Gradient blends of English PP verbal dependents

Najoung Kim[1], Kyle Rawlins[1] & Paul Smolensky[1,2]

Dept. of Cognitive Science, Johns Hopkins University[1], Microsoft Research AI[2]

(Acceptability judgments in current linguistic theory, 25 October 2018)



Complements and Adjuncts

Adjuncts (or modifiers) of “open”

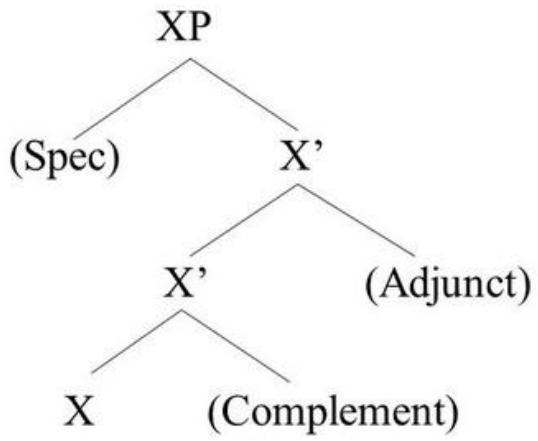
John opened the window this morning with Mary.

Argument (or complement) of “open”

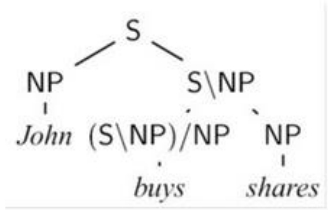
Arguments/Complements and Adjuncts/Modifiers = “Verbal dependents”

Complements and Adjuncts

- Important theoretical concepts in linguistics and also has psychological reality (Tutunjian and Boland, 2008)
- Often given distinct representations in many formalisms

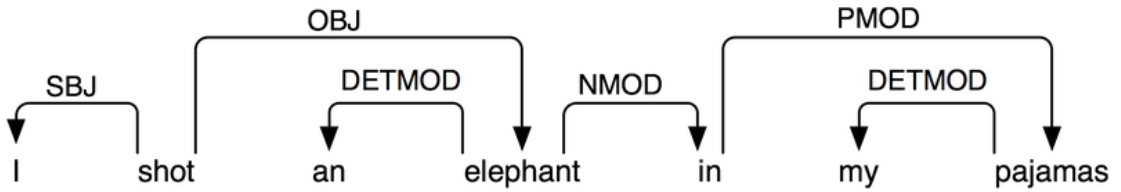


John ⊢ NP
shares ⊢ NP
buys ⊢ (S\NP)/NP
sleeps ⊢ S\NP
well ⊢ (S\NP)\(S\NP)



From Levy, Eisner & Klein

HOWEVER…



From Bird, Klein & Loper (2017)

Problematic dichotomy

- Simple argument (complement) vs adjunct dichotomy turns out not to be sufficient

Sue cut the bread with a knife >> Sue cut the bread with John, Sue cut the bread on Saturday
Tom walked to a park >> Tom walked with a friend

- Pustejovsky (1995): true argument, default argument, shadow argument, true adjunct
- Grimshaw and Vikner (1993): obligatory adjuncts

Schütze (1995): "argumenthood is not an all-or-nothing phenomenon, but [...] it comes in degrees"

Motivation

- Gradience in argumenthood is well-recognized (Donohue et al. 2004, Rissman, Rawlins & Landau 2015, Lewis 2004, *inter alia*), but not often formally described

- No systematic way to clearly separate arguments from adjuncts (Schütze 1995)

- Acceptability judgments on "Diagnostic tests" : (neither necessary nor sufficient)
  - Omissibility / obligatoriness: Is the expression acceptable without the verbal dependent under question? Is it necessary for the expression to be acceptable? (John threw (*the cookie) (out of the window))
  - Iterability: Can the same type of verbal dependent appear repeatedly? (John ran on Monday this week at 3pm…)
  - Separability: Can something intervene between the main verb and the verbal dependent?
  - Pseudo-clefting: Is is acceptable to pseudo-cleft the verbal dependent?
  - Latency, co-occurrence restrictions, case-marking, position restriction, etc… (Forker 2014)

  ***** Big problem: different tests can produce conflicting results *****

Desiderata based on observations

- We want..
  - **a model that accounts for gradient argumenthood**

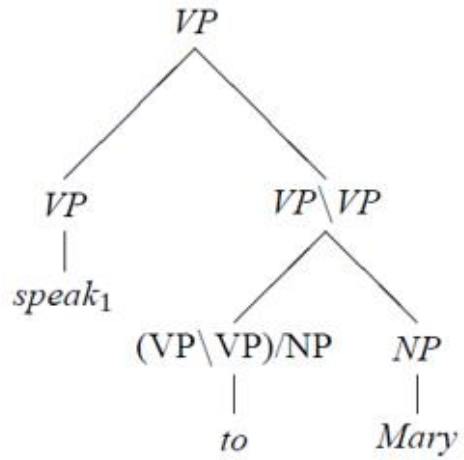
    Literature suggests that a dichotomy does not suffice.

  - **a coherent explanation for conflicting diagnostic tests**

    What are these tests doing? They are not everything (i.e., does not deterministically separate complements/adjuncts), but they must not be nothing (i.e., there is partial success---there must be reasons that these tests were proposed in the first place, and work for some cases)

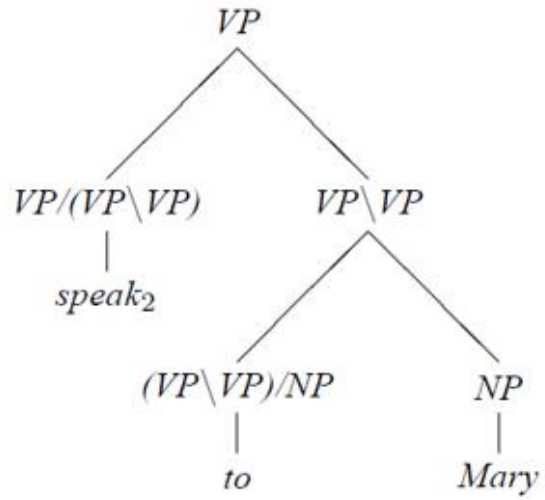
Dowty (2000)'s dual analysis

(10) a. **adjunct structure:**



Sem: $(to'(Mary))(speak'_1)$

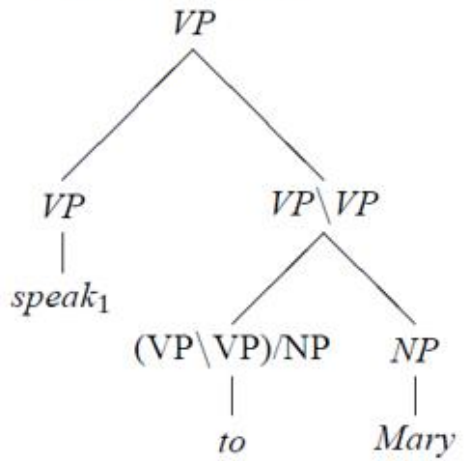
b. **complement reanalysis:**[9]



Sem: $speak'_2(to'(Mary'))$

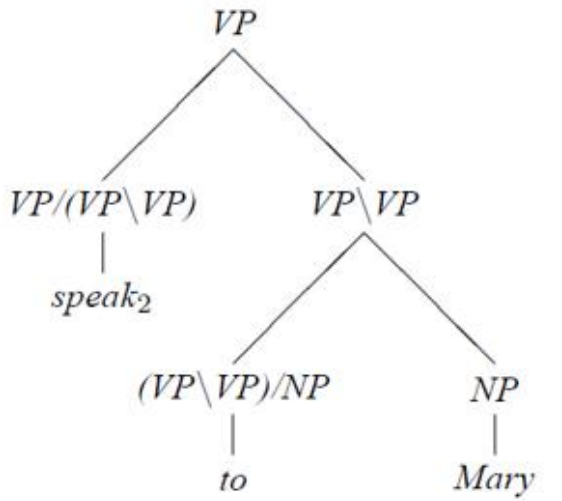
We are saying…

(10) a. **adjunct structure:**



Sem: $(to'(Mary))(speak'_1)$

b. **complement reanalysis:**[9]

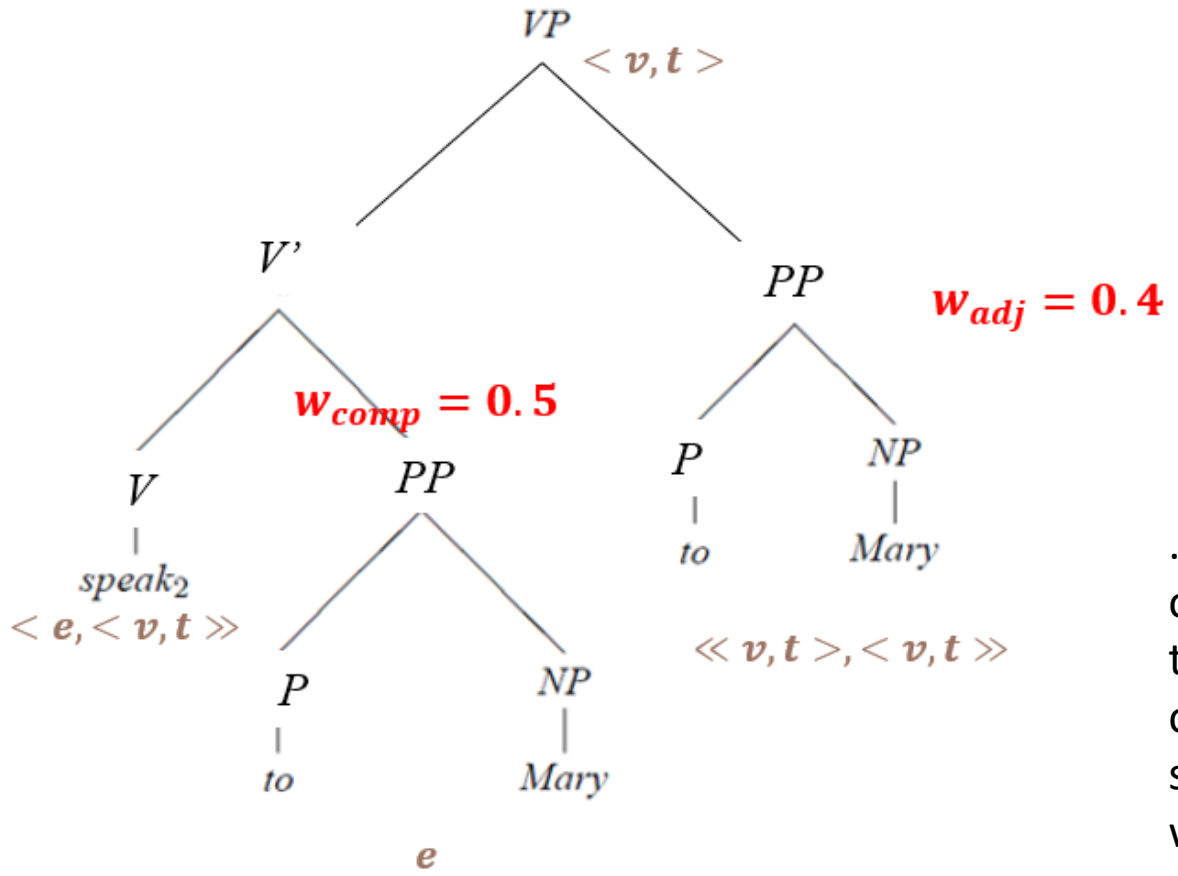


Sem: $speak'_2(to'(Mary'))$

They're not one or the other, but both at the same time!

“Weighted blend” analysis

Properties of complements and adjuncts derive from a **blend** of two partially (and simultaneously) active structures

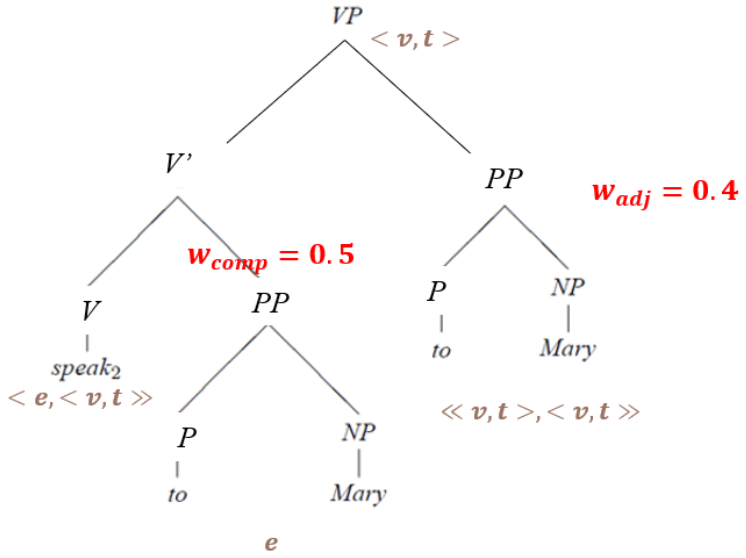


… in a conceptual sense that the idea of two related items (coindexing for traces, two parts of the PP blend here) occupying different positions at the same time, with potentially different weights, isn’t entirely novel

Not an entirely crazy idea—think of traces

Benefits

Clean representation of inherent lexical meaning of the preposition ("function" meaning) and idiosyncratic, predicate-dependent meaning ("scene role" meaning")



ORGROLE⇝LOCUS

(5) a. Vernon works **at** Grunnings.
b. Vernon works **for** Grunnings.

ORGROLE⇝BENEFICIARY

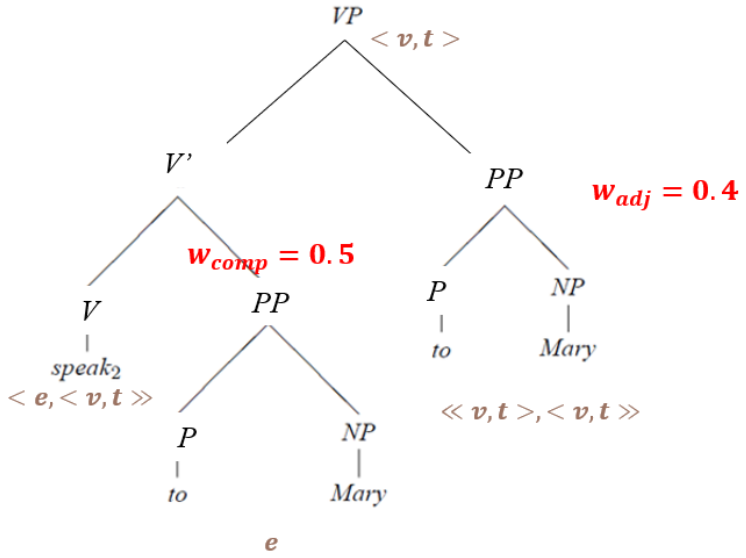
(6) Put it **on/by/behind/on_top_of/**... the door
GOAL⇝LOCUS

(Schneider et al. 2018)

$$\boldsymbol{\lambda e.\, speaking'(e)} \wedge \underbrace{\boldsymbol{Theme(e) = M'}}_{\text{Complement meaning; conjoined if } w_{comp} \geq 0.5} \wedge \underbrace{\boldsymbol{Goal(e) = M'}}_{\text{Adjunct meaning; conjoined if } w_{adj} \geq 0.3}$$

Benefits

Clean representation of inherent lexical meaning of the preposition ("function" meaning) and idiosyncratic, predicate-dependent meaning ("scene role" meaning")



ORGROLE⇝LOCUS

(5) a. Vernon works **at** Grunnings.
b. Vernon works **for** Grunnings.

ORGROLE⇝BENEFICIARY

(6) Put it **on/by/behind/on_top_of/**... the door.
GOAL⇝LOCUS

(Schneider et al. 2018)

$$\lambda x.\lambda e.\, verb\_meaning(e) \wedge v\_dependent\_meaning(x,e) \wedge v\_independent\_meaning(x,e)$$

Complement meaning; conjoined if $w_{comp} \geq n$

Adjunct meaning; conjoined if $w_{adj} \geq m$

Benefits

- Principled explanation for **conflicting diagnostic test acceptability**
- i.e., “What is happening when speakers are judging whether a test sentence is acceptable?”

Explanation by blend model: Different diagnostic tests use different optimization process, but the same weighted blend structures are used in all of them

Test 1 (omissibility)

	(a_C, a_A), OmitPP	FAITH-OMITPP	FAITH-a_C	H
a.	V PP (=*O): $a_C C + a_A A$	$-m_0(a_C + a_A)$		$-m_0(a_C + a_A)$
b.	V (=O): $a = 0$		$-m_1 a_C$	$-m_1 a_C$

Test 2 (pseudo-cleftability)

	(a_C, a_A), Pseudocleft	FAITH-PSUEDOCLEFT	FAITH-a_C	DO-PP$_A$	H
a.	V PP (=*P): $a_C C + a_A A$	$-k_0$			$-k_0$
b.	what X did PP [..] (=P): $a_A A$		$-k_1 a_C$	$k_2 a_A$	$-k_1 a_C + k_2 a_A$

Diagnostic tests

- We looked into some of these tests in more detail, especially focusing on the failure cases, in order to analyze what these diagnostics are probing

- Two diagnostics, **Pseudo-clefting** and **Omissibility** selected based on observations in the pilot study and also based on universal applicability

- Point of diagnostic tests: to find out whether something (here, the PP) is a complement or an adjunct

Pseudo-clefting test

Transform the original sentence into the form *What X did [PP] was […].*

Steve tossed the ball [for fun].
What Steve did [for fun] is toss the ball. ✓

(=> PP is an adjunct [P])

Steve strangled the victims [into a coma].
**What Steve did [into a coma] is strangle the victims.* ✗

(=> PP is a complement [*P])

Omissibility test

Remove the target PP from the original sentence and decide whether the remainder sounds grammatical/acceptable

If the PP is NOT omissible (*O), it is a complement.

I put the eggs [on the table].
**I put the eggs.*

Complement [*O]

John collaborated [with Paul].
?John collaborated.

Anna danced [with Elsa].
Anna danced.

Adjunct [O]

Conflicting case

*O or *P = Complement
O or P = Adjunct
(according to diagnostic test)

Steve pelted Anna [with acorns].
**Steve pelted Anna.*

Complement [*O]

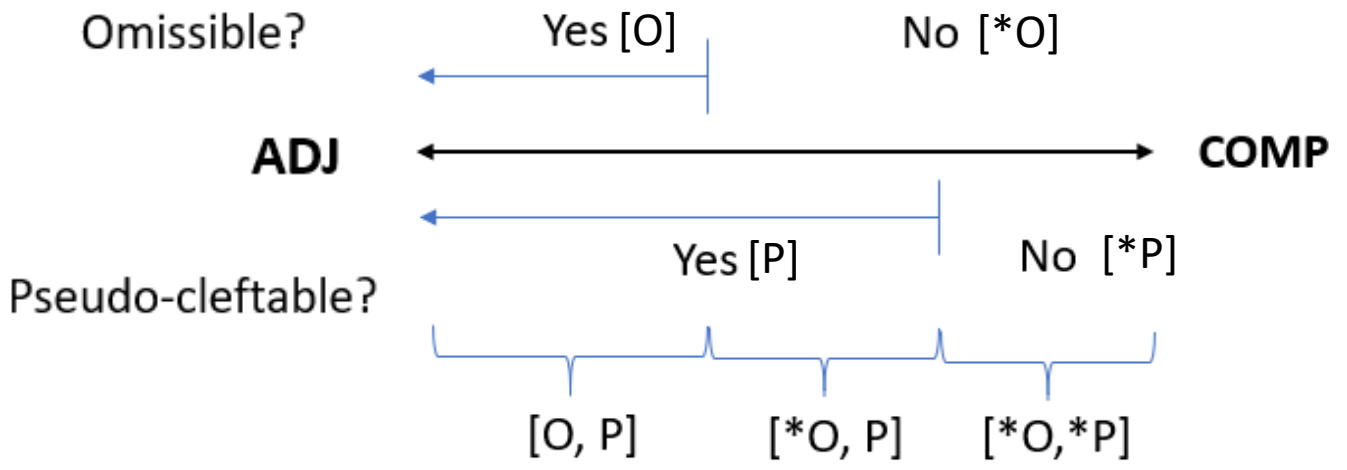
Steve pelted Anna [with acorns].
What Steve did [with acorns] was pelt Anna.

Adjunct [P]

Omissibility diagnostic says [with acorns] is a complement (it is not omissible; *O), but pseudo-clefting diagnostic says [with acorns] is an adjunct (it is pseudo-cleftable; P).

Should not happen if complement/adjunct are discrete categories! We need gradience (already knew this), but we claim this is insufficient.

Maybe a gradient Comp-Adj scale works..



Q: [O, *P]?

Steve strangled the victims into a coma.
Steve strangled the victims. **[O]**
*****What Steve did into a coma was strangle the victims.* **[*P]**

Steve tossed the ball [for fun].
Steve tossed the ball. **[O]**
What Steve did for fun was toss the ball. **[P]**

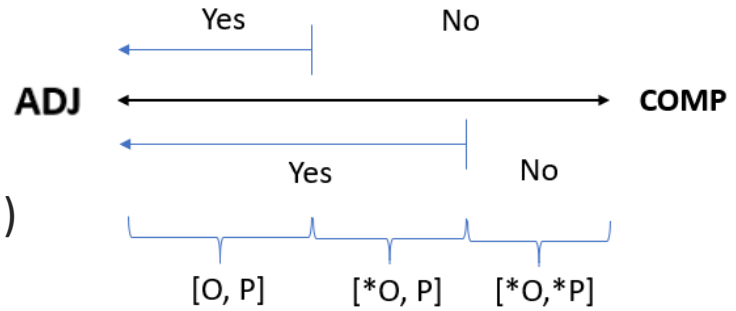
Steve pelted Anna [with acorns].
*****Steve pelted Anna.* **[*O]**
What Steve did [with acorns] was pelt Anna. **[P]**

Jackie chased [after the thief].
*****Jackie chased.* **[*O]**
*****What Jackie did [after the thief] was chase.* **[*P]**

The gradient (weighted) blend model comes in…

Diagnostics under gradient blend model

- With the gradience-only model, we speculated:
  - Different diagnostic tests are sensitive to different levels of complement-adjuncthood (or targets different points on the continuum)



- Under the gradient blend model, we hypothesize **in addition to** the previous assumption:
  - A diagnostic test can be sensitive to either the activation of the complement structure (a_C) or the activation of the adjunct structure (a_A)

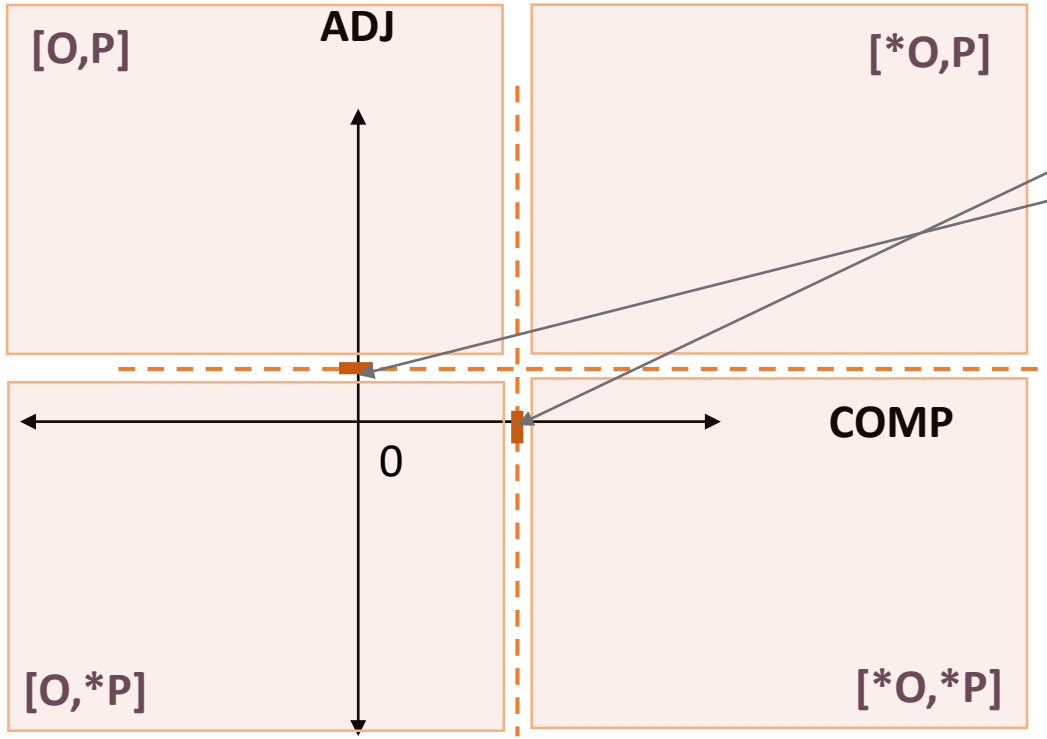
if a PP is not omissible, it is more complement-like

$$*\text{OMISSIBLE} \implies \text{high } a_C$$

$$\text{PSEUDO-CLEFTING} \implies \text{high } a_A$$

if a PP is pseudo-cleftable, it is more adjunct-like

Diagnostics under gradient blend model



$^*\text{OMISSIBLE} \implies \text{high } a_C$

$\text{PSEUDO-CLEFTING} \implies \text{high } a_A$

Justification

- Why might the two diagnostics be sensitive to different parts of the blend?

  [In bonus slides, ask if interested!]

Experimental verification: problems

- No gradient Comp/Adj judgment data available (need to collect data)
- How to elicit such judgments is unclear (need to develop collection procedure)

Experiments

For the same set of sentences:

1. **Collect gradient complement-adjunct judgments**

    1. Pilot: Discrete, relatively clear-cut complement-adjunct judgments
    (purpose: justifying centrality as a good proxy for linguists' CompAdj judgments)
    2. Main: Scaled comp-adj judgments

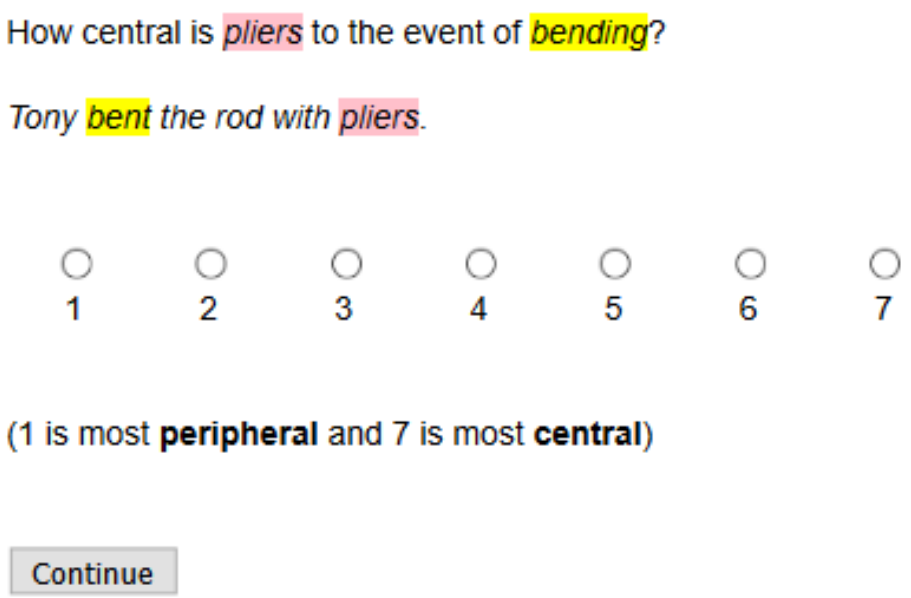
2. **Collect gradient acceptability judgments for diagnostic tests** (Omissibility and Pseudo-clefting)

    1. Linguist judgment (ternary pilot; what was reported in the abstract)
    2. Gradient acceptability judgment (scaled)

*** Judgments are EXPECTED to be noisy! PPs are rarely clear arguments, and dataset was designed to contain variable degrees of argumenthood ***

Collecting gradient Comp/Adj judgments

- Participants (non-linguists) were shown a single sentence per question
- Asked to pick a point on a 7-point Likert scale, according to how central they thought the highlighted NP under PP was, after a quick training question (2 sentences simultaneously given for training)



Stimuli

*** Judgments are EXPECTED to be noisy! PPs are rarely clear arguments, and dataset was designed to contain variable degrees of argumenthood ***

- 305 sentences containing at least one PP
- 120 unique verbs
- Mainly from VerbNet subcat frames and PropBank examples
- More adjunct-like examples added manually by PP substitution (intuition: subcat frame examples would be more complement-like)
- Varying degrees of complement- or adjunct-likeliness, at least according to the authors' intuition
- Superset of pilot stimuli

He withdrew [from the trip].

Claire took a train [to Colorado].

They participated [as a good gesture].

Nora pushed her [with the biggest smile].

Bill repaired the tractor [for a road trip].

I whipped the sugar [with cream].

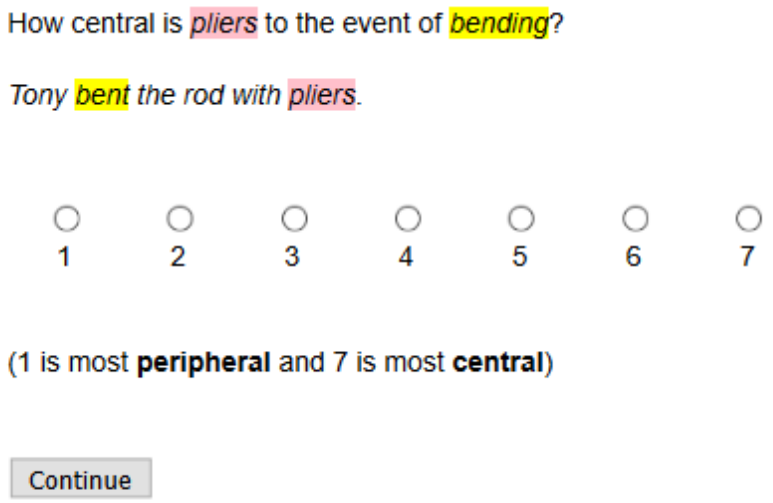
It clamped [on his ankle].

The witch turned him [into a frog] .

The children hid [in a hurry].

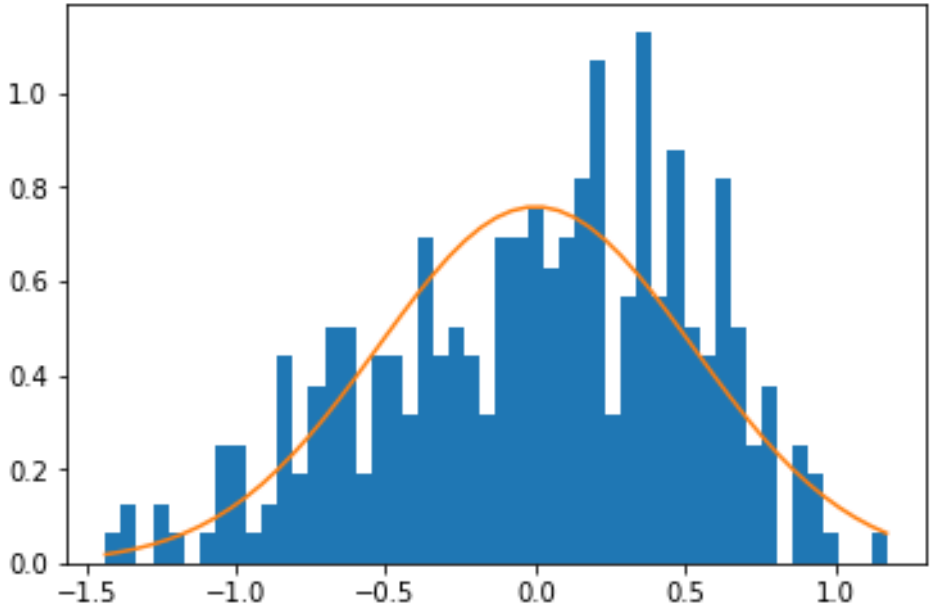
Procedure

- Conducted in subsets of around 50 sentences to reduce task load
  - Pilot with 77 questions took 38 minutes on average
- 305 sentences randomly split into 6 subsets
- 25 potentially overlapping participants for each subset
- Questions presented in random order



C-scores

- "C(entrality) score" of a PP given a V = mean of within-subject normalization using the mean and standard deviation of each individual participant
- $\mu = 1.967e-11, \sigma = 0.526, \text{range} = [-1.435, 1.172]$
- Slightly left-skewed / right-leaning $(z = -2.877, p < 0.01)$ distribution
- Interesting properties…



Alignment with thematic role hierarchy

Table 2: Centrality scores by thematic roles (mostly as annotated in VerbNet)

	Theme	Source	Topic	Instrument	Recipient	Trajectory	Result
μ	0.426	0.393	0.341	0.276	0.225	0.190	0.181
range	1.561	0.804	0.752	1.393	0.755	1.534	0.972

	Co-agent	Goal	Beneficiary	Location	Initial location	Manner	Time
μ	0.131	0.008	-0.133	-0.228	-0.333	-0.437	-0.565
range	1.296	1.323	1.648	2.302	1.231	1.427	1.820

- Agent > Benefactive/Goal > Theme > Location (Kiparsky 1987, Bresnan and Kanerva 1989)
- Agent > Goal/Experiencer/Location > Theme (Jackendoff 1972, Grimshaw 1990)
- **Agent > Theme > Goal/Benefactive/Location (Carrier-Duncan 1985, Larson 1988, Baker 1989) *****

Alignment with thematic role hierarchy

Table 2: Centrality scores by thematic roles (mostly as annotated in VerbNet)

	Theme	Source	Topic	Instrument	Recipient	Trajectory	Result
μ	0.426	0.393	0.341	0.276	0.225	0.190	0.181
range	1.561	0.804	0.752	1.393	0.755	1.534	0.972

	Co-agent	Goal	Beneficiary	Location	Initial location	Manner	Time
μ	0.131	0.008	-0.133	-0.228	-0.333	-0.437	-0.565
range	1.296	1.323	1.648	2.302	1.231	1.427	1.820

- Nevertheless, very large range for most roles (even location/manner/time, even though the means are at the lower end)
- Adds support to the previous finding that dependents more integral to the event display term-like properties and may not line up with the standard thematic hierarchy (Donohue et al. 2004; instrumentals)

Back to diagnostic tests…

- Recall that…

$$*\text{OMISSIBLE} \implies \text{high } a_C$$
$$\text{PSEUDO-CLEFTING} \implies \text{high } a_A$$

Reasonable hypothesis about the relation between C-scores and complement / adjunct activations (a_C, a_A)

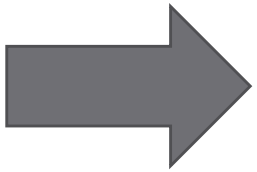
$$\boldsymbol{C} \propto w_1 * a_C + w_2 * a_A \; (\mathrm{w}_1 > 0, \mathrm{w}_2 < 0)$$

$$[*\text{OMISSIBLE}, *\text{PSEUDO-CLEFTING}] \implies \text{high centrality}$$
$$[\text{OMISSIBLE}, *\text{PSEUDO-CLEFTING}] \implies \text{moderate centrality}$$
$$[*\text{OMISSIBLE}, \text{PSEUDO-CLEFTING}] \implies \text{moderate centrality}$$
$$[\text{OMISSIBLE}, \text{PSEUDO-CLEFTING}] \implies \text{low centrality}$$

Magnitude of w1,w2 will affect “high” “moderate” and “low”. The two “moderate”s will likely differ, if w1 and w2 are not equal

Result

[*OMISSIBLE, *PSEUDO-CLEFTING] ⟹ high centrality
[OMISSIBLE, *PSEUDO-CLEFTING] ⟹ moderate centrality
[*OMISSIBLE, PSEUDO-CLEFTING] ⟹ moderate centrality
[OMISSIBLE, PSEUDO-CLEFTING] ⟹ low centrality



Pattern	μ
[*O, *P]	0.294
[O, *P]	0.007†
[*O, P]	-0.069**
[O, P]	-0.250***

- Significant effect of pattern group on mean C-scores
- Alignment of means equivalent to our predictions

Revised predictions

- Recall that…

$$*\text{OMISSIBLE} \implies \text{high } a_C$$
$$\text{PSEUDO-CLEFTING} \implies \text{high } a_A$$

We said "moderate"-ness will probably differ depending on w1 and w2.
The results suggest $|w_1| > |w_2|$, which means…

Reasonable hypothesis about the relation between C-scores and complement / adjunct activations (a_C, a_A)

$$\boldsymbol{C} \propto w_1 * a_C + w_2 * a_A \; (\mathrm{w}_1 > 0, \mathrm{w}_2 < 0)$$

$$[*\text{OMISSIBLE}, *\text{PSEUDO-CLEFTING}] \implies \text{high centrality}$$
$$[\text{OMISSIBLE}, *\text{PSEUDO-CLEFTING}] \implies \text{Moderate-to-high}$$
$$[*\text{OMISSIBLE}, \text{PSEUDO-CLEFTING}] \implies \text{Moderate-to-low}$$
$$[\text{OMISSIBLE}, \text{PSEUDO-CLEFTING}] \implies \text{low centrality}$$

Revised predictions

- Recall that…

$$*\text{OMISSIBLE} \implies \text{high } a_C$$

$$\text{PSEUDO-CLEFTING} \implies \text{high } a_A$$

Reasonable hypothesis about the relation between C-scores and complement / adjunct activations (a_C, a_A)

$$\boldsymbol{C} \propto w_1 * a_C + w_2 * a_A \; (\mathrm{w}_1 > 0, \mathrm{w}_2 < 0)$$

In-words-summary: higher complement activation (a_C) makes a big positive contribution to C-scores, and higher adjunct activation (a_A) makes a small negative contribution to C-scores. ($|w_1| > |w_2|$)

How are a_C and a_A are computed exactly?

(Gradient complement and adjunct activations in the output)

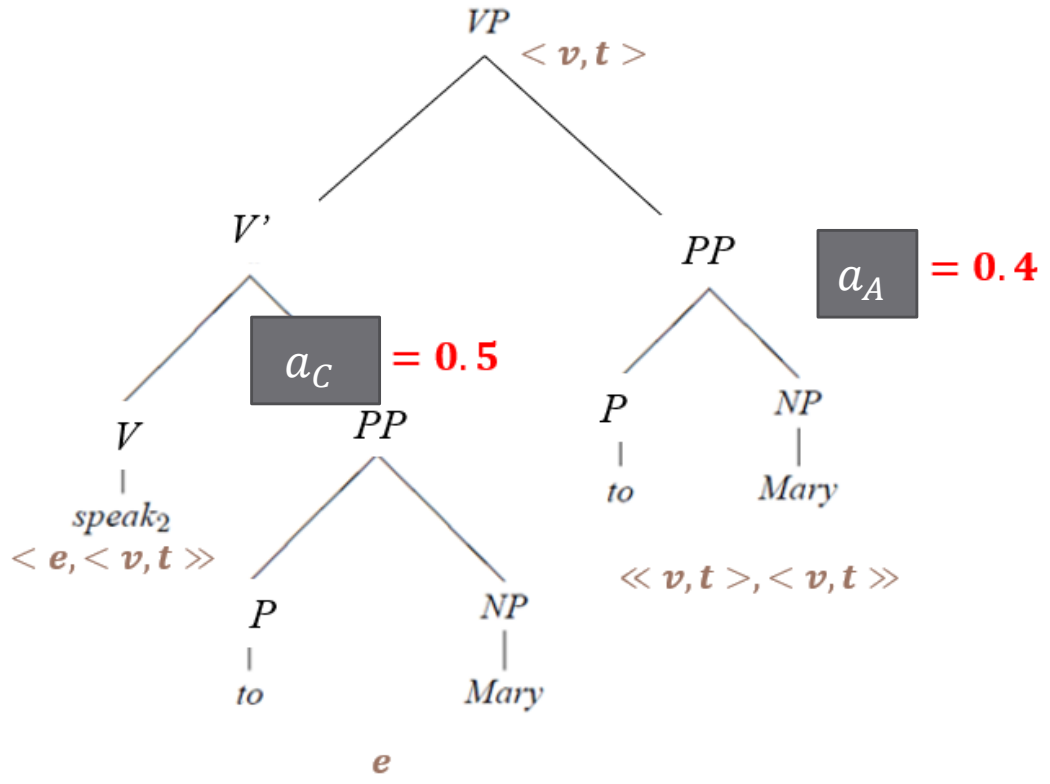
Gradient Symbolic Computation (GSC) framework

Smolensky, Goldri[illegible] & Mathis (2014): Computational model of cognition

- **Optimization** u[illegible] to OT/Harmonic Grammar
- **Quantization** d[illegible]

GSC formu[illegible]

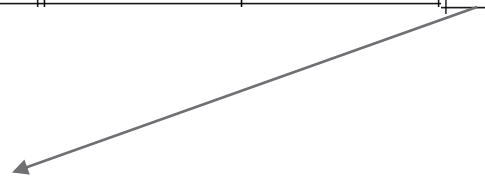
"Find[illegible] ize Harmony"



		$\alpha \implies a_A$	**H**
a.	OC (Some 'V PP'	$\alpha \cdot a_A$	$\gamma \cdot a_C + \alpha \cdot a_A$

Optimization with GSC (w/ quantization)

(γ, α), V, PP	$\gamma \implies a_C$	$\alpha \implies a_A$	**H**
a. OC (Some 'V PP' is optimal): $\{a_C \geq 0, a_A > 0\}$	$\gamma \cdot a_C$	$\alpha \cdot a_A$	$\gamma \cdot a_C + \alpha \cdot a_A$



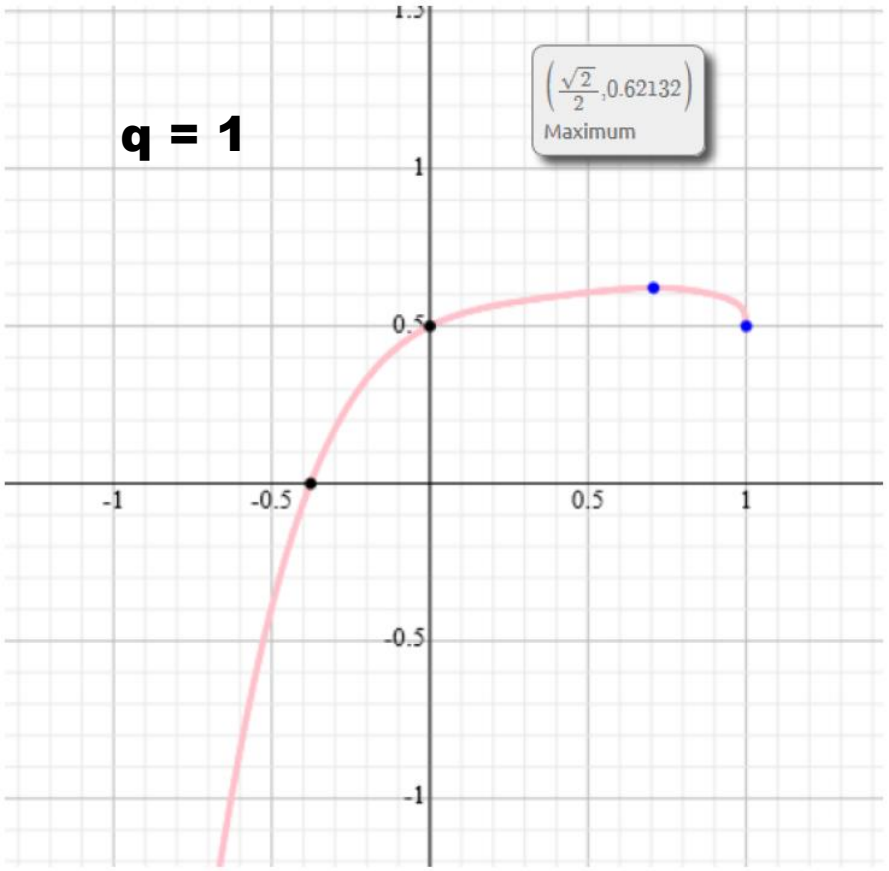
$$Harmony\ (H) = \gamma \cdot a_C + \alpha \cdot a_A$$

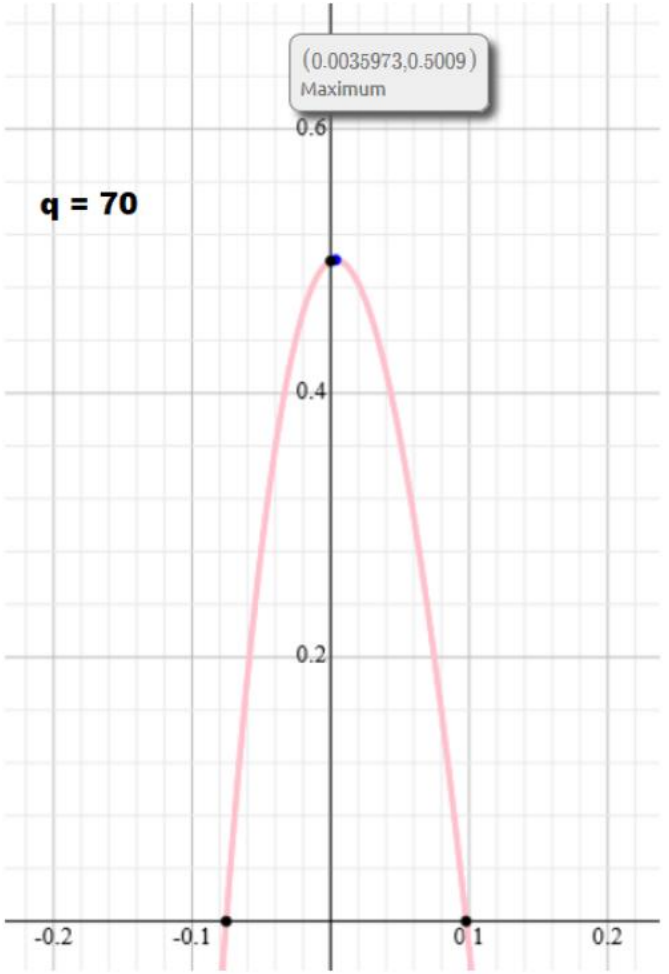
$$Quantization\ constraint\ (Q) = -q \cdot [k \cdot (a_C^2 + a_A^2\ - 1)^2 + (a_C^2(1 - a_C)^2 + a_A^2(1 - a_A)^2)\ ]$$

Optimal $\{a_C, a_A\}$ are s.t. maximize $H + Q$!!

Example case

For $(\gamma, \alpha) = (0.5,0.5)$, we always get gradient optima between $0 \leq q < 80$ (nice, because of Cho & Smolensky 2016)





Optimization with diagnostic sentences

Diagnostic tests use what we just computed—the blend weights (a_C, a_A)-- as inputs to separate, test-dependent optimization processes
(conscious comparison of a certain diagnostic construction ("b") with respect to the original sentence with ("a") the weights (a_C, a_A))

Test 1 (omissibility)

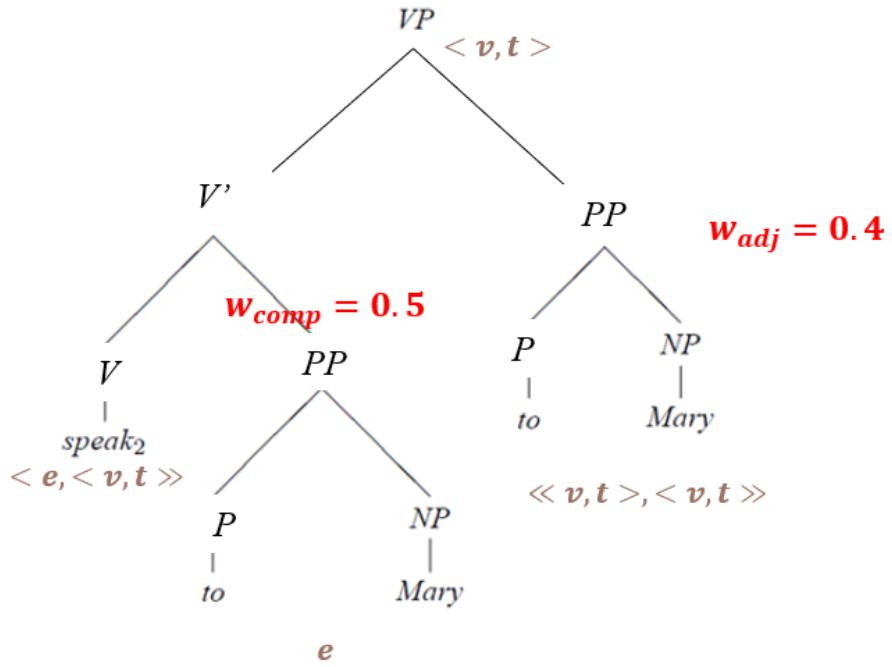
(a_C, a_A), OmitPP		FAITH-OMITPP	FAITH-a_C	H
a.	V PP (=*O): $a_C C + a_A A$	$-m_0(a_C + a_A)$		$-m_0(a_C + a_A)$
b.	V (=O): $a = 0$		$-m_1 a_C$	$-m_1 a_C$

Test 2 (pseudo-cleftability)

(a_C, a_A), Pseudocleft		FAITH-PSUEDOCLEFT	FAITH-a_C	DO-PP$_A$	H
a.	V PP (=*P): $a_C C + a_A A$	$-k_0$			$-k_0$
b.	what X did PP [..] (=P): $a_A A$		$-k_1 a_C$	$k_2 a_A$	$-k_1 a_C + k_2 a_A$

Summary

- Proposed an analysis of English {V, PP} constructions as gradient blends (weighted partial activations of both Comp and Adj structures)



Summary

- Explained *why* conflicting diagnostic tests are partially and only partially successful
  - Why are they sometimes informative? (because they involve optimizing for the same target structure $\{a_C, a_A\}$ with the same lexical inputs {V, PP})
  - Why do they sometimes conflict? (because different diagnostic tests are different optimization processes, although they make use of some common variables. Exploiting different structural environments necessarily evoke additional constraints)

(γ, α), V, PP	$\gamma \Longrightarrow a_C$	$\alpha \Longrightarrow a_A$	**H**
a. OC (Some 'V PP' is optimal): $\{a_C \geq 0, a_A > 0\}$	$\gamma \cdot a_C$	$\alpha \cdot a_A$	$\gamma \cdot a_C + \alpha \cdot a_A$

Test 1 (omissibility)

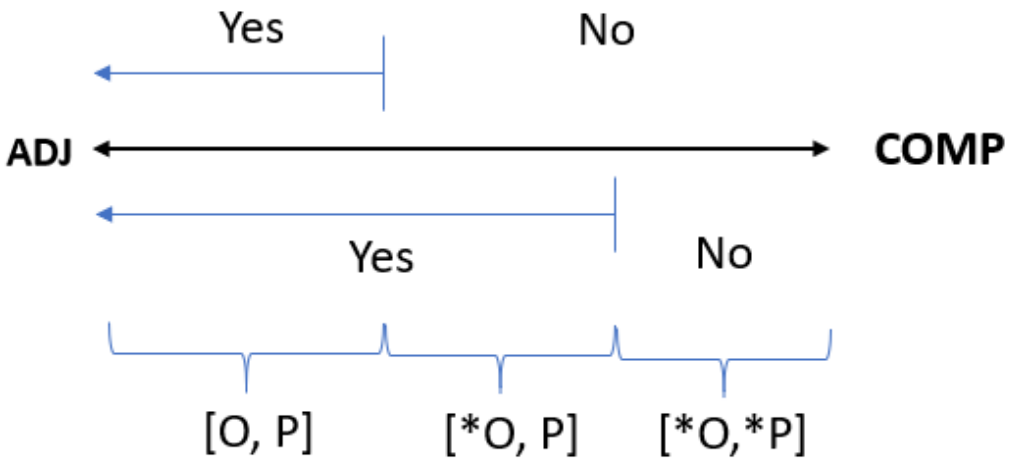
(a_C, a_A), OmitPP	FAITH-OMITPP	FAITH-a_C	**H**
a. V PP (=*O): $a_C C + a_A A$	$-m_0(a_C + a_A)$		$-m_0(a_C + a_A)$
b. V (=O): $a = 0$		$-m_1 a_C$	$-m_1 a_C$

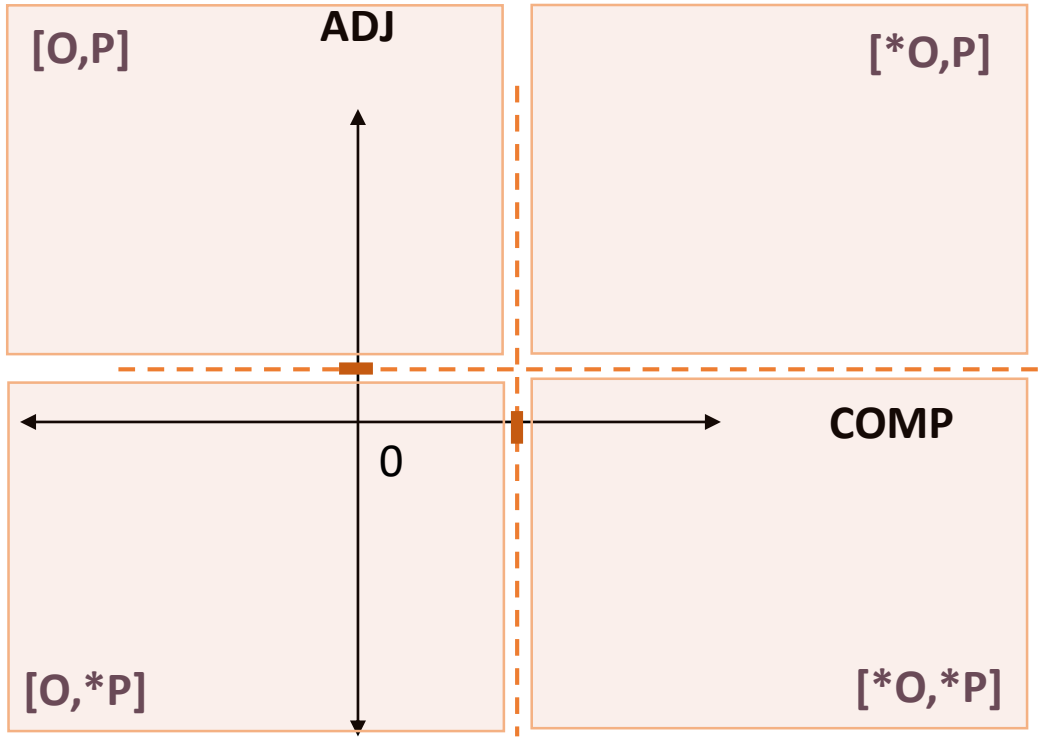
Test 2 (pseudo-cleftability)

(a_C, a_A), Pseudocleft	FAITH-PSUEDOCLEFT	FAITH-a_C	DO-PP$_A$	**H**
a. V PP (=*P): $a_C C + a_A A$	$-k_0$			$-k_0$
b. what X did PP [..] (=P): $a_A A$		$-k_1 a_C$	$k_2 a_A$	$-k_1 a_C + k_2 a_A$

Summary

- Provided experimental evidence for blended representations (dichotomous or gradience-only models do not cover the observed diagnostic pattern typology fully)





References

Barbu, R. M., & Toivonen, I. (2016). Event participants and linguistic arguments. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Cho, P. W., & Smolensky, P. (2016). Bifurcation analysis of a Gradient Symbolic Computation model of incremental processing. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society. Cognitive Science Society, Austin, TX*.

Donohue, C., Donohue, M., Butt, M., & King, T. H. (2004). On the special status of instrumentals. In *Proceedings of the LFG04 Conference, On–line proceedings. Stanford, CA: CSLI Publications*.

Grimshaw, J., & Vikner, S. (1993). Obligatory adjuncts and the structure of events. In *Knowledge and language* (pp. 143-155). Springer, Dordrecht.

Lewis, H. (2004). The with-phrase theme in English: Argument or adjunct". PhD thesis. University of Canterbury.

Manning, C. D. (2003). “Probabilistic syntax". In: Probabilistic linguistics, pp. 289-341.

Pollard, C. and Sag, I. (1987). Information-Based Syntax and Semantics, Vol. 1. CSLI Publications.

References

Pustejovsky, J. (1995). The Generative Lexicon. MIT Press.

Schütze, C. T. (1995). PP attachment and argumenthood. *MIT working papers in linguistics*, *26*(95), 151.

Rissman, L., Rawlins, K., & Landau, B. (2015). Using instruments to understand argument structure: Evidence for gradient representation. *Cognition*, *142*, 266-290.

Schneider et al. (2018). Comprehensive Supersense Disambiguation of English Prepositions and Possessives. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 185—196.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cognitive science*, *38*(6), 1102-1138.

Tutunjian, D., & Boland, J. E. (2008). Do we need a distinction between arguments and adjuncts? Evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass*, *2*(4), 631-646.

Thanks!
Questions?

Slides will be available at https://najoungkim.github.io

Look out for our very relevant paper on Nov 1st 🤞 (but it is already on arXiv!)

Authors



Jane Lutken
(our syntactician)

Ben Van Durme
(very relevant collaborator)

Funding: JHU CogSci & NSF INSPIRE #1344269

JOHNS HOPKINS
KRIEGER SCHOOL
of ARTS & SCIENCES

NSF