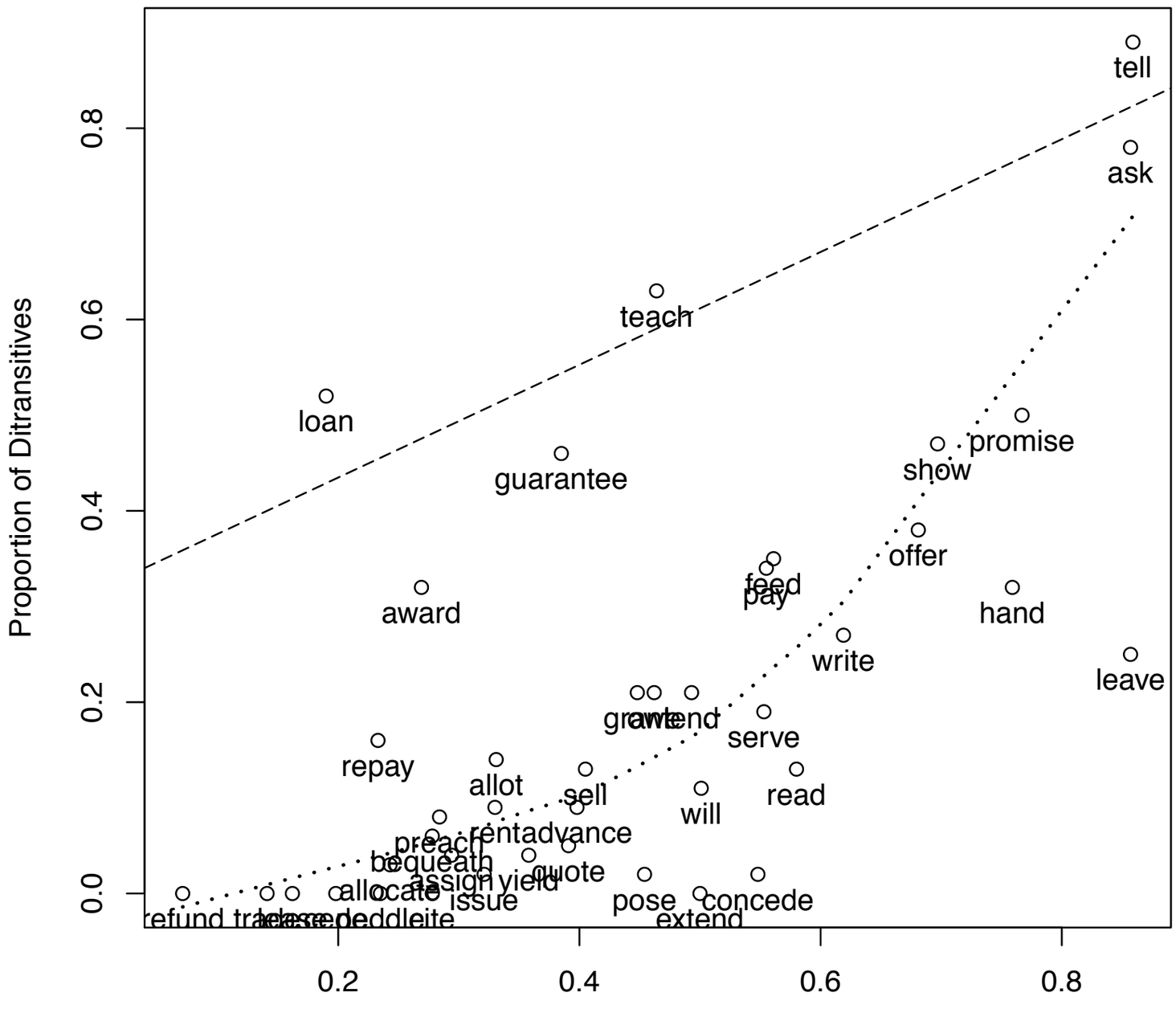
Restricted “entailment”

- The notion of entailment Rappaport and Levin have in mind is not standard
- Verbs like *offer*, *promise*, *deny*, … do not entail transfer *stricto sensu*, but only in a subset of possible worlds
    - Koenig and Davis (2001) provide an analysis of such restricted “entailments” making use of a notion of sublexical modality

Correlation study I



Correlation study II



The VAH is empirically supported

- Verbs that are more semantically similar to *give do* occur more in the ditransitive

- This is even more true of verbs that entail caused possession irrespective of syntactic frame

- First evidence we know of that lexical meaning affects *choice* of semantic frame (and aggregatively, relative frequency of occurrence in one of two alternating frames)

Adding verb external factors

- Two logistic regression models:
  - One on our BNC data, only used predictors we can determine automatically (pronominality of theme/recipient; length difference)
  - One on Bresnan's data with most of their predictors

Semantic similarity is independent of verb external factors

- All predictors are independently significant predictors of the sentences' syntactic frame:
    - Verb similarity to *give*, entailing caused possession irrespective of syntactic frame, pronominal recipient, pronominal theme, and recipient-theme length difference
- This is true fo both datasets

Comparable results with Bresnan et al.'s dataset

Predictors	A Replication of Bresnan et al.'s	B ⊕ sim-to-*give* ⊕ csd-poss	C ⊕ prorec-by-prothm interaction
inanimate recipient	3.59 ***	3.64 ***	3.82 ***
inanimate theme	-1.20 *	-1.26 *	-1.28 *
nonpronominal recipient	1.21 ***	1.33 ***	-0.31 ns
nonpronominal theme	-0.70 *	-0.88 **	-1.32 ***
nongiven recipient	1.38 ***	1.34 ***	1.34 ***
nongiven theme	-1.14 ***	-1.18 ***	-1.14 ***
indefinite recipient	0.56 *	0.70 **	0.65 **
indefinite theme	-1.25 ***	-1.19 ***	-1.20 ***
transfer semantic class	0.05 ns	-0.30 ns	-0.29 ns
communication semantic class	-2.62 ***	-2.15 ***	-2.05 ***
future having semantic class	-1.36 **	-1.32 **	-1.29 **
length difference	-0.91 ***	-0.91 ***	-0.93 ***
verb similarity to *give* (LSA cosines)	-	-3.54 ***	-3.58 ***
verb caused-possession entailment	-	-1.20 ***	-1.28 ***
nonpron recipient : nonpron theme	-	-	1.88 ***

Are a couple of frequent verbs causing the verb anchor effect?

Dataset	Entire data	Subsets					
		[1]	[2]	[3]	[4]	[5]	[6]
# of verbs	**104 verbs**	**103 verbs** *(- tell)*	**102 verbs** *(- tell, take)*	**101 verbs** *(- tell, take, bring)*	**100 verbs** *(- tell, take, bring, send)*	**104 verbs (**max freq. = 100/verb)	**57 verbs** **(**max freq. = 100 & excl. D=0 verbs)
D sentences	10 732	8 030	5 986	5 406	4 748	851	851
P sentences	28 268	27 929	22 309	17 382	14 248	3 965	2 900
D + P total	39 000	35 959	28 295	22 788	18 996	4 816	3 751
D:P	28:72	22:78	21:79	24:76	25:75	18:82	23:77
# of sentences *per* verb mean sd	375 1072.5	350 1044.6	277 752.9	226 544.4	190 411.8	46 40.1	66 35.7

The anchor effect is not tied to one or two very frequent verbs

Dataset	Entire data	Subsets					
		[1]	[2]	[3]	[4]	[5]	[6]
	0.50 ***	0.31 ***	0.33 ***	0.36 ***	0.40 ***	0.63 ***	0.55 ***
b coefficient & significance							
Odds ratio	1,65	1,37	1,39	1,44	1,49	1,88	1,73
95% C.I.	1.59~1.71	1.32~1.42	1.33~1.45	1.37~1.51	1.41~1.56	1.66~2.13	1.53~1.96

Only one consistent interaction

- The VAH does not make any prediction regarding interactions between predictors

- Only one interaction significant across data sets: semantic similarity to give X pronominality of theme X pronominality of recipient

- This three-way interaction is driven by the two-way interaction between pronominality of theme and pronominality of recipient

(1) *Joann sent him it/Joann sent it to him

VAH does not depend on particular estimate of semantic similarity (WordNet)

- We ran a model where we used a Wordnet-based vector measure of semantic similarity

- 49 verbs in four subclasses of alternating verbs (*give* verbs, message transfer verbs, future having verbs, *send* verbs)

- Used the mean of the vectors between verb senses as measure of similarity

- We replicated the effect of semantic similarity to *give* on syntactic frame selection

VAH does not depend on particular estimate of semantic similarity (GloVe)

- We ran a model using GloVe as a measure of semantic similarity (Pennington et al. 2014)

- The GloVe measure of semantic similarity was trained on a 6-billion word corpus (Wikipedia and Gigaword 5) and used 300 dimensions

- All verb-internal and verb-external predictors remained significant when the GloVe measure is substituted for an LSA measure

- Strong negative correlation between GloVe semantic distance and LSA cosines (Pearson's r = -0.669, $p < .001$)

Narrow classes

- Pinker argues that more semantically coherent subclasses are the locus of predictability in valence alternations

- Does the VAH hold of narrow classes too?

- Three classes considered: *give*-class, message transfer class, and verbs of future having

Narrow classes (from Beavers 2011)

(6) a. **Verbs that inherently signify acts of giving:** *give, pass, hand, sell, pay, trade, lend, loan, serve, feed*

b. **Verbs of sending:** *send, mail, ship*

c. **Verbs of instantaneous causation of ballistic motion (verbs of throwing):** *throw, toss, flip, slap, kick, poke, fling, shoot, blast*

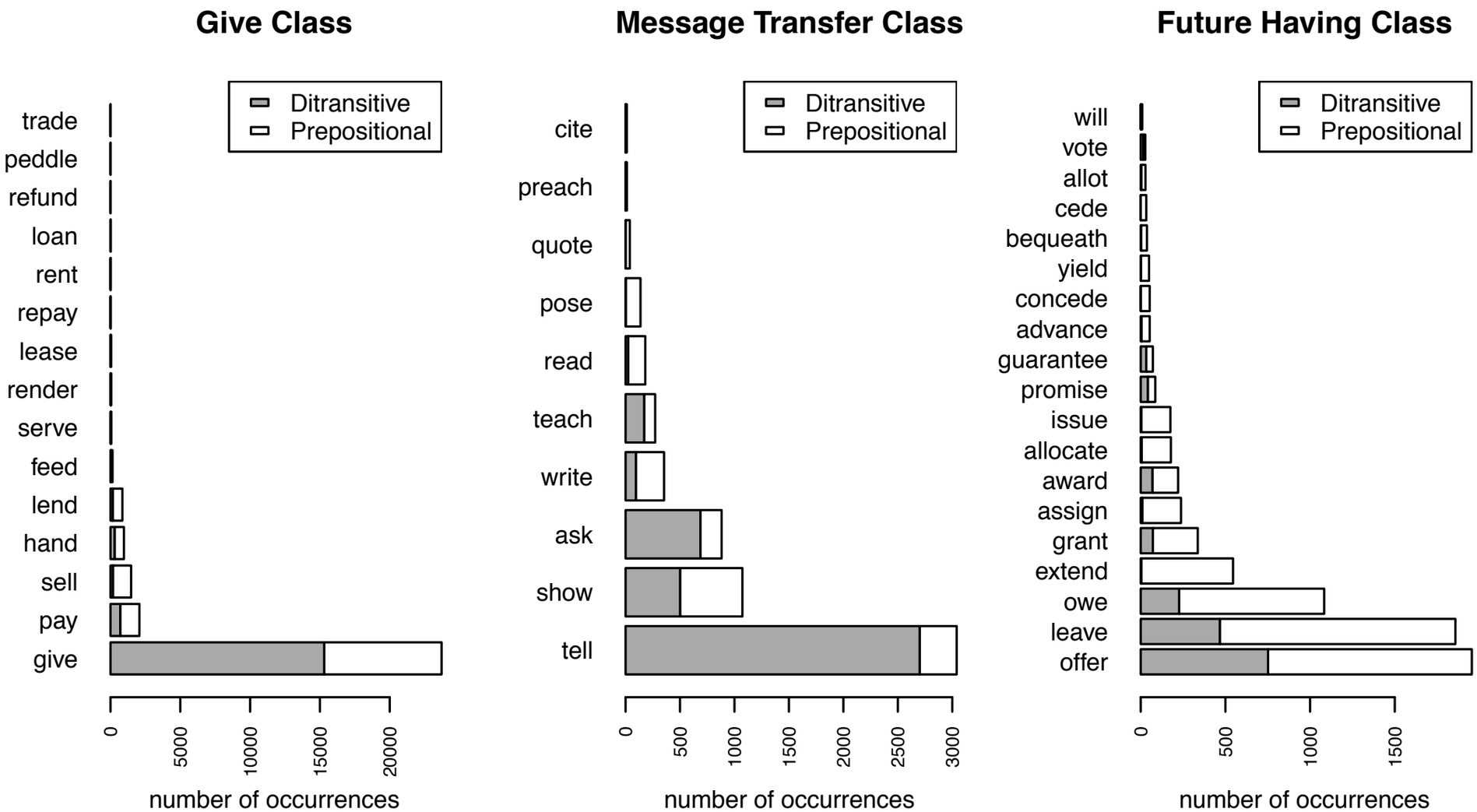
d. **Verbs of continuous causation of accompanied motion in a deictically specified direction:** *bring, take*

e. **Verbs of future having:** *offer, promise, bequeath, leave, refer, forward, allocate, guarantee, allot, assign, allow, advance, award, reserve, grant*

f. Verbs of type of communicated message: *tell, show, ask, teach, pose, write, spin, read, quote, cite*

g. **Verbs of instrument of communication:** *radio, email, telegraph, wire, telephone, netmail, fax*

Verb frequency differs across narrow classes



Give-class

pay, sell, hand, lend, feed, serve, lease, repay, loan, rent, refund, peddle, trade

- Semantic similarity to *give* is a significant predictor of the selection of the ditransitive frame

- Confirms that our results on the entire class of alternating verbs are not due to a few frequent verbs

Verbs of transfer of message (*tell* is the anchor)

show, ask, write, teach, read, pose, quote, preach, cite

- Similarity to *tell* is a significant predictor of ditransitive use
- … but only if similarity to *give* is not included in the predictors
- Extremely high correlation between similarity to *give* and *tell* (Pearson's $r = .94$, $p < .001$)
- *Give* is a better predictor because it is a better anchor (given the difference in raw frequency between the two verbs)

Verbs of future having (*leave* as anchor)

offer, owe, extend, grant, assign, award, allocate, issue, promise, guarantee, advance, concede, yield, bequeath, cede, allot, and *will*

- Similarity to *give* is a significant predictor, but not similarity to *leave*

- Significant interaction between similarity to *give* and *leave*:

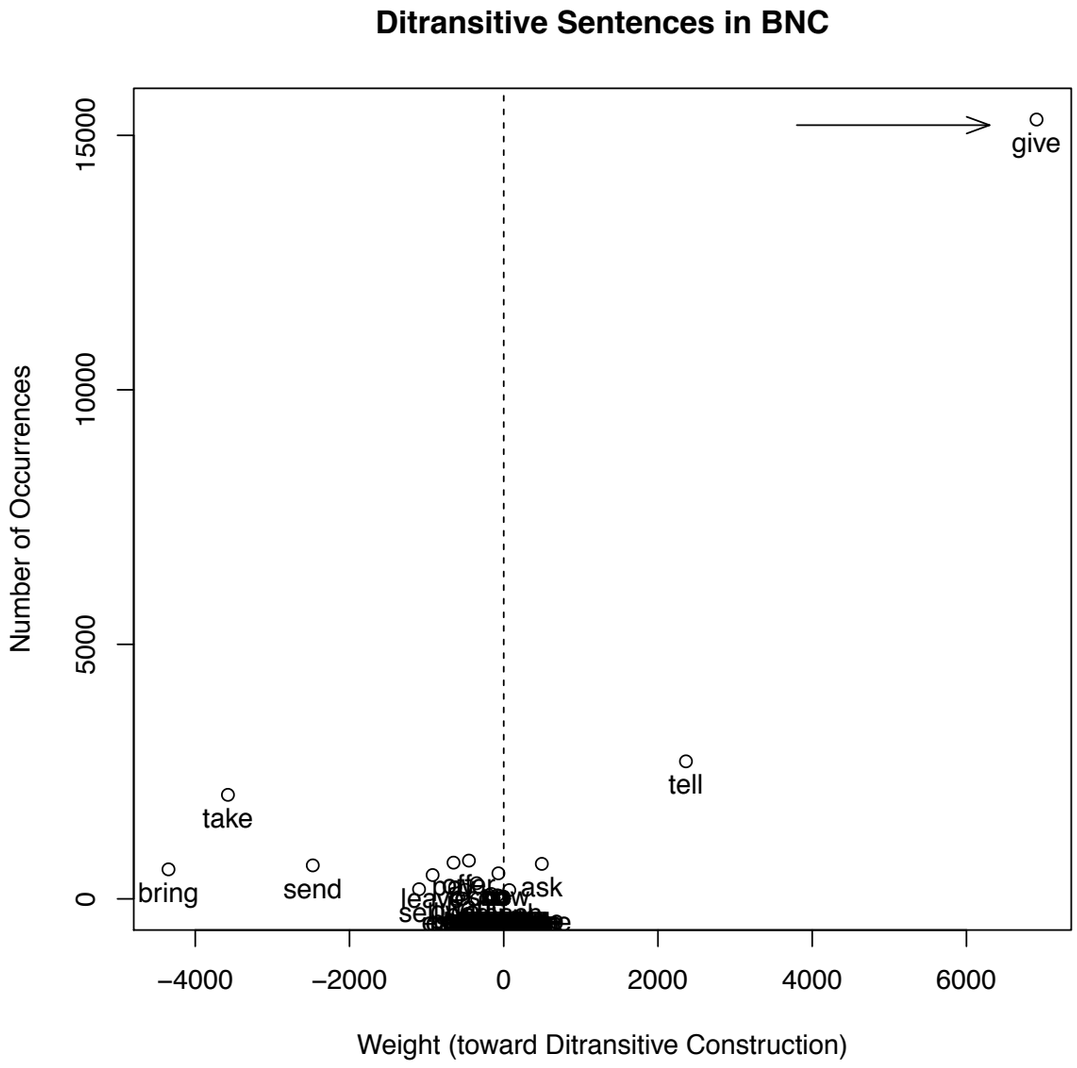
  - High similarity to *leave* slightly increases likelihood of selecting prepositional frame for verbs of low-similarity to *give* but slightly increases likelihood of ditransitive frame for verbs of high-similarity to *give*

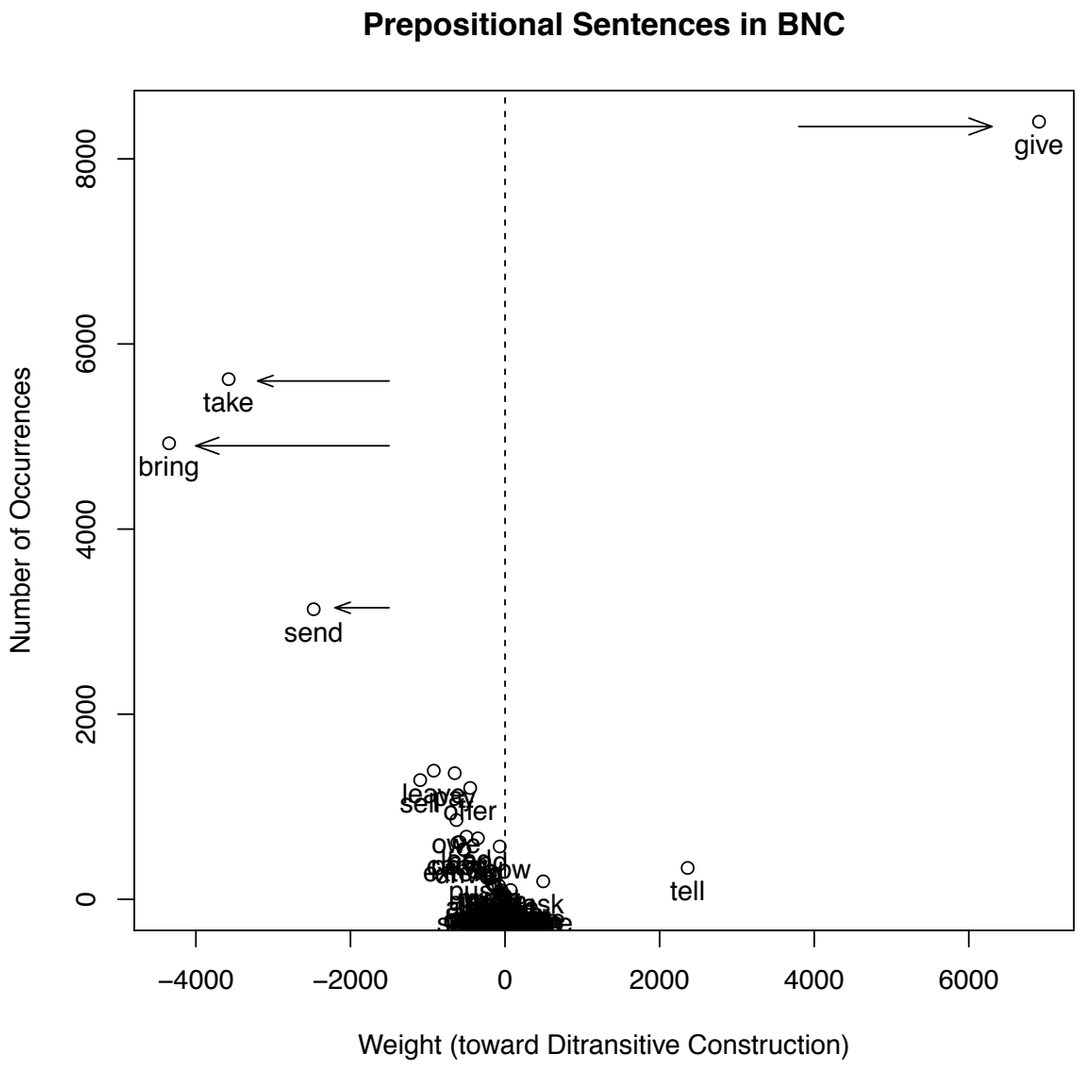
Does the VAH apply beyond the ditransitive frame?

- The VAH predicts that anchors to the prepositional frame should attract semantically similar verbs to the prepositional frame: *Bring*

- Anchors are not as strong as *give* for the ditransitive

How different *give* is!





Similarity to *bring* predicts selection of prepositional frame

- Our new model included as predictor similarity to *give*, similarity to *bring*, caused possession, pronominal theme, pronominal recipient, and difference in length between recipient and theme phrases

- All predictors are significant

- There is an interaction between semantic similarity to *bring* and *give*:

    - Semantic similarity to *bring* increases the likelihood of speakers choosing the prepositional frame only when the main verb is not highly similar to *give*

Beyond the ditransitive alternation

- Does the VAH apply beyond the ditransitive alternation?
- The locative alternation:
    1. Joann loaded the truck with hay
    2. Joann loaded hay onto the truck
- What is the anchor of the theme-object variant? *Put* (assuming anchors need not be alternating)

Put attracts verbs to the caused motion frame

- 45 verbs listed as alternating in Levin (1993)
- Not much study yet of other factors that could affect choice of caused motion frame so only one predictor included in the model (semantic similarity to *put*)

- Semantic similarity to *put* is a significant predictor of the occurrence of verbs in the locative PP frame

There is something unique about the ditransitive (Sun and Koenig 2017; Sun 2018)

	Child-directed speech	Child-directed speech	Child-directed speech	Adult-directed speech	Adult-directed speech	Adult-directed speech
quartiles	0,25	0,5	0,75	0,25	0,5	0,75
ditran	1	1	1	1	1	3
spray	1	2	3	2	4	7
caus	1	3	6	2	5	12
conative	1	3	4	1	2	4
passive	2	5	11	3	7	15

There is something unique about the ditransitive (Sun and Koenig 2017; Sun 2018)

Alternation	Verb type inventory size	Token Frequency
Ditransitive	17/34	6289/12428
Spray-load	5/19	175/224
Causative-inchoative	33/39	8159/8151
Conative	10/10	1158/3373
Active-Passive	45/77	112046/179705