Prediction and priming across levels: Fall 2022

Brian Dillon
Shota Momma

University of Massachusetts, Amherst
Department of Linguistics

4/26/2022

Why care?

— **P**rocessing involves **P**rediction,
— **P**rediction is **P**roduction,
— **P**rediction leads to **P**rediction error,
— **P**rediction error creates **P**riming,
— **P**riming is im**P**licit learning,
— im**P**licit learning is the mechanism for acquisition/adaptation of **P**rocessing, **P**rediction and **P**roduction, and
— **P**roduction provides the input for training **P**rocessing.

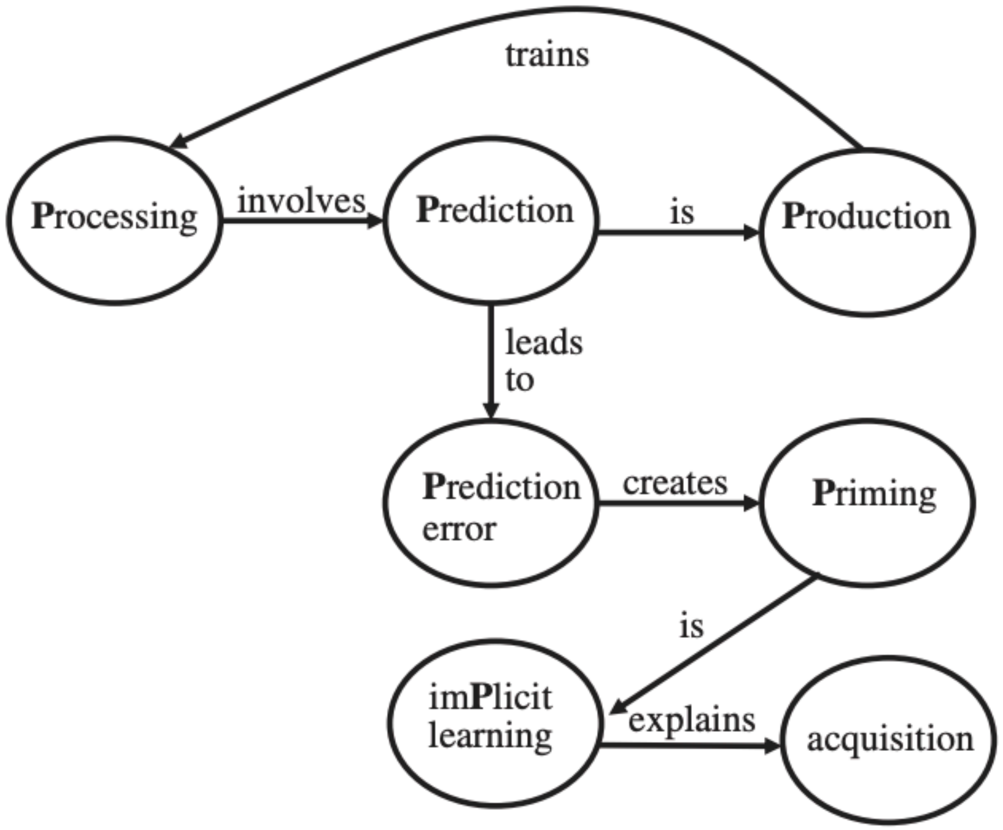


Figure 1. The P-chain framework for psycholinguistics.

An integrated theory of comprehension, production and acquisition (with prediction as the central component)?

Why care?

What do RNN Language Models Learn about Filler–Gap Dependencies?

Ethan Wilcox, Roger Levy, Takashi Morita, Richard Futrell

Mechanisms for handling nested dependencies in neural-network language models and humans☆

Yair Lakretz [a,*], Dieuwke Hupkes [c], Alessandra Vergallito [b], Marco Marelli [b], Marco Baroni [c,d,1], Stanislas Dehaene [a,e,1]

Neural Language Models Capture Some, But Not All, Agreement Attraction Effects

Suhas Arehalli
Johns Hopkins University

Tal Linzen
Johns Hopkins University

Accounting for Agreement Phenomena in Sentence Comprehension with Transformer Language Models: Effects of Similarity-based Interference on Surprisal and Attention

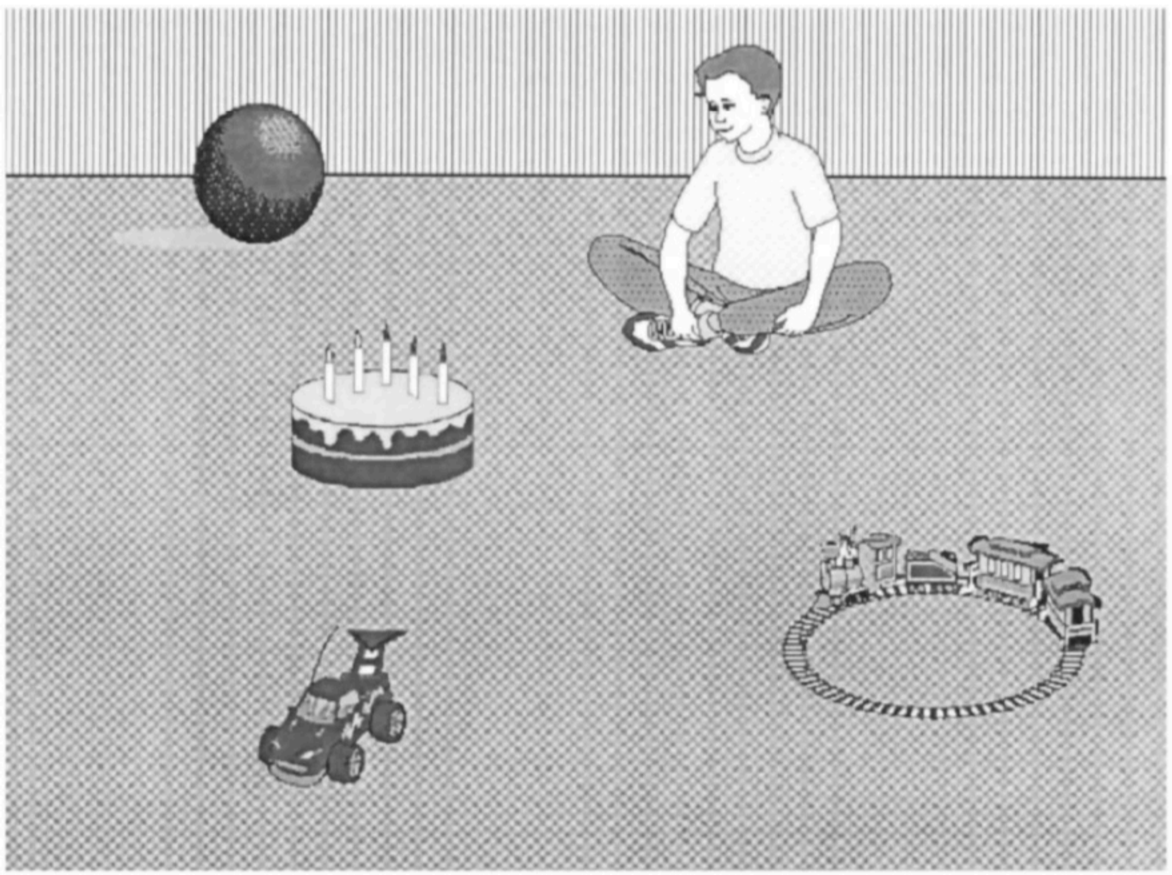
Soo Hyun Ryu, Richard L. Lewis

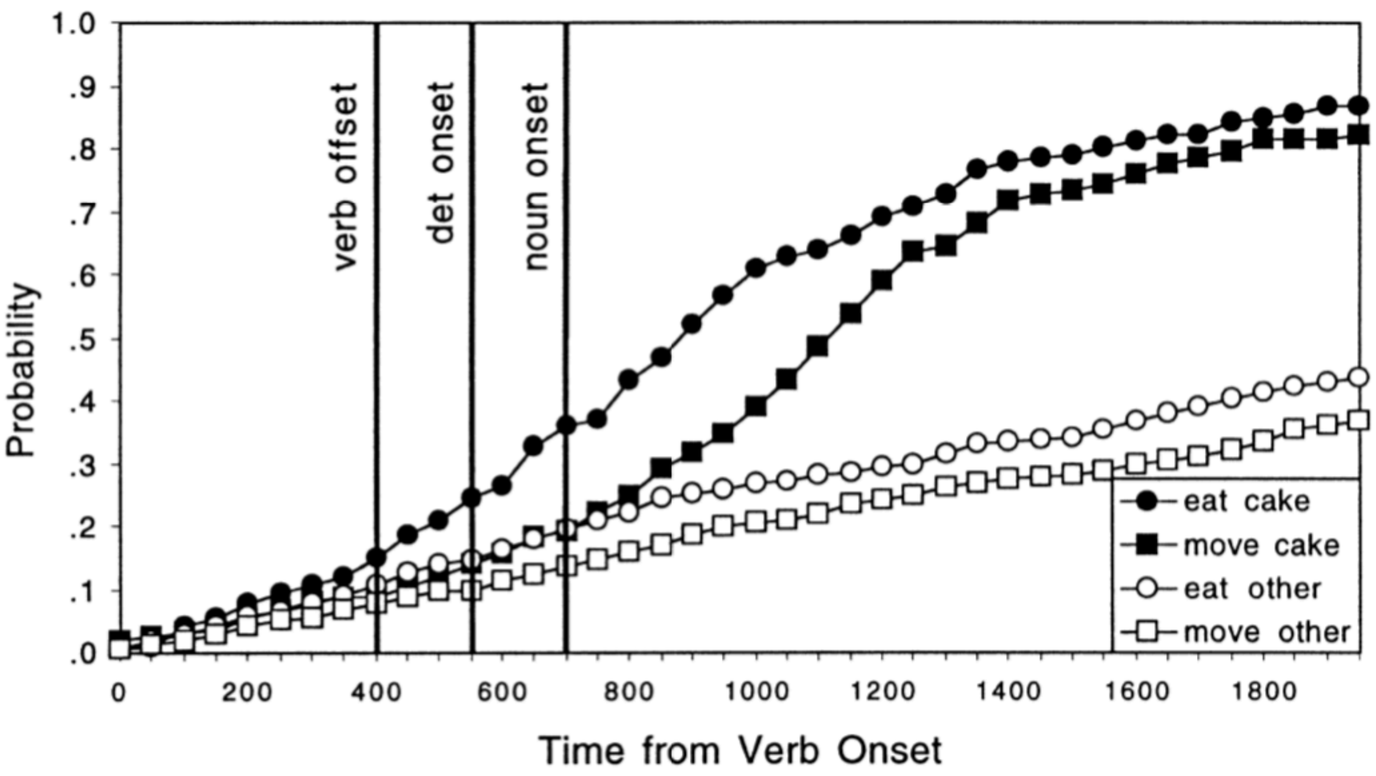
Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies

Tal Linzen[1,2] **Emmanuel Dupoux**[1]
LSCP[1] & IJN[2], CNRS,
EHESS and ENS, PSL Research University
{tal.linzen,
emmanuel.dupoux}@ens.fr

Yoav Goldberg
Computer Science Department
Bar Ilan University
yoav.goldberg@gmail.com

Evidence for prediction

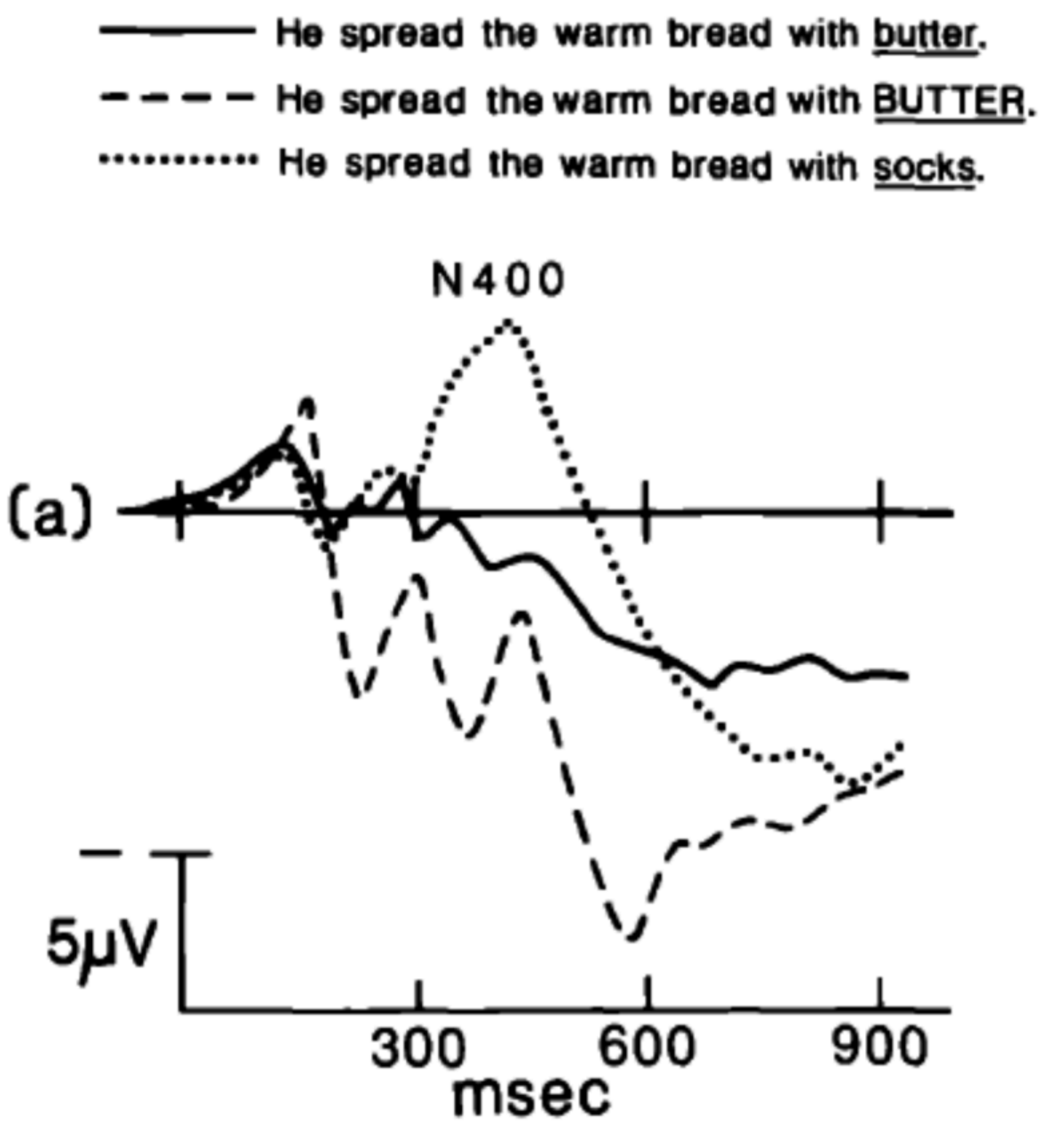




Altmann & Kamide (1999)

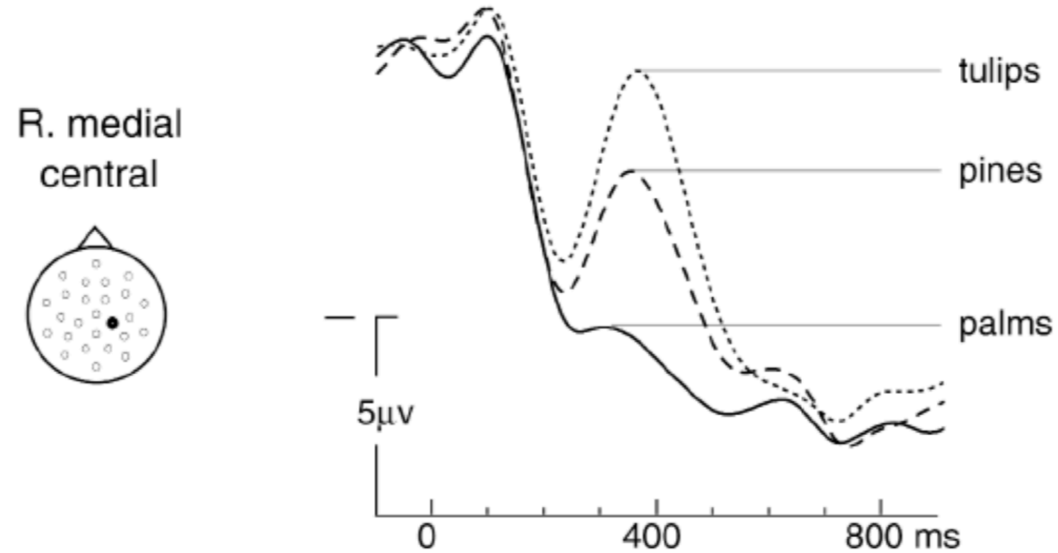
N400 and prediction

ERPs



"They wanted to make the hotel look more like a tropical resort.

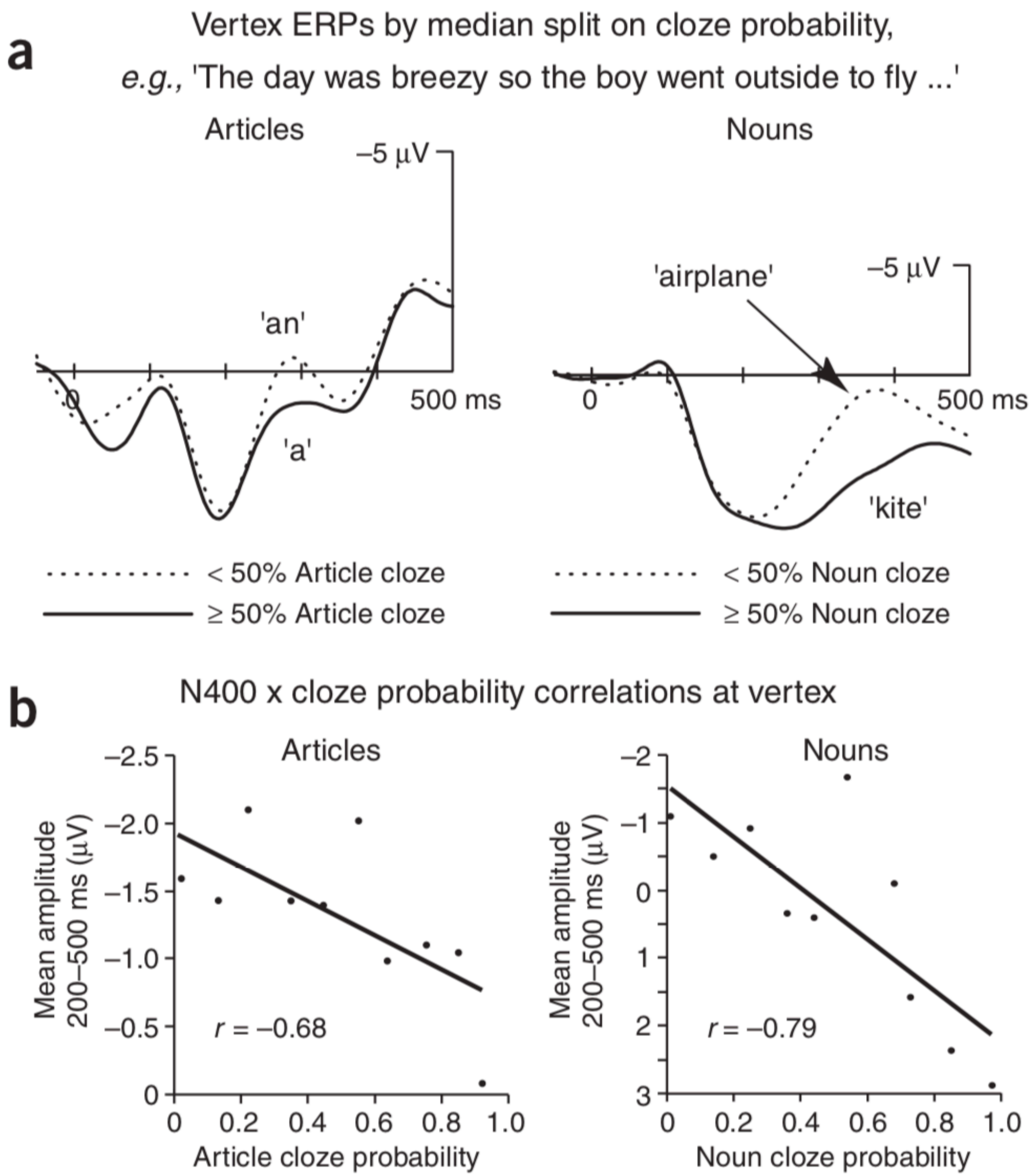
So along the driveway they planted rows of _______ ."



Kutas & Hillyard (1983)

Federmeier & Kutas (1999)

Evidence for prediction



De Long et al . (2005)

Parallel architecture: a claim

Prediction via preactivation (across all levels) is a by-product of lexical access.

Lexical access is not just about accessing words, but it's about accessing any (linguistic) item in long term memory: Extended lexicon

Cat		-s		Cats	
Semantics:	$[CAT_1]$	Semantics:	$[PLUR\ (X_x)]_y$	Semantics:	$[PLUR\ (CAT_1)]_7$
Syntax:	N_1	Morphosyntax:	$[N_x\ PLUR_6\ ]_y$	Morphosyntax:	$[N_1\ PLUR_6\ ]_7$
Phonology:	$/kæt_1/$	Phonology:	$/\ \ldots_x\ s_6\ /_y$	Phonology:	$/kæt_1\ s_6\ /_7$

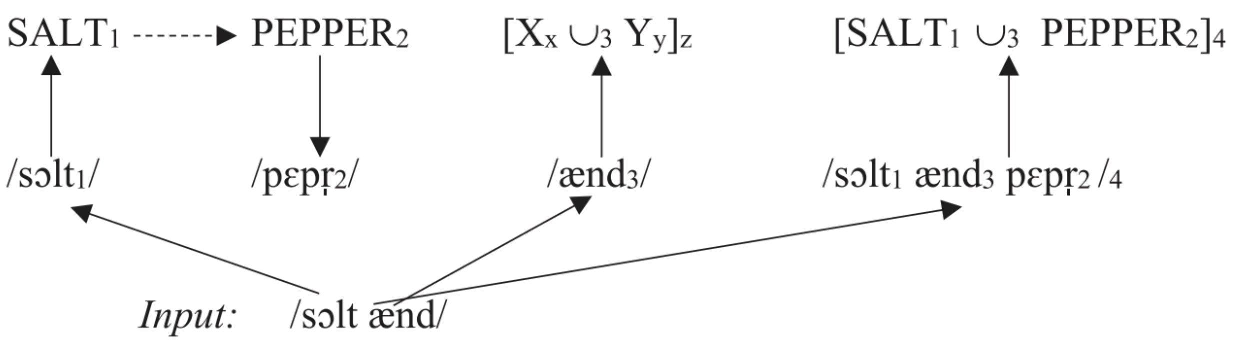
Priming and prediction both arise from activation spreading in extended lexicon

Extended lexicon and prediction

Syntactic schema

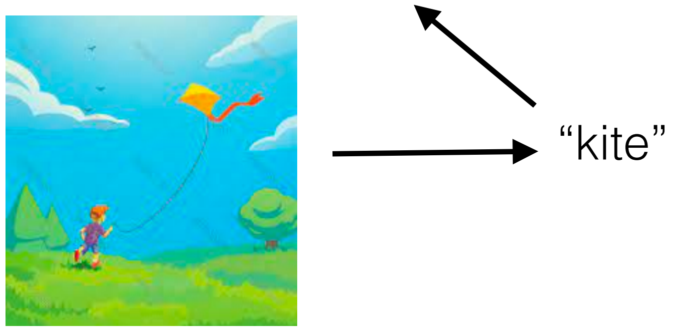
Syntax: [$_{NP}$ Det <A> N …]

Collocation



Phonology

Semantics:	a. INDEF$_{12}$	b. INDEF$_{13}$
Syntax:	Det$_{12}$	Det$_{13}$
Phonology:	/ə$_{12}$ C/	/ən$_{13}$ V/



Discourse/
world knowledge

“We do not explicitly formalize this here, but we do assume that discourse event and world knowledge, and visual and spatial information can prime lexical items, contingent on the contextual situation.”

Case study 1: syntactic prediction

Left corner parsing

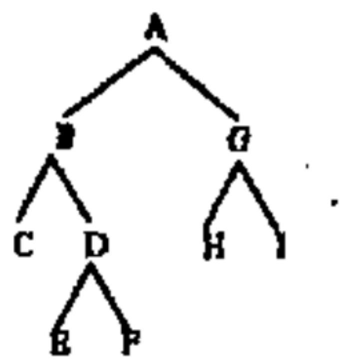


Figure 1: *A parse tree*

top-down:
ABCDEFGHI

bottom-up:
CEFDBHIGA

left-corner:
CBEDFAHGI

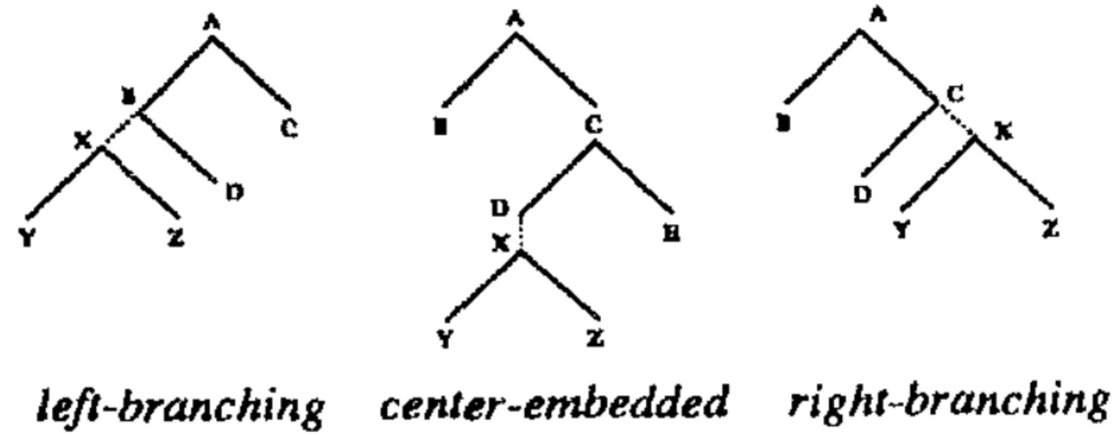


Figure 2: *Branching structures*

Left-branching (head-final): Easy

Center-embedded: Hard

Right-branching (head-initial): Easy

Strategy		Space required		
Nodes	*Arcs*	*Left*	*Center*	*Right*
Top-down	either	$O(n)$	$O(n)$	$O(1)$
Bottom-up	either	$O(1)$	$O(n)$	$O(n)$
Left-corner	standard	$O(1)$	$O(n)$	$O(n)$
Left-corner	eager	$O(1)$	$O(n)$	$O(1)$
What people do		$O(1)$	$O(n)$	$O(1)$

Resnik (1992)
Chomsky & Miller (1961)

PCFG

Garden-path

1.0	S	→	NP VP .
0.876404494831	NP	→	DT NN
0.123595505169	NP	→	NP VP
1.0	PP	→	IN NP
0.171428571172	VP	→	VBD PP
0.752380952552	VP	→	VBN PP
0.0761904762759	VP	→	VBD
1.0	DT	→	*the*
0.5	NN	→	*horse*
0.5	NN	→	*barn*
0.5	VBD	→	*fell*
0.5	VBD	→	*raced*
1.0	VBN	→	*raced*
1.0	IN	→	*past*

SRC vs. ORC

0.33	NP	→	SPECNP NBAR
0.33	NP	→	*you*
0.33	NP	→	*me*
1.0	SPECNP	→	DT
0.5	NBAR	→	NBAR S[+R]
0.5	NBAR	→	N
1.0	S	→	NP VP
0.86864638	S[+R]	→	NP[+R] VP
0.13135362	S[+R]	→	NP[+R] S/NP
1.0	S/NP	→	NP VP/NP
1.0	VP/NP	→	V NP/NP
1.0	VP	→	V NP
1.0	V	→	*saw*
1.0	NP[+R]	→	*who*
1.0	DT	→	*the*
1.0	N	→	*man*
1.0	NP/NP	→	ϵ

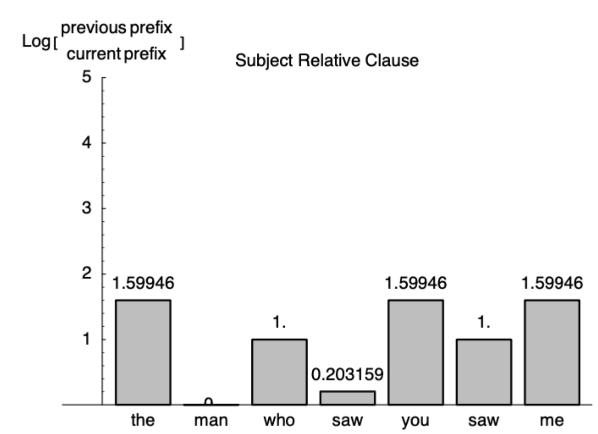


Figure 7: Subject relative claus

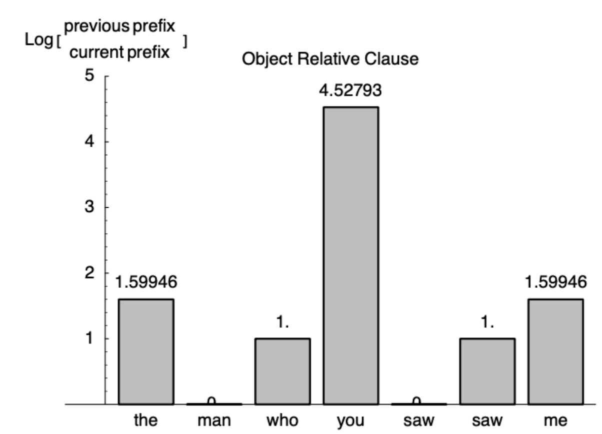
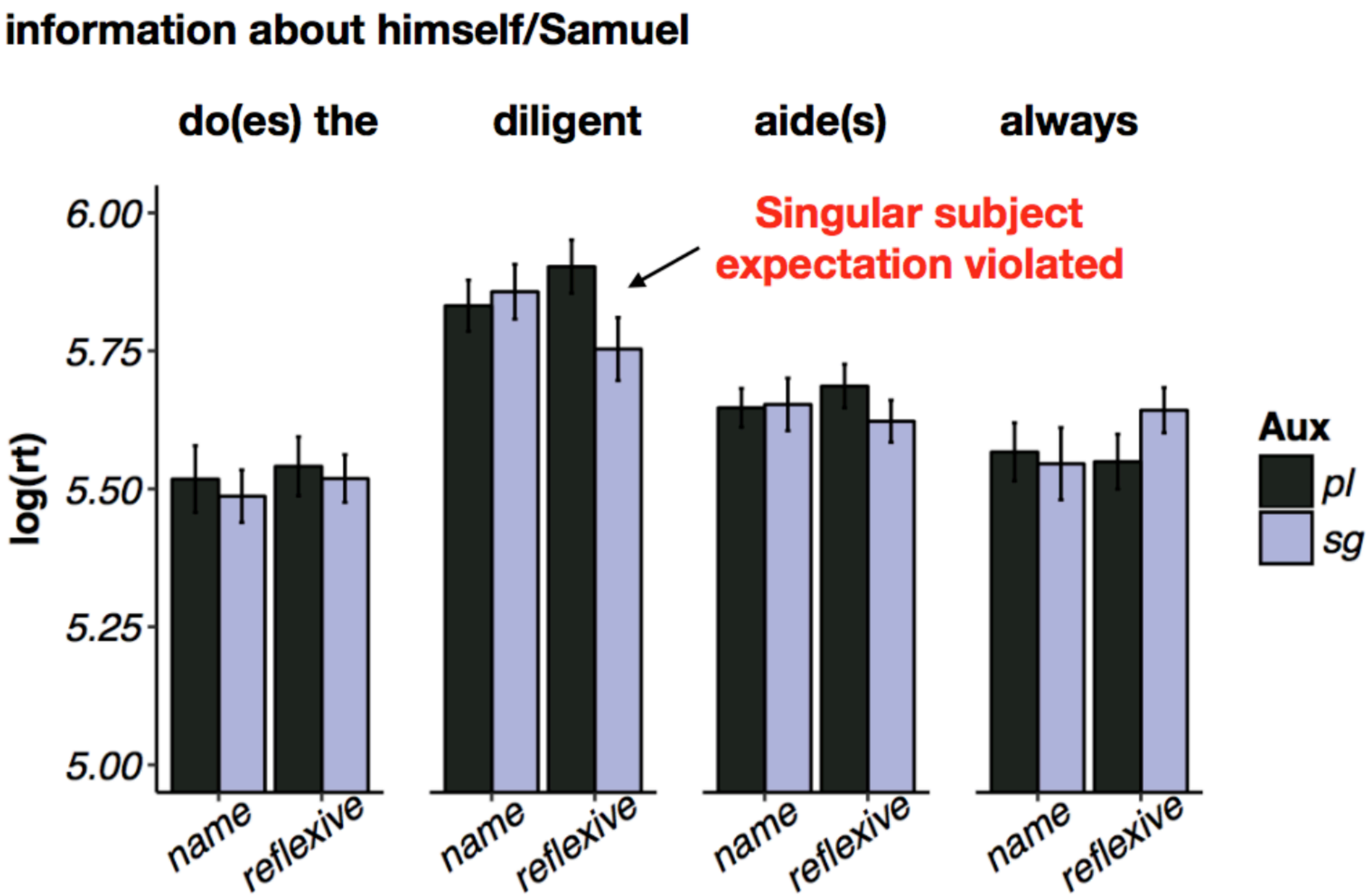


Figure 8: Object relative clause

Hale (2001)

Syntactic prediction as treelet activations?

[$_{NP}$ Which information about {***himself*** / ***Samuel***}] {**do** / **does**} the deligent aide(s) always remind the researcher about?

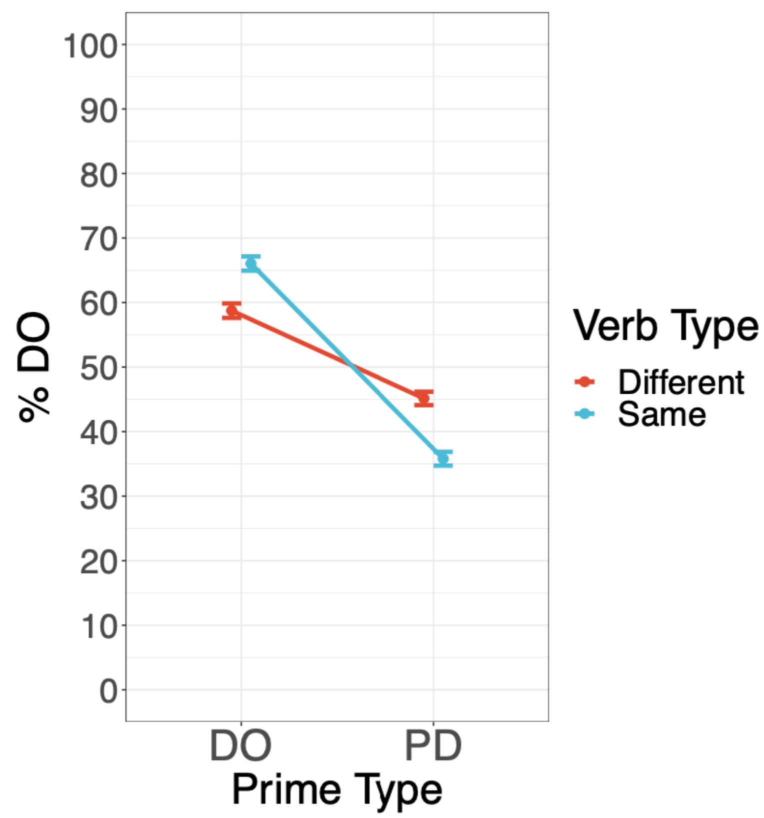


Yoshida et al. (in prep)

Syntactic prediction & syntactic priming

Syntactic priming is also reflecting heightened activation of treelet/syntactic schema?
Prediction and priming arise from the same cause (activated lexical item in the extended lexicon)

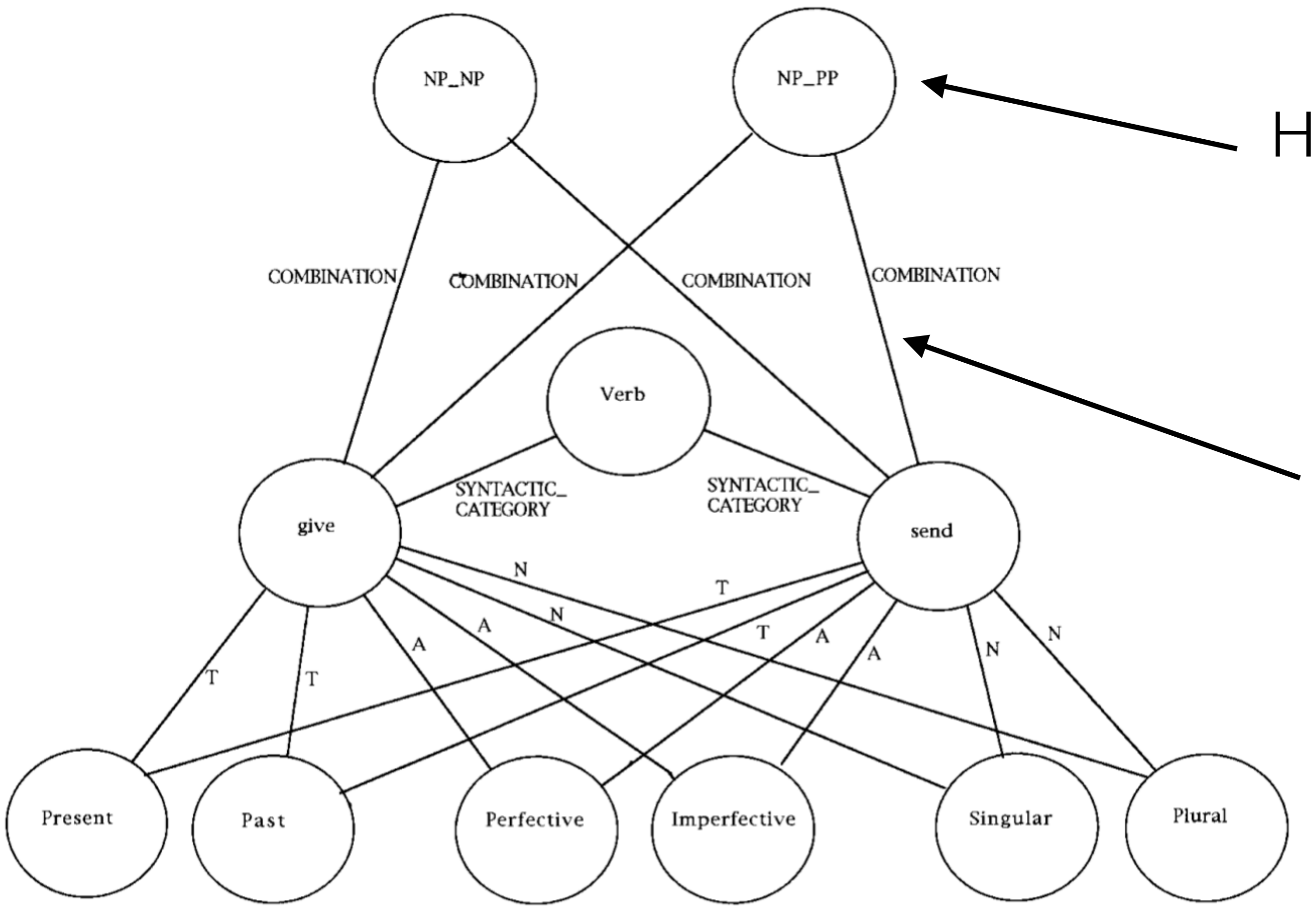
Prime	Prime Type	Verb Type
The girl gave the boy the book.	DO	Same
The girl showed the boy the book.	DO	Different
The girl gave the book to the boy.	PD	Same
The girl showed the book to the boy	PD	Different



Pickering & Branigan (1998)
Momma (under review)

Syntactic prediction & syntactic priming

Syntactic priming & prediction are reflecting heightened activation of treelet/syntactic schema?

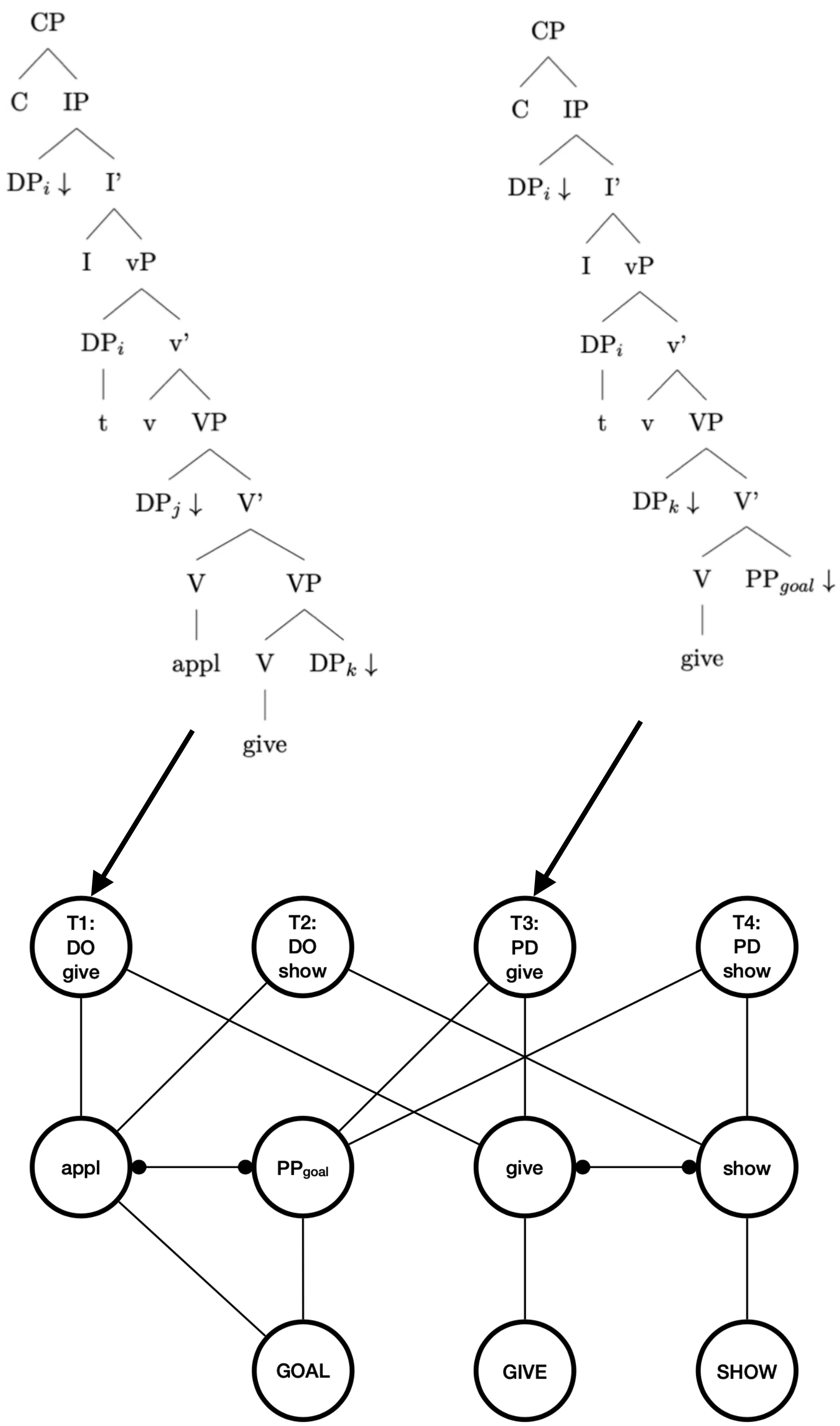


Heightened activation of NP_PP node -> PP syntactic priming

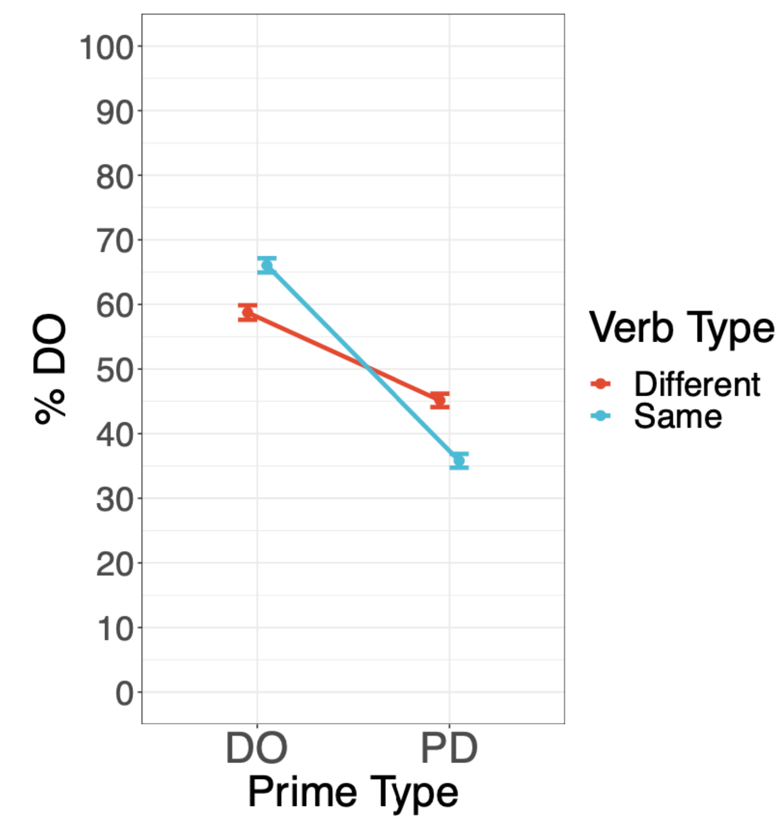
Heightened activation of the link between a verb & NP_PP node ->
Lexical boost effect
(Note: there is not such things as ‘activation’ of nodes)

Pickering & Branigan (1998)

Syntactic priming & ‘treelet’



Prime	Prime Type	Verb Type
The girl gave the boy the book.	DO	Same
The girl showed the boy the book.	DO	Different
The girl gave the book to the boy.	PD	Same
The girl showed the book to the boy	PD	Different



Momma (under review)

Parallel architecture: a claim

Prediction via preactivation (across all levels) is a by-product of lexical access.

Lexical access is not just about accessing words, but it's about accessing any (linguistic) item in long term memory: Extended lexicon

Cat

Semantics: $[\text{CAT}_1]$
Syntax: N_1
Phonology: $/\text{kæt}_1/$

-s

Semantics: $[\text{PLUR}\ (\text{X}_x)]_y$
Morphosyntax: $[\text{N}_x\ \text{PLUR}_6\ ]_y$
Phonology: $/\ \ldots_x\ \text{s}_6\ /_y$

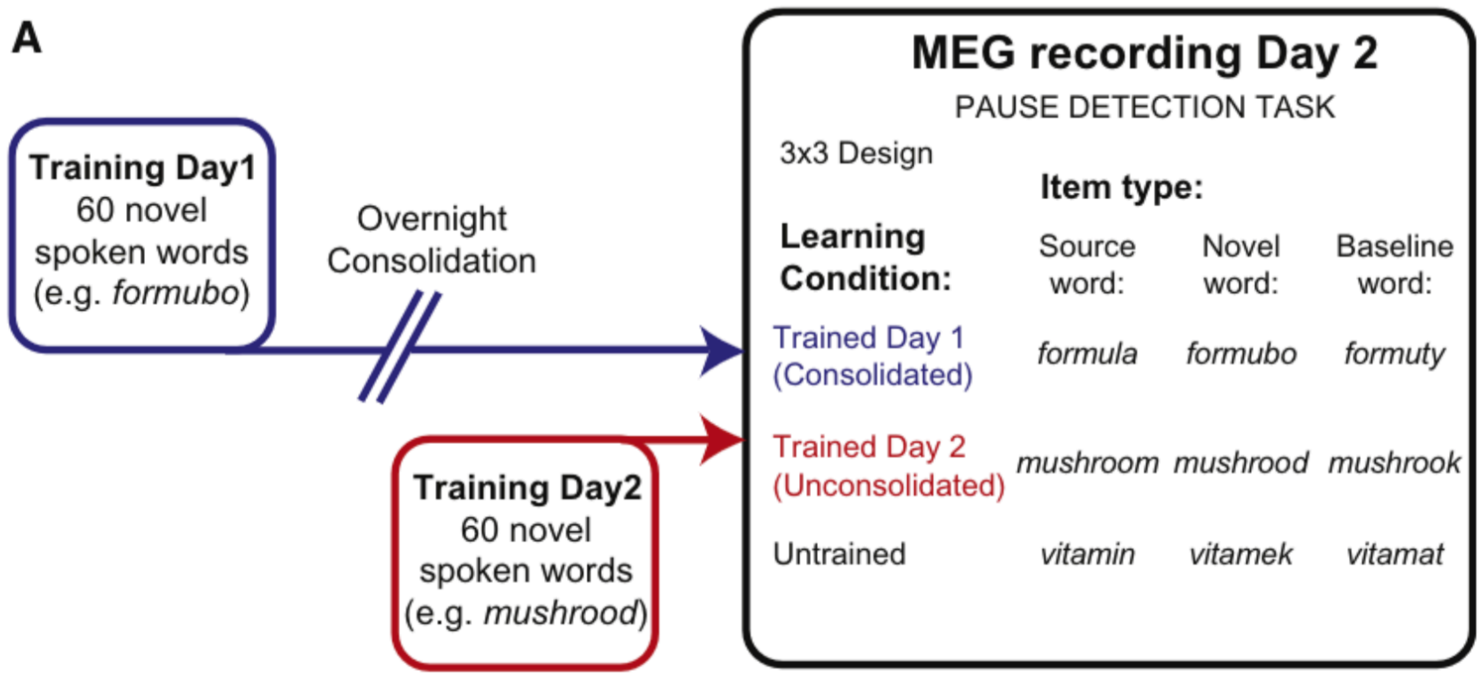
Cats

Semantics: $[\text{PLUR}\ (\text{CAT}_1)]_7$
Morphosyntax: $[\text{N}_1\ \text{PLUR}_6\ ]_7$
Phonology: $/\text{kæt}_1\ \ \text{s}_6\ /_7$

Priming and prediction both arise from activation spreading in extended lexicon

Case study 2: phoneme prediction

Phoneme prediction

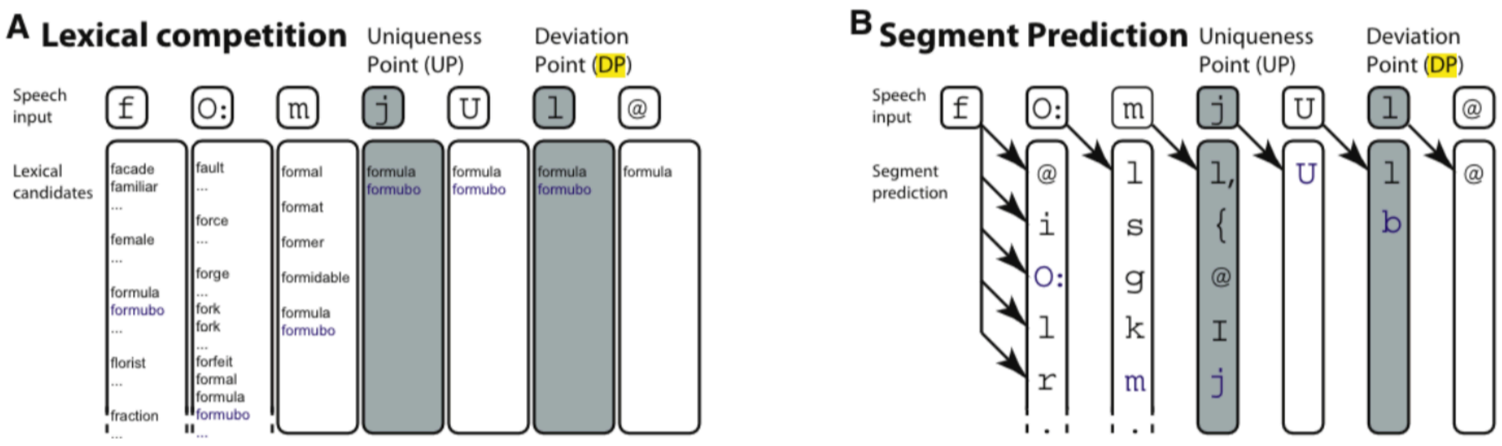


Newly learned words become part of the lexicon after a day (consolidation)

Gagenepain et al. (2008)

Phoneme prediction

Lexical competition vs. phoneme prediction models

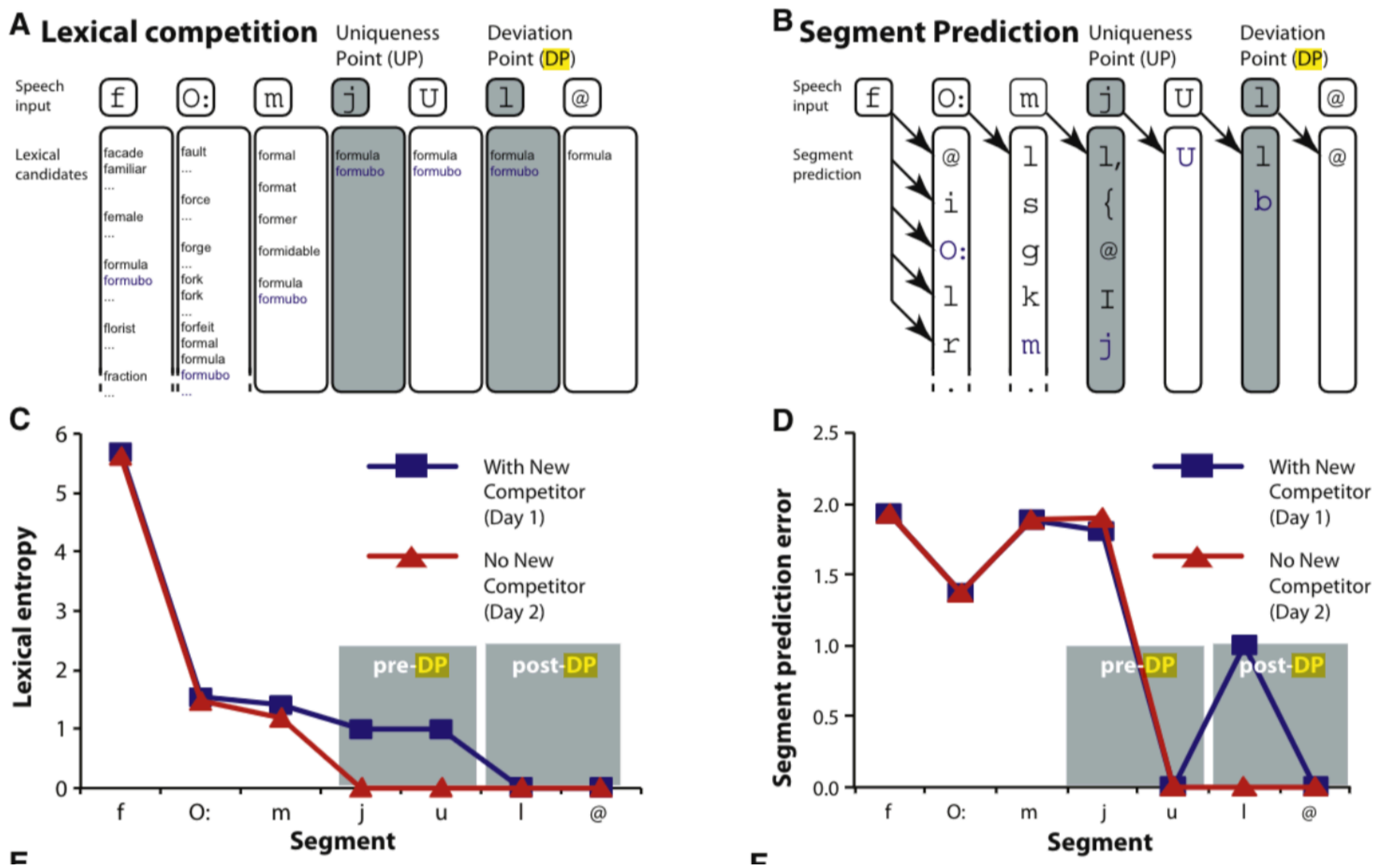


After a new word ("formuba") becomes a part of lexicon, uniqueness point shifts.

Gagenepain et al. (2008)

Phoneme prediction

Lexical competition vs. phoneme prediction models



After a new word ("formuba") becomes a part of lexicon, uniqueness point shifts.

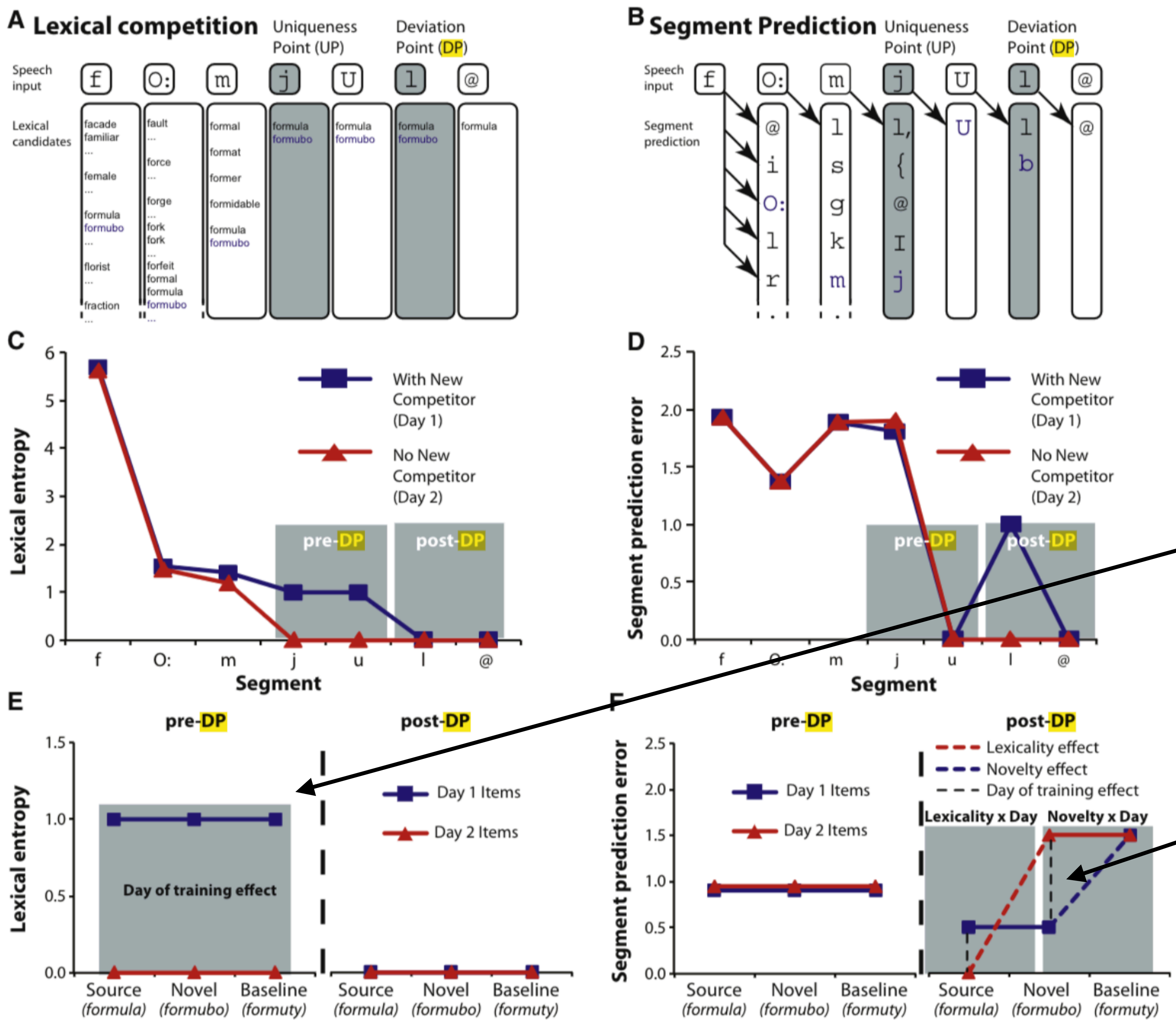
Lexical competition (cohort) model predicts that, adding 'formuba' to lexicon increases competition before the deviation point (the shifted uniqueness point)

In comparison, phoneme prediction model predicts that adding 'formuba' to lexicon makes the prediction error of /b, l/ greater at the deviation point.

Gagenepain et al. (2008)

Phoneme prediction

Lexical competition vs. phoneme prediction models

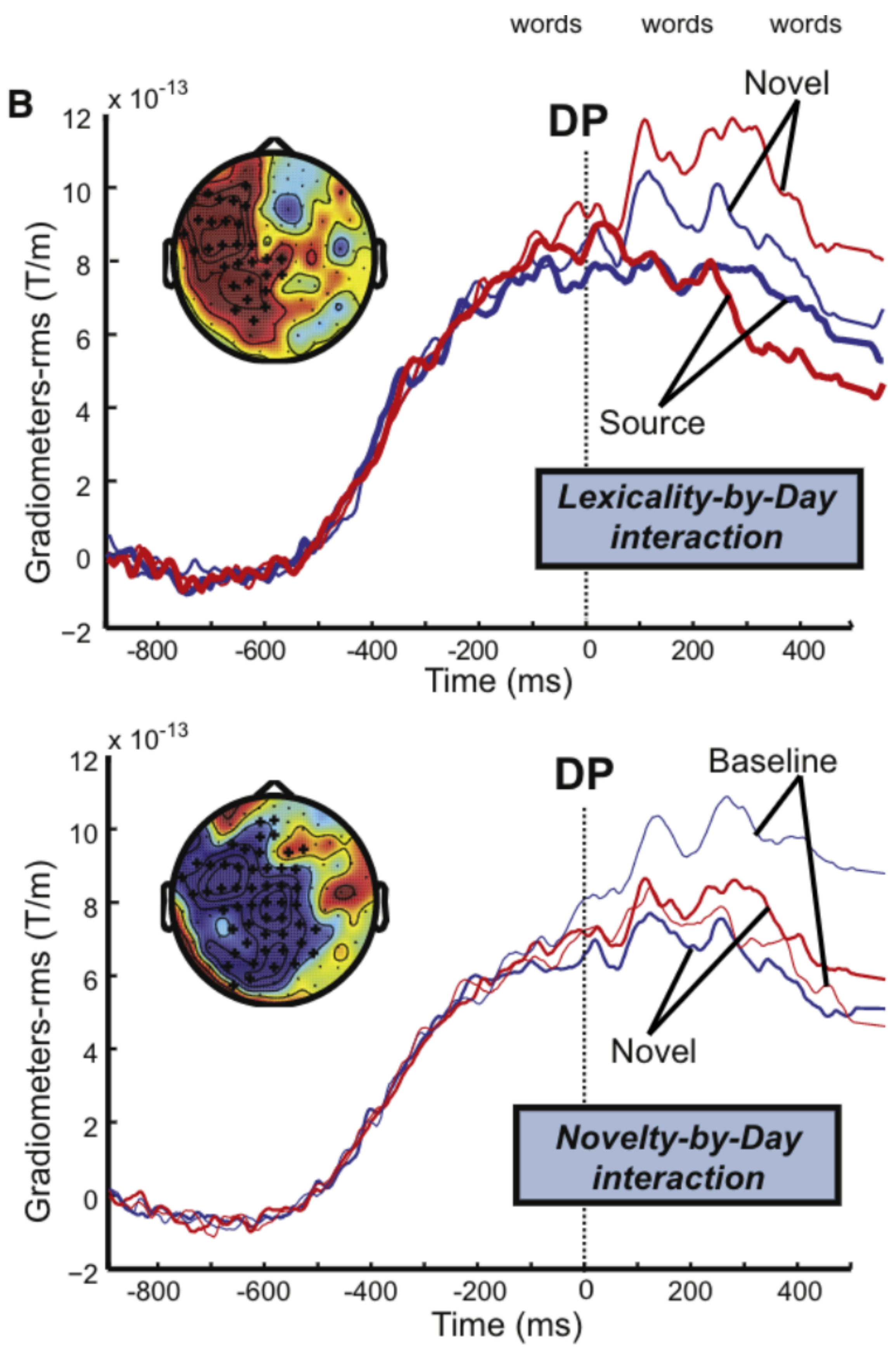


Lexical competition model => Competition effect before the deviation point

Phoneme prediction model => surprisal effect at the deviation point (only for the consolidated items)

Gagenepain et al. (2008)

Phoneme prediction

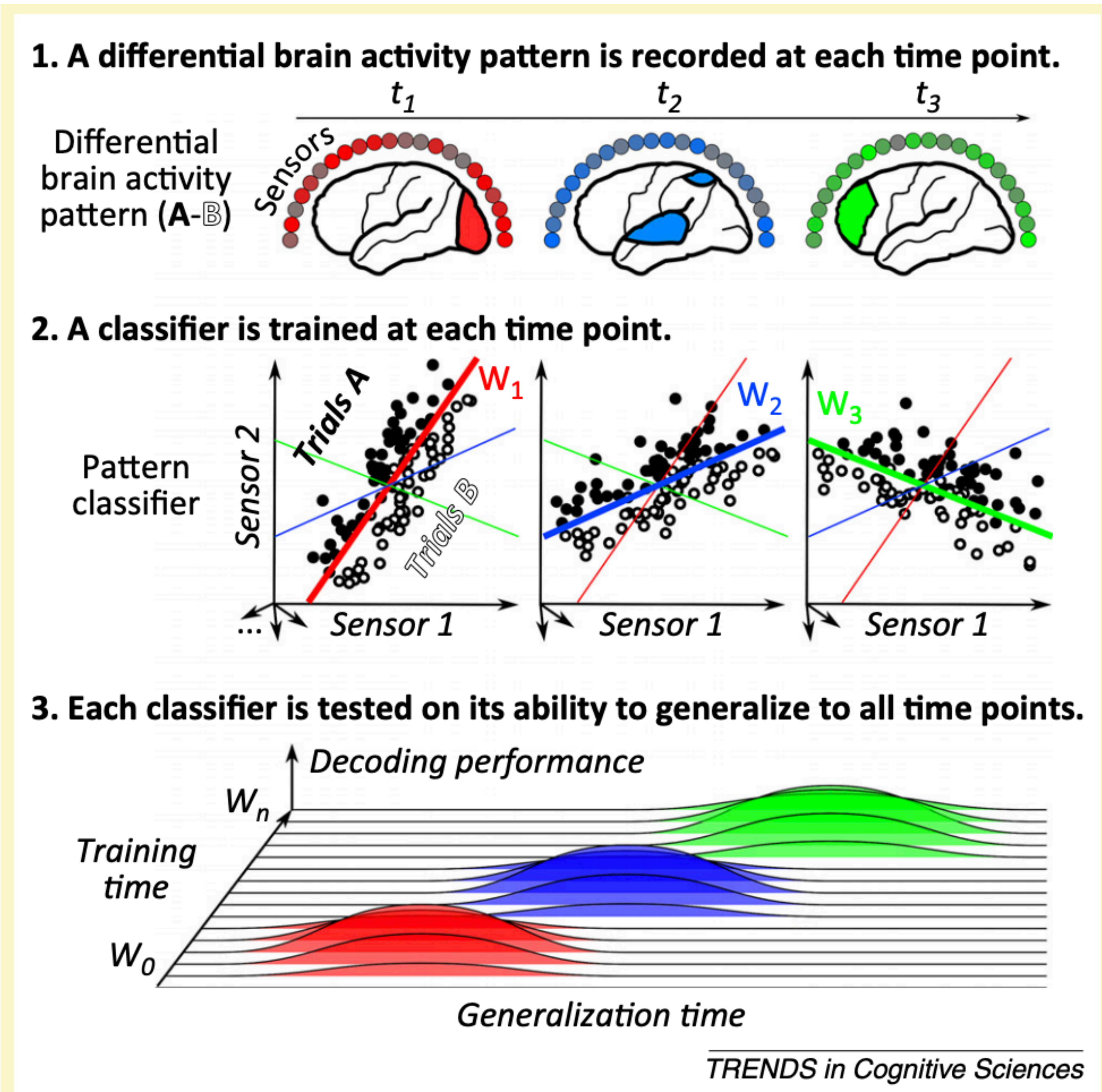


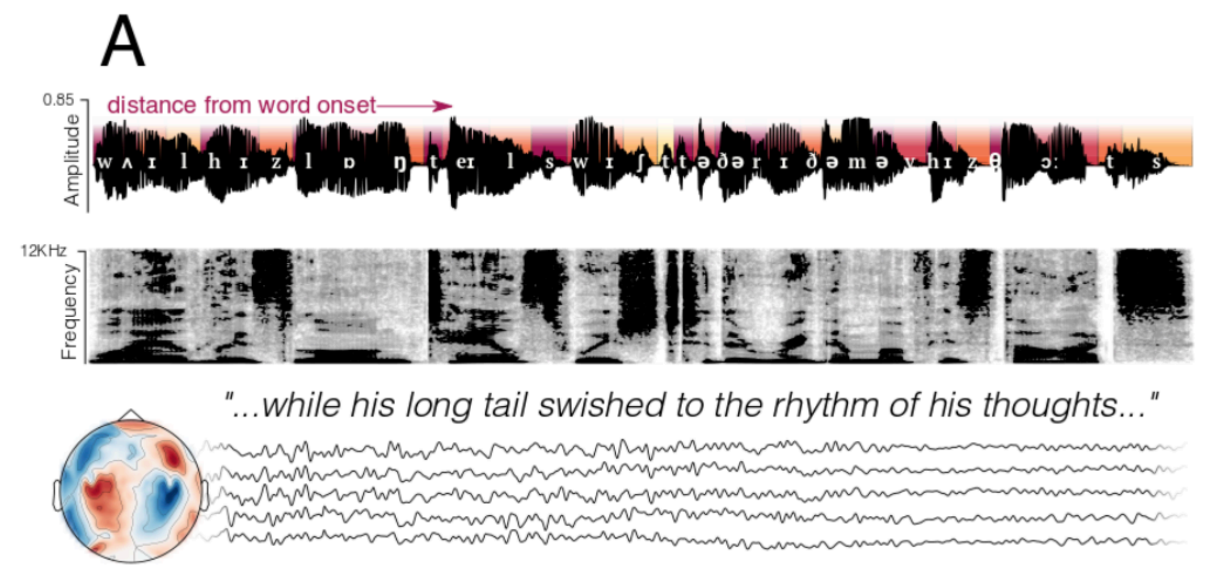
Right after the deviation point, divergence starts right after the deviation point.

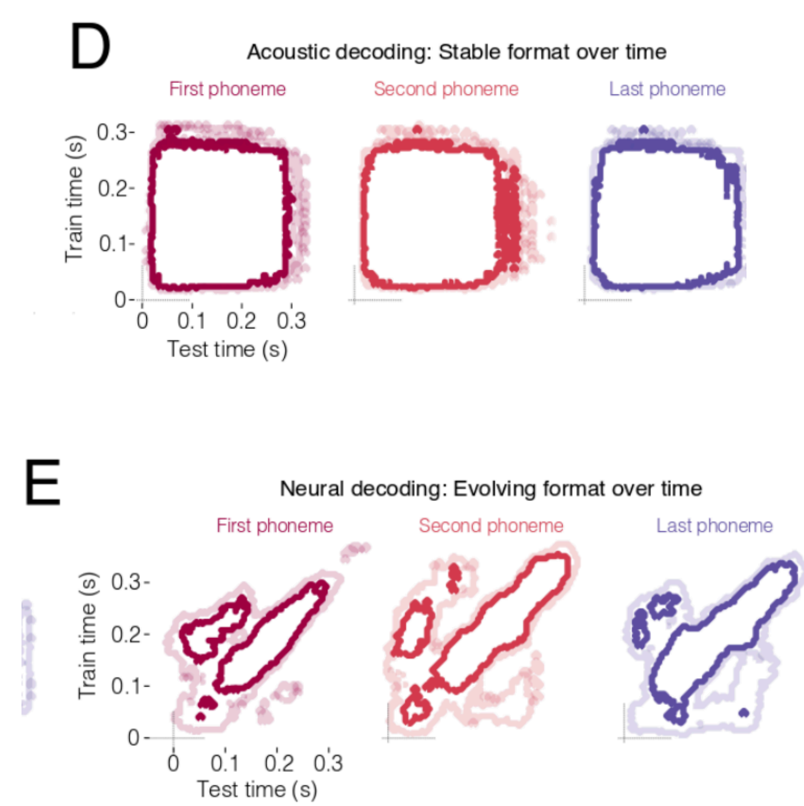
Difference between ‘formula’ vs. ‘formuba’ smaller for consolidated (Day 1) than unconsolidated (Day 2) items.

Difference between ‘formula’ and ‘formuty’ (total nonword) greater for consolidated (Day 1) than unconsolidated (Day 2) items.

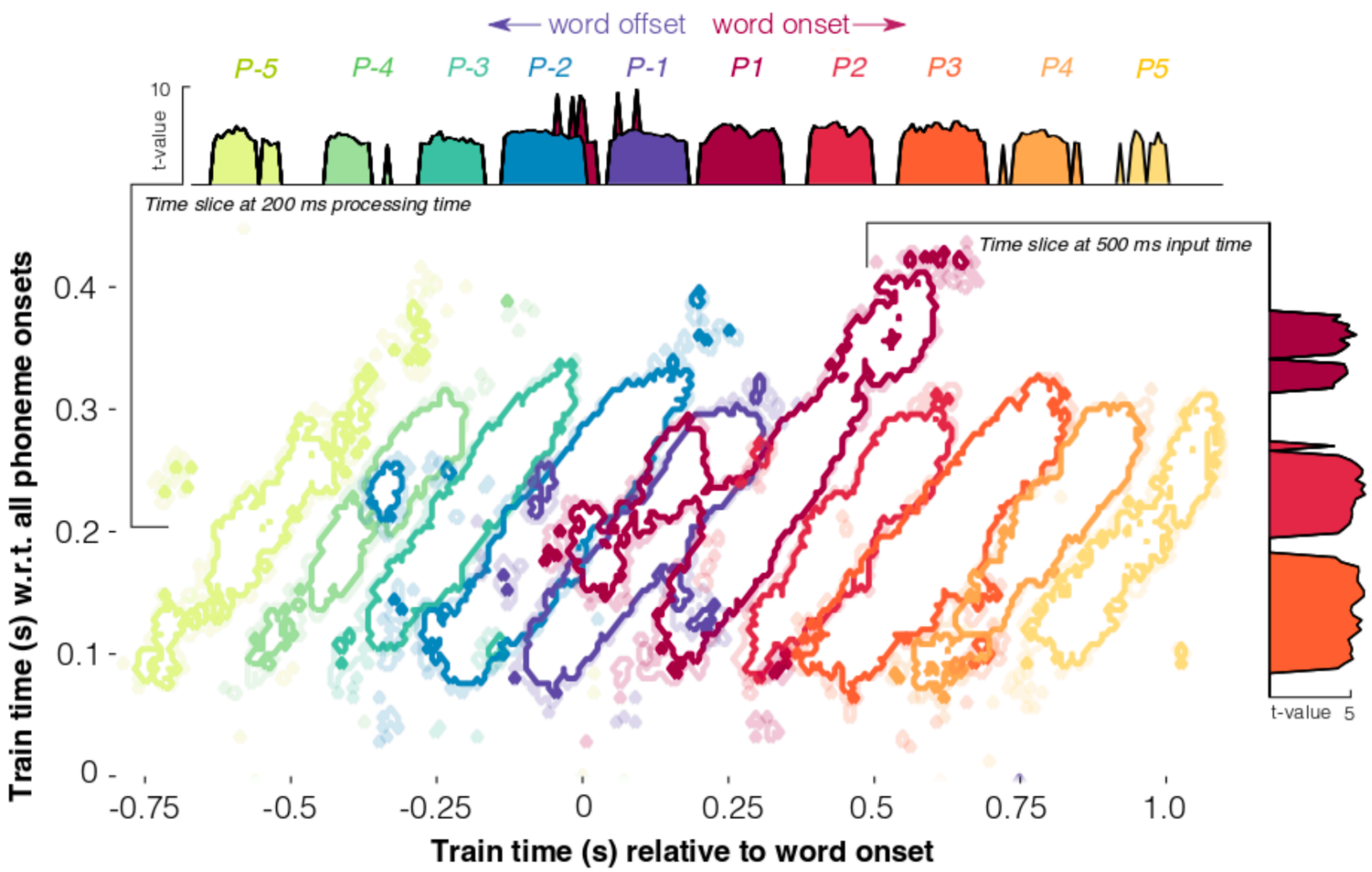
Gagenepain et al. (2008)

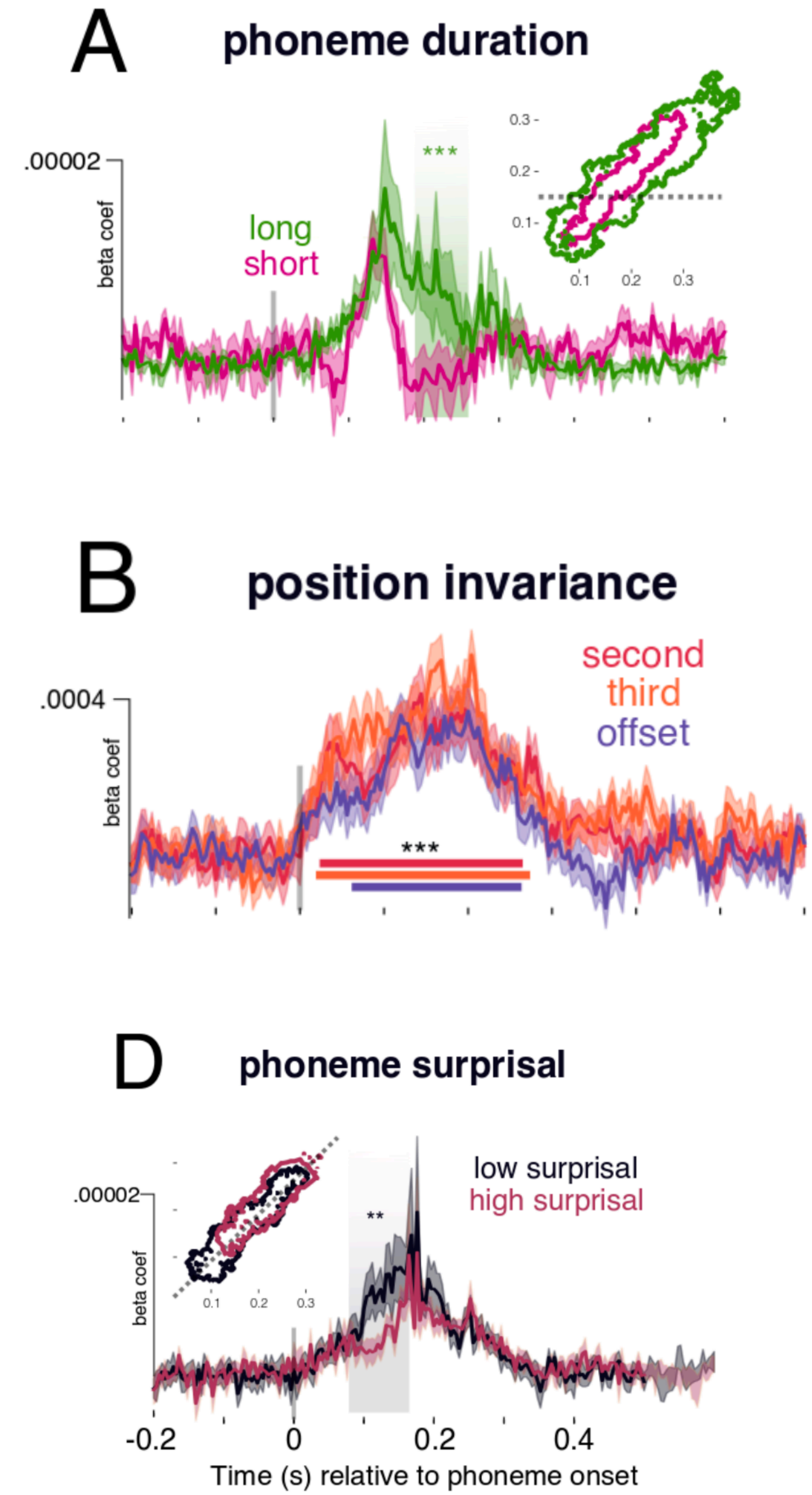






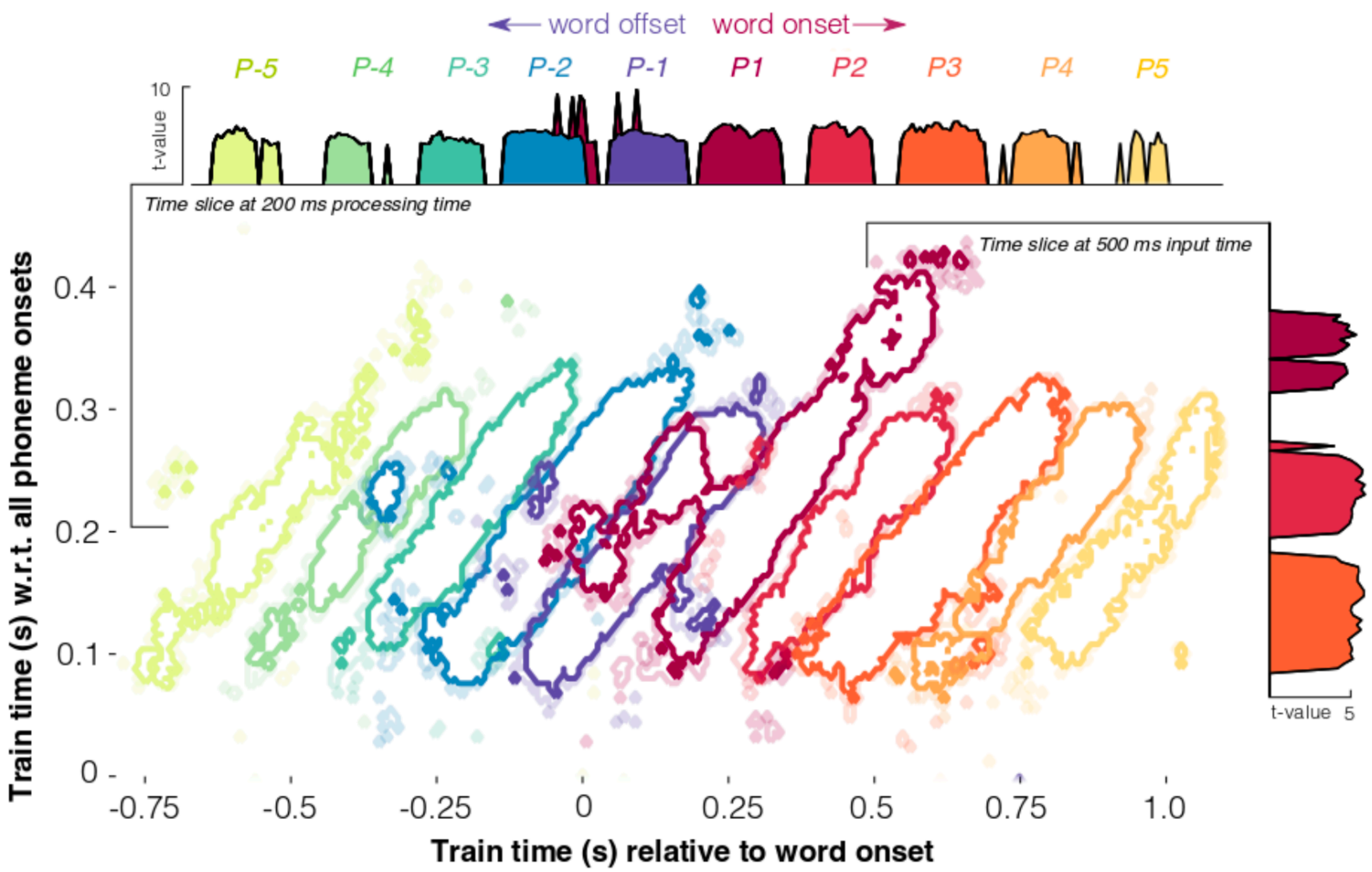
Gwilliams et al. (under review?)

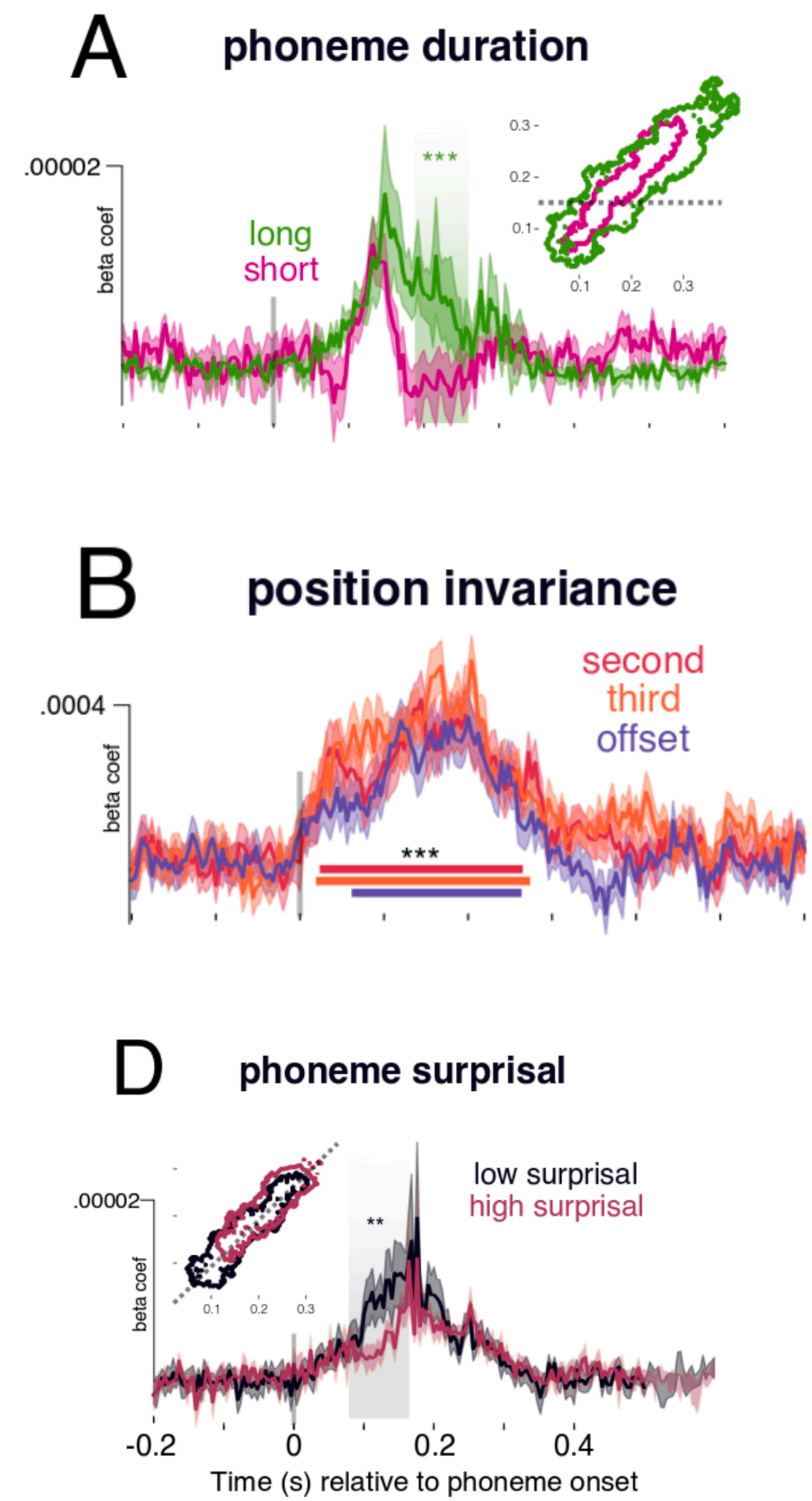




A: short phonemes have narrower diagonal shape -> The brain representation of short phonemes evolves faster

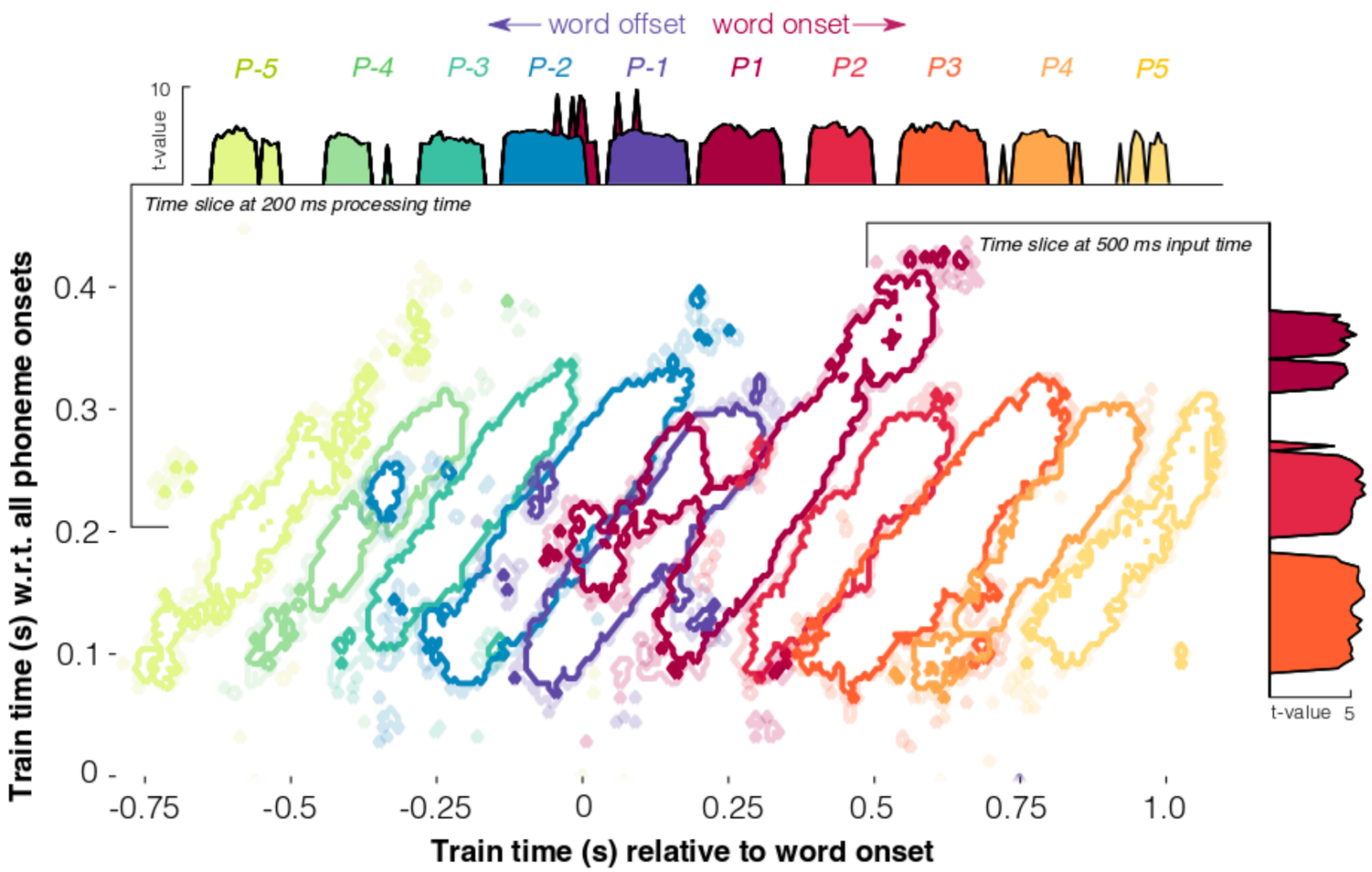
Gwilliams et al. (under review?)

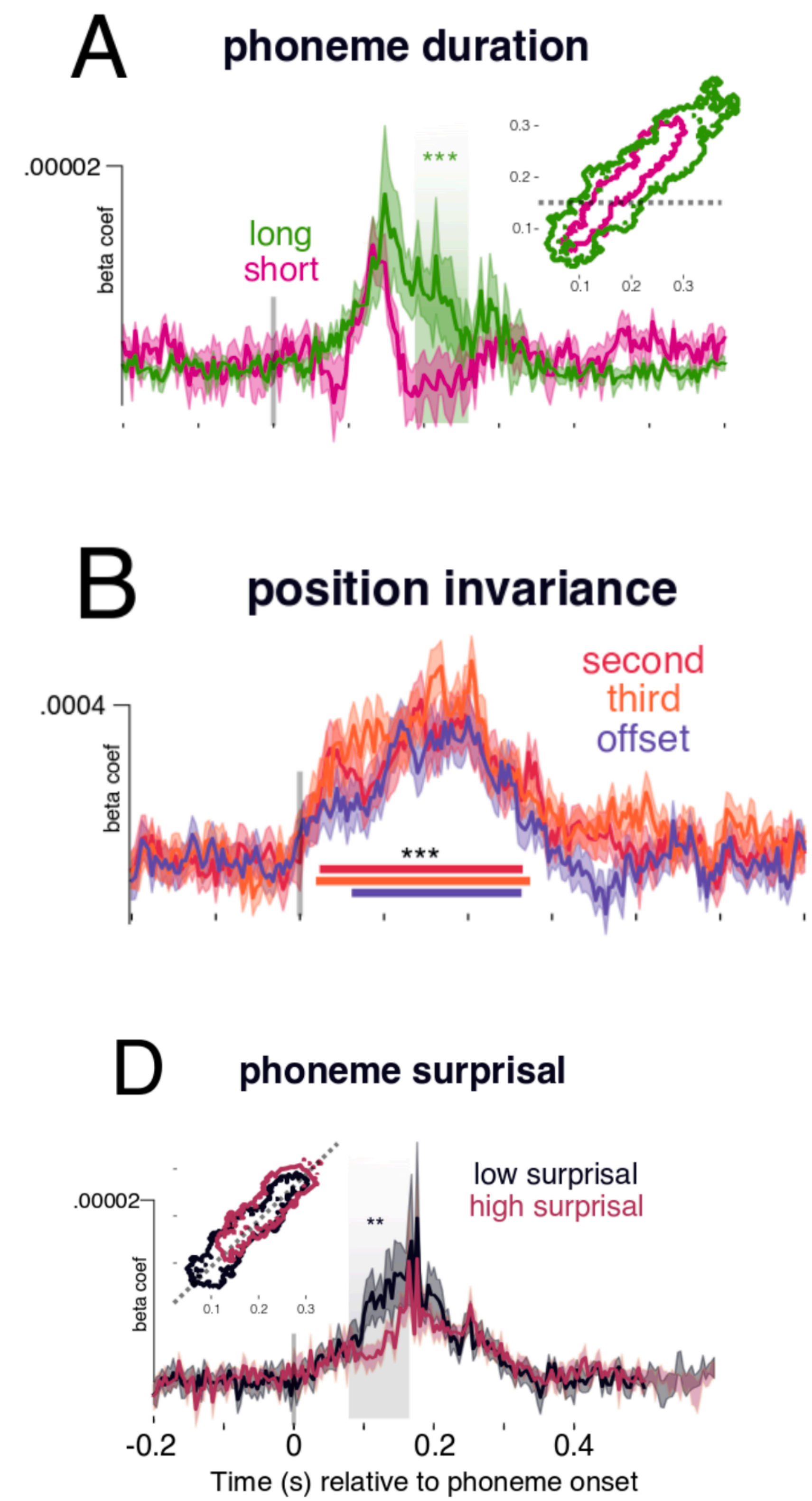




B: Classifier trained on first phoneme good at decoding the second, third and last phonemes -> evidence for position-invariant neural representations

Gwilliams et al. (under review?)





C: more surprising (= less predictable) phonemes have later onset of decodable time region -> evidence for phoneme prediction?

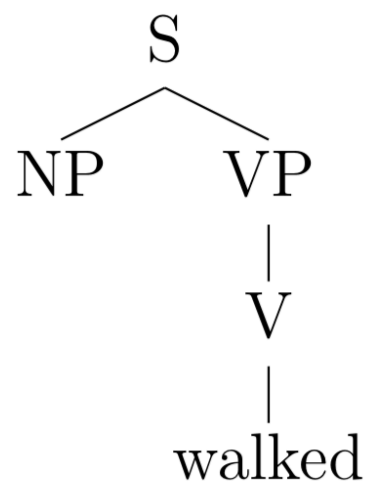
Gwilliams et al. (under review?)

Tree adjoining grammar

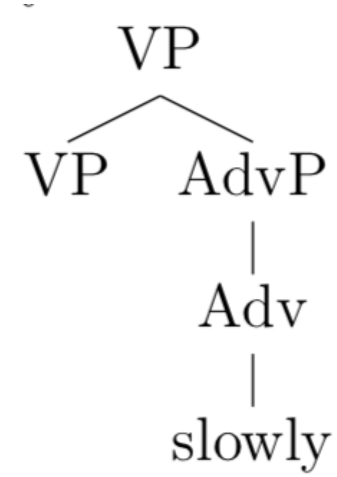
Basic unit: Elementary trees

(1)

(2)



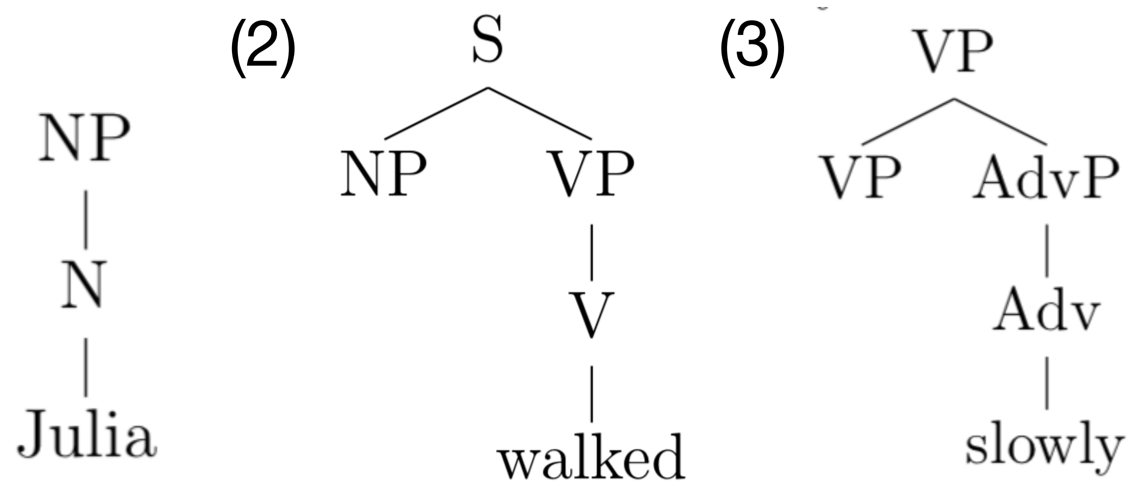
(3)



Tree adjoining grammar

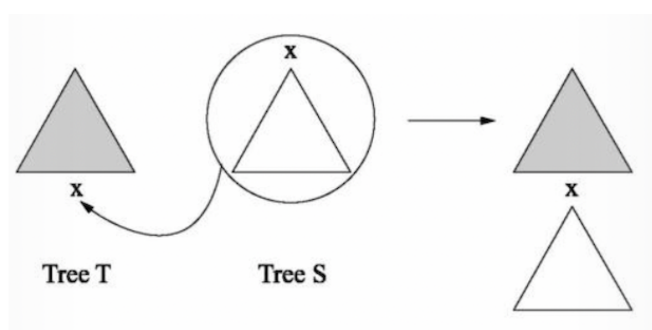
Basic unit: Elementary trees

(1) (2) (3)



Combinatorial operations

Substitution

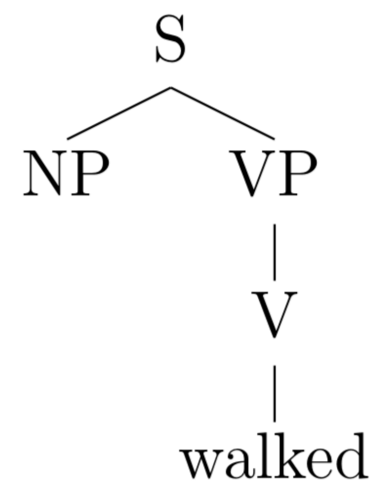


Tree adjoining grammar

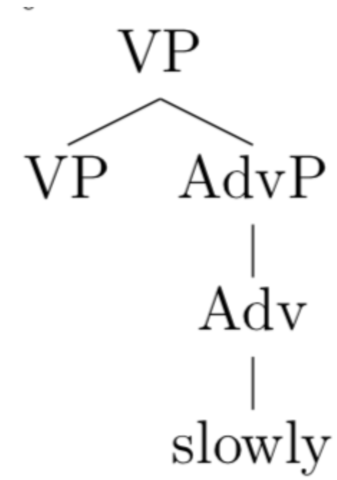
Basic unit: Elementary trees

(1)

(2)

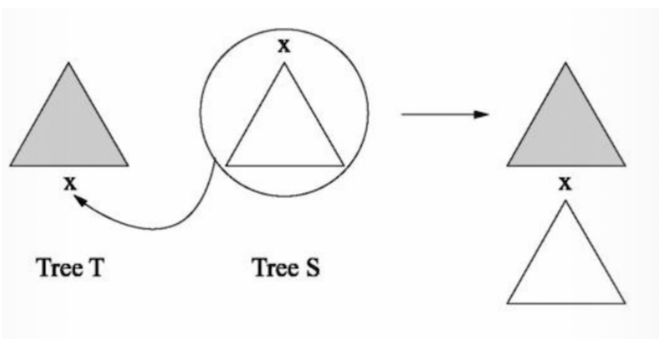


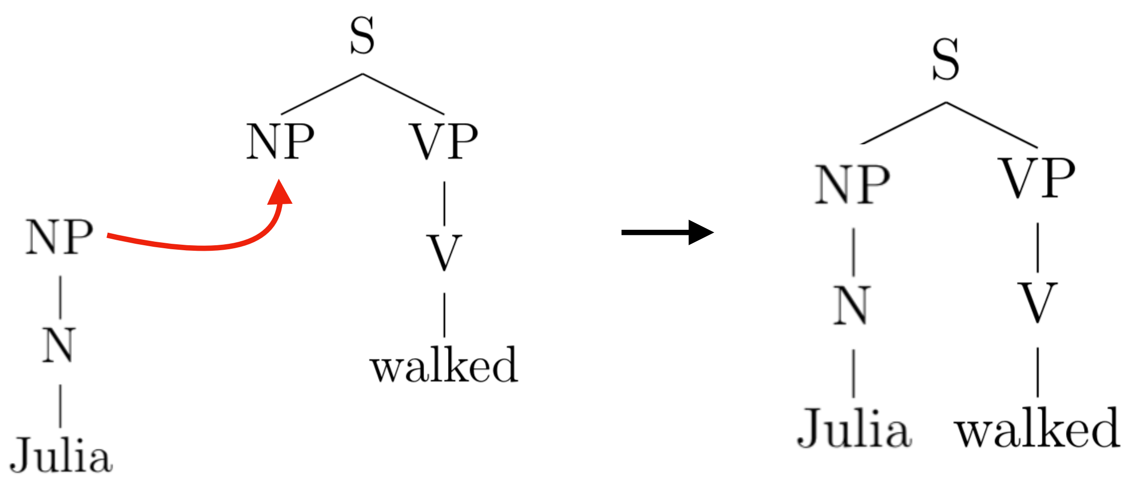
(3)



Combinatorial operations

Substitution

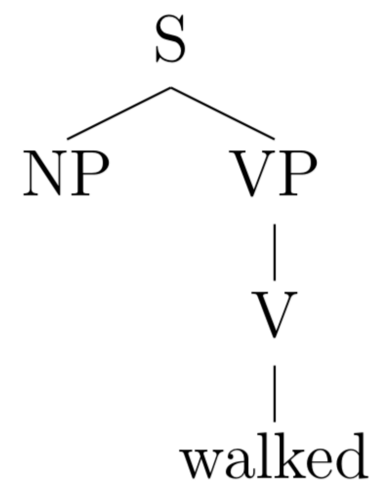


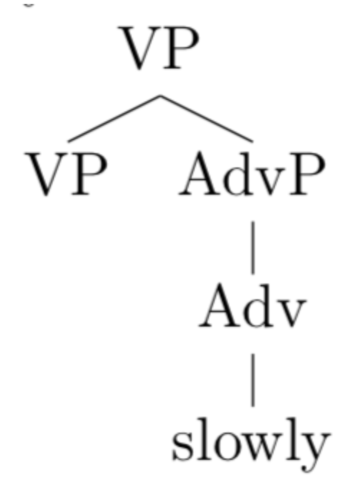


Tree adjoining grammar

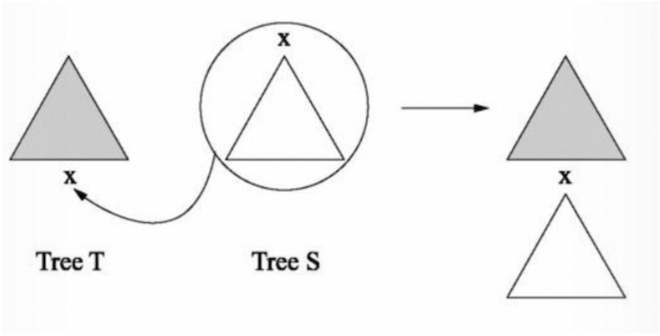
Basic unit: Elementary trees

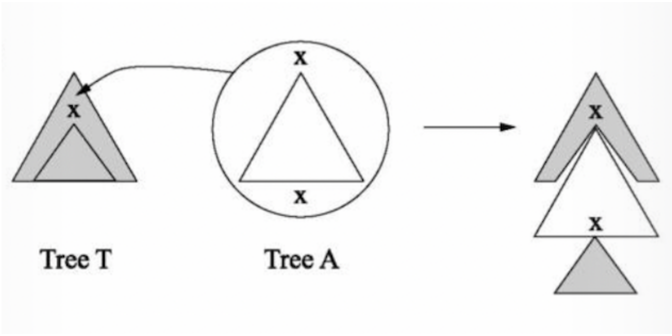
(1)


(2)


(3)


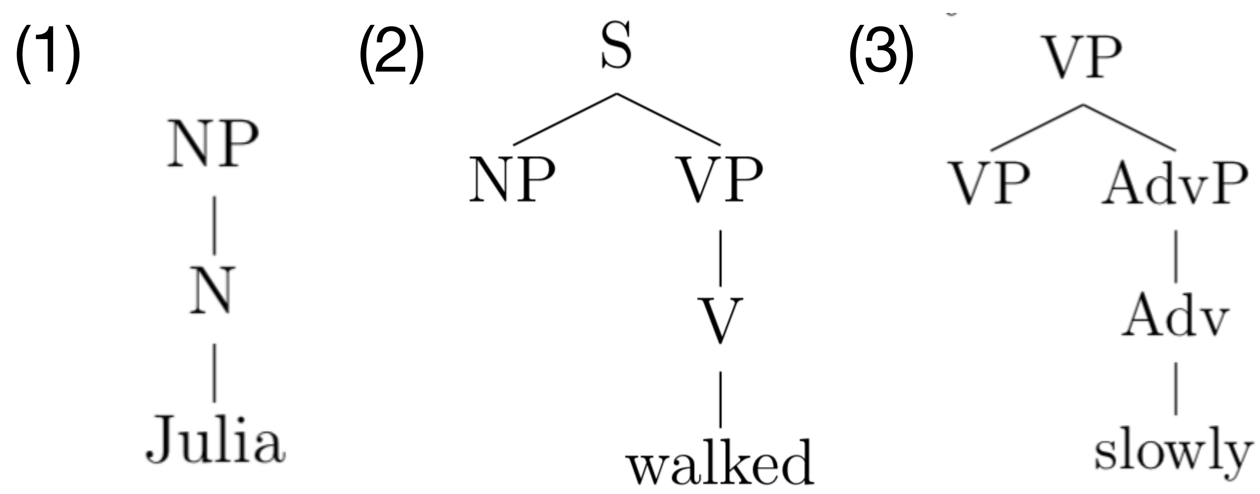
Combinatorial operations

Substitution


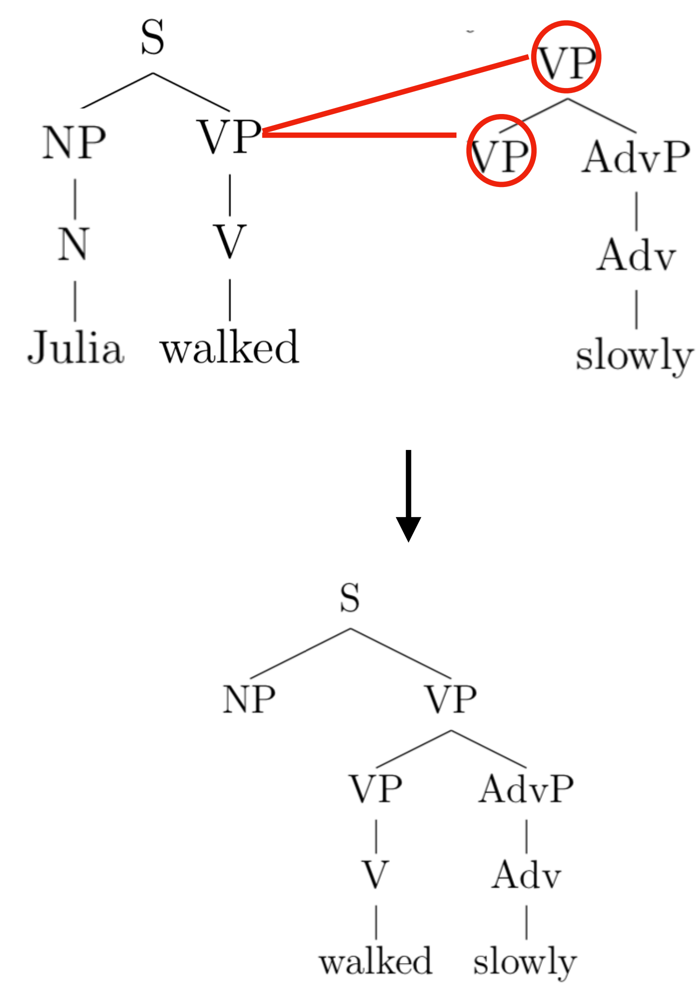
Adjoining


Tree adjoining grammar

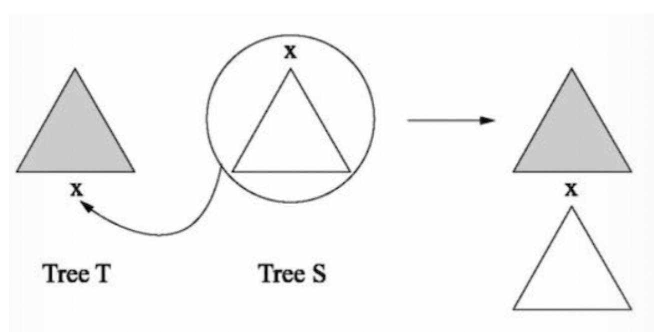
Basic unit: Elementary trees



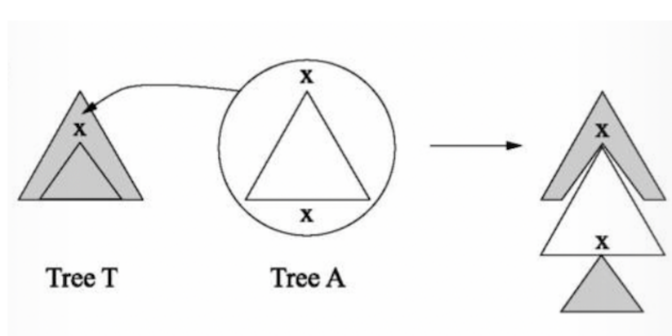
Combinatorial operations



Substitution

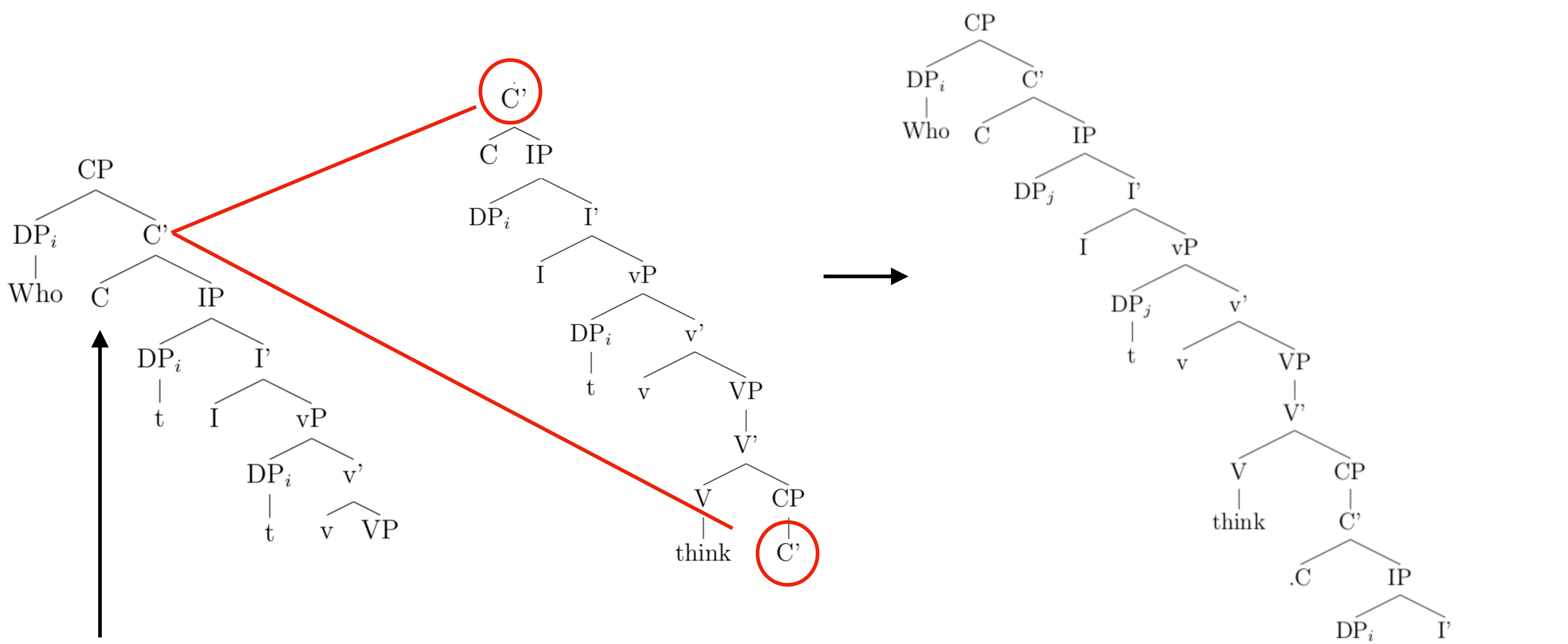


Adjoining



Tree adjoining grammar

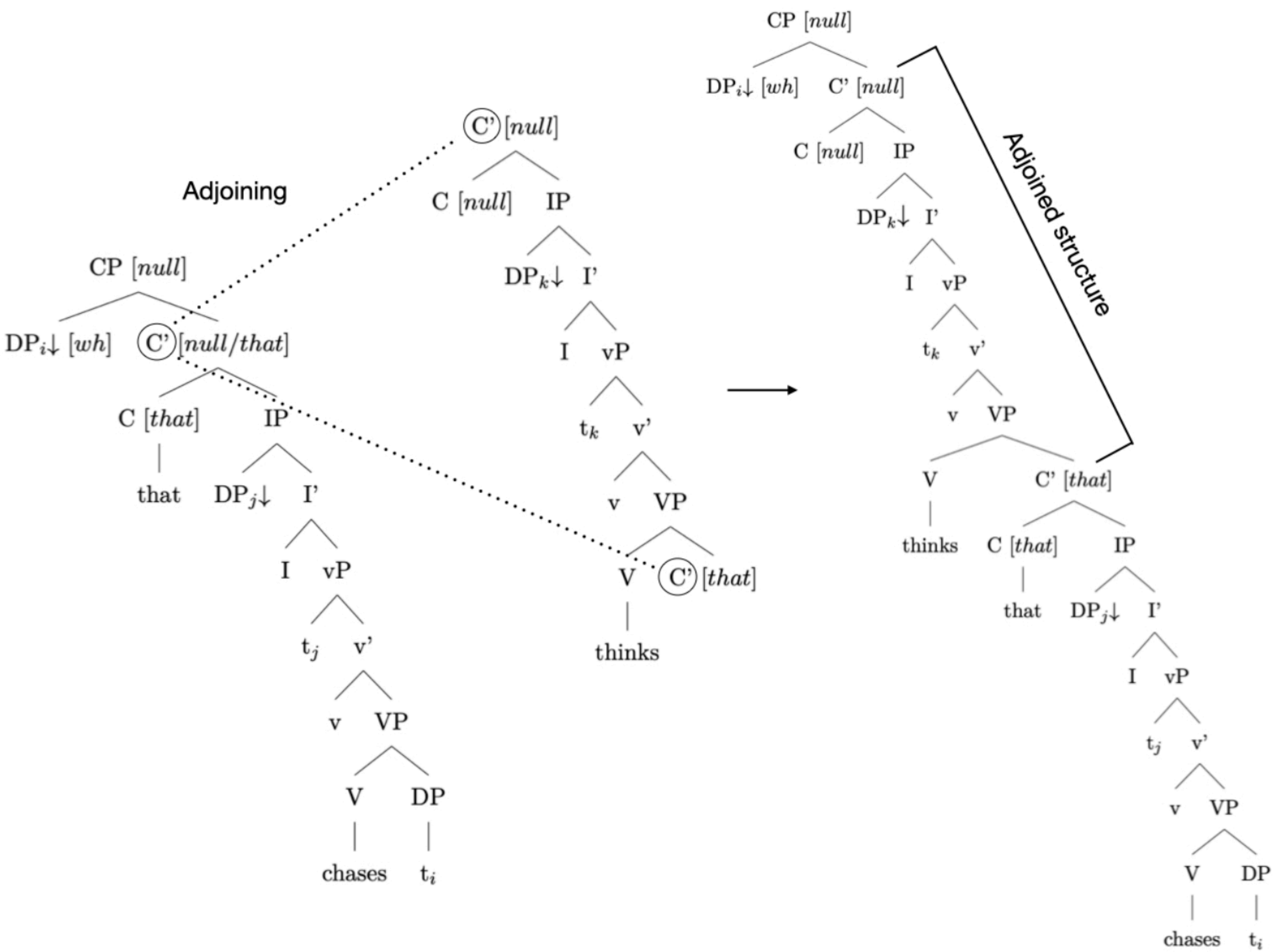
Empirical generalization: Speakers plan gap structures as soon as they represent the filler (and later insert the materials in between).



The decision about the embedded complementizer is already made here

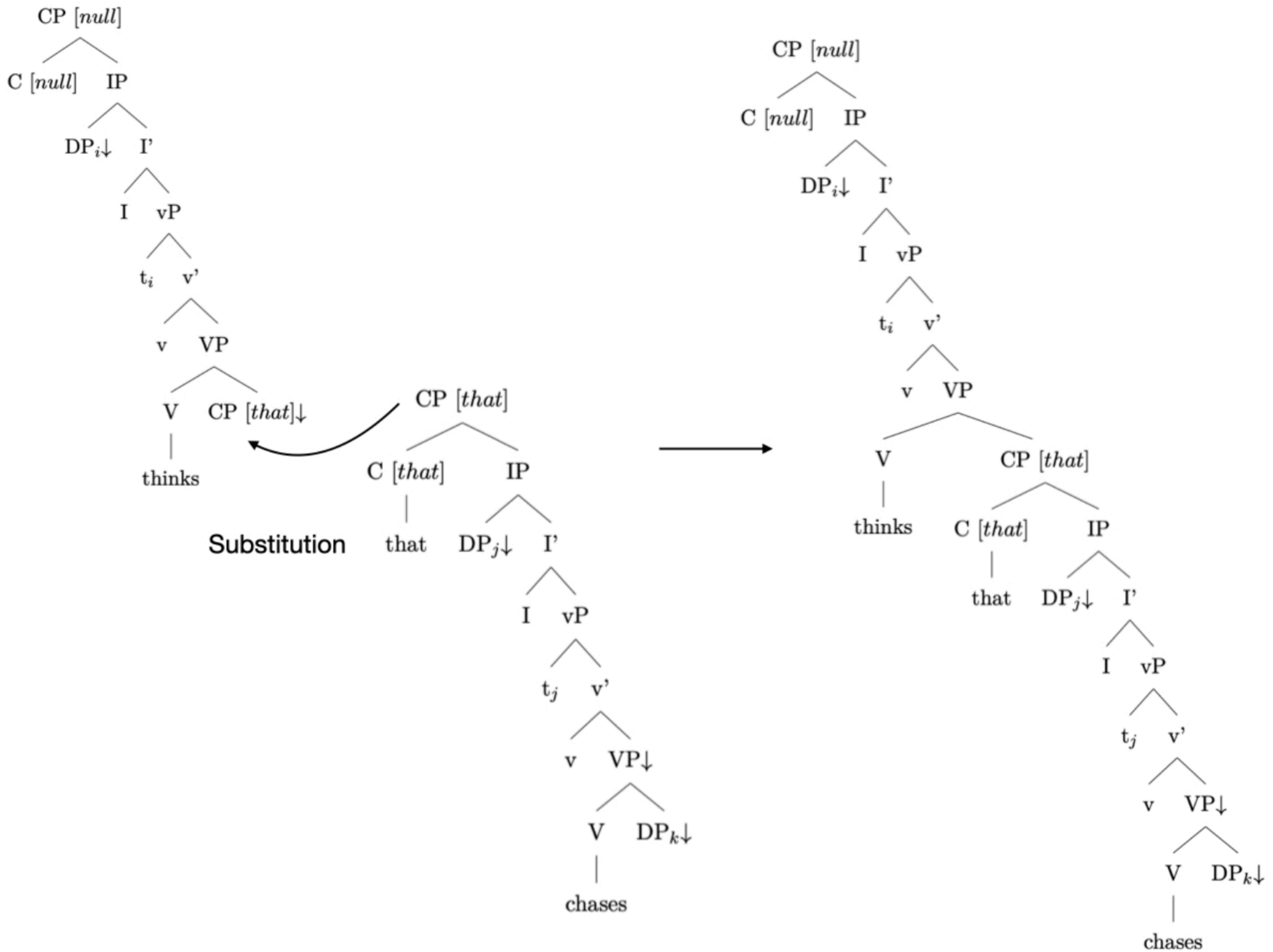
Filler-gap dependencies in TAG

<u>Who</u> does the girl think the dog chased __?



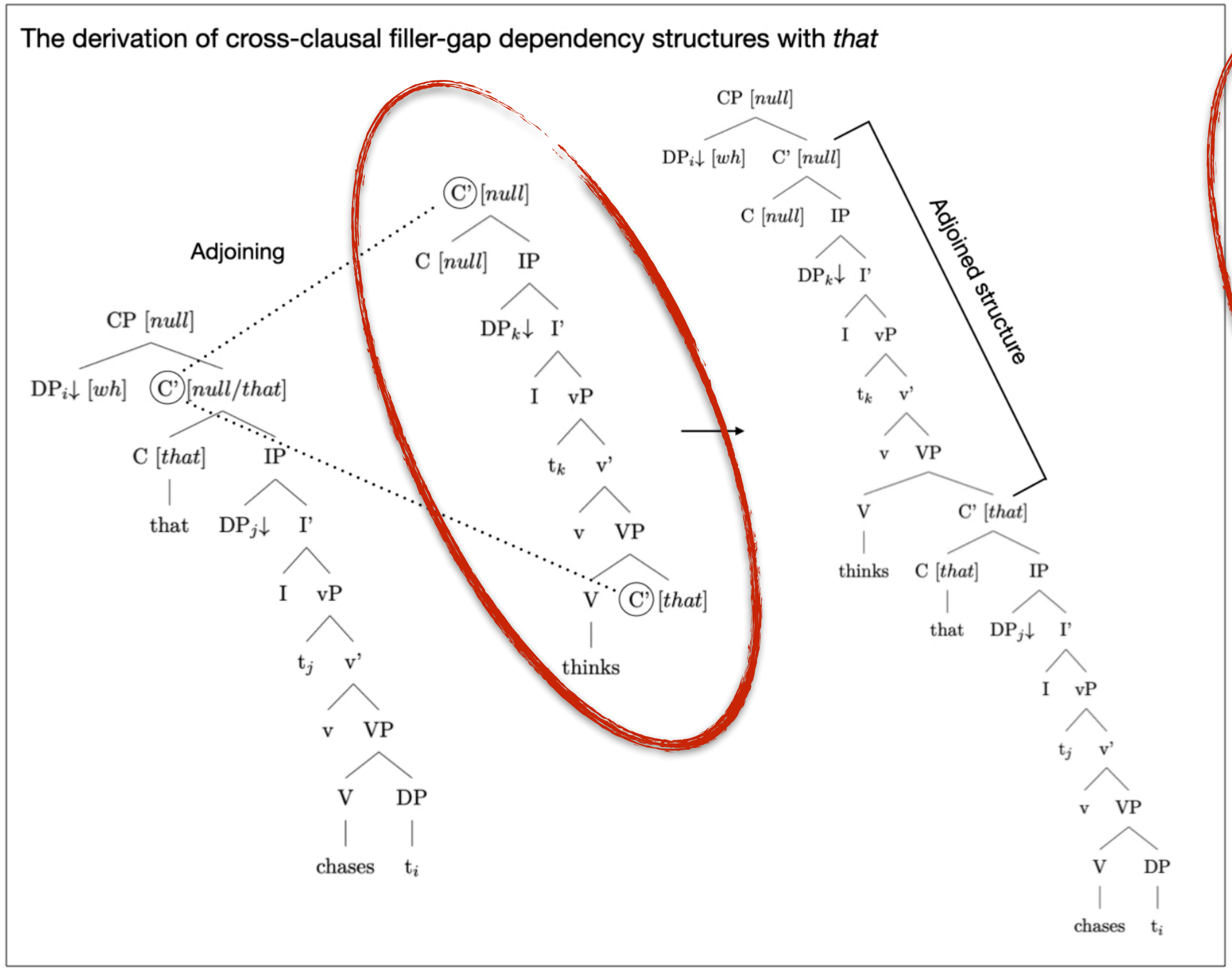
Filler-gap dependencies in TAG

<u>Who</u> __ thinks that the dog chases the cat?

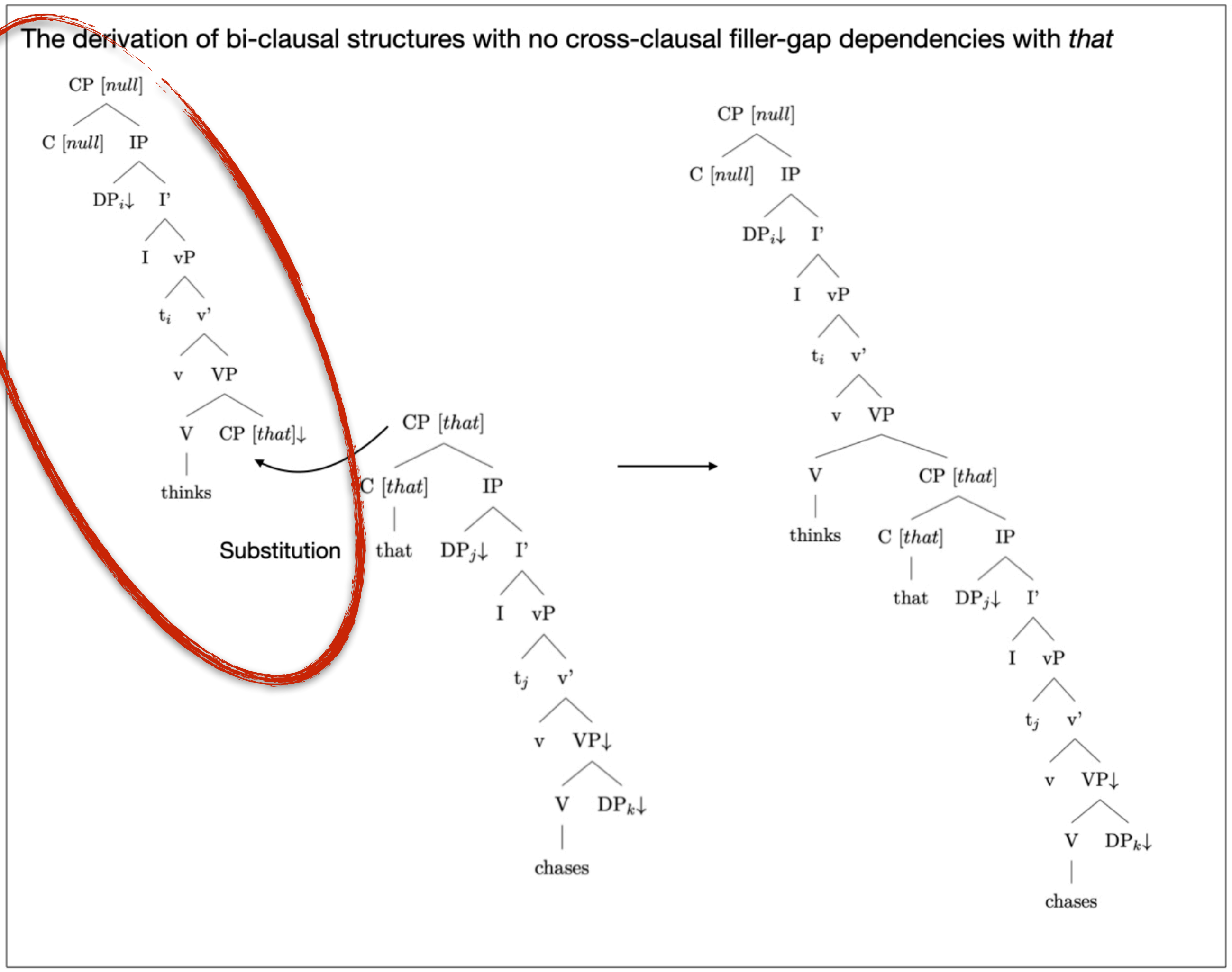


Filler-gap dependencies in TAG

Adjoining



Substitution



Two distinct elementary trees, both headed by *think* but one for adjoining, the other for substitution.

Both elementary trees may be primed by repetition, but they don't prime each other.

Filler-gap dependencies in TAG

Adjoining requires a structural representation (***elementary tree***) that contains *think* and *that* specifically used for adjoining.

When a sentence does not contain a cross-clausal filler-gap dependency, an elementary tree that contains *think* and *that* is distinct from the one used for adjoining.

Prediction: elementary trees containing verbs like *think* and *that* can be primed, but only when both prime and target sentences contain a cross-clausal filler-gap dependency or when neither does.

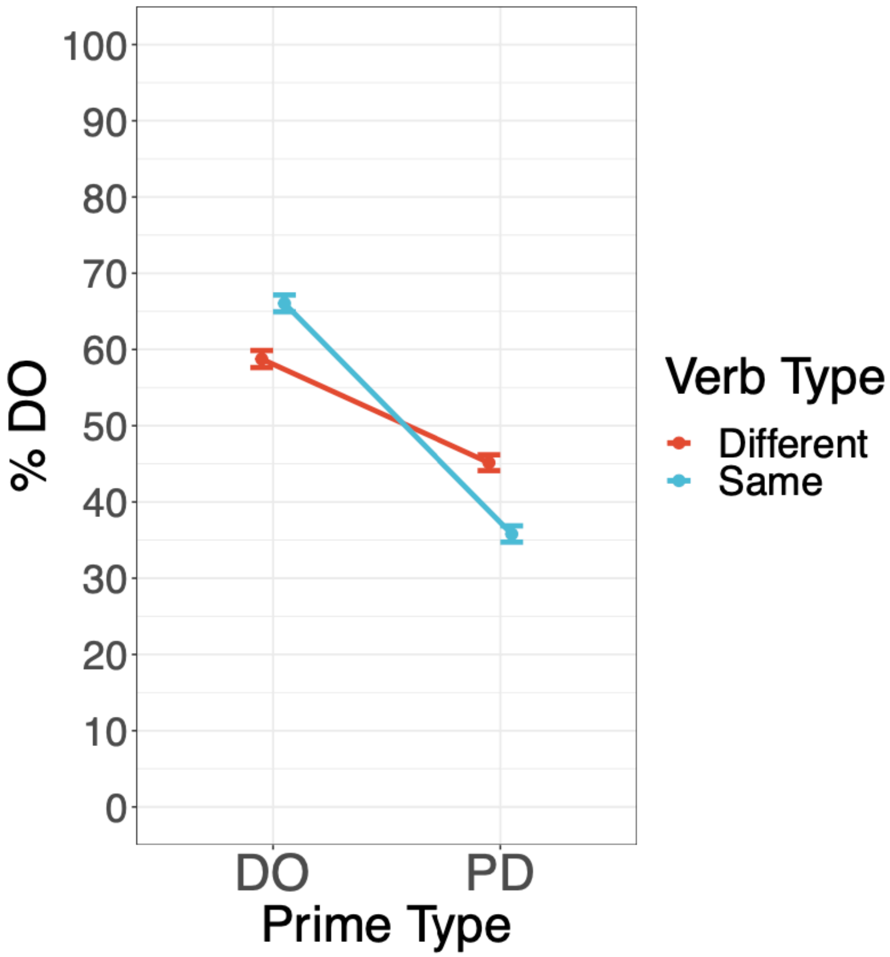
Structural priming and lexical-boost

The magnitude of structural priming effect can be increased by repeating a head of the primed structure (Pickering & Branigan, 1998; cf. Scheepers et al. 2017 but see Calminati et al. 2019).

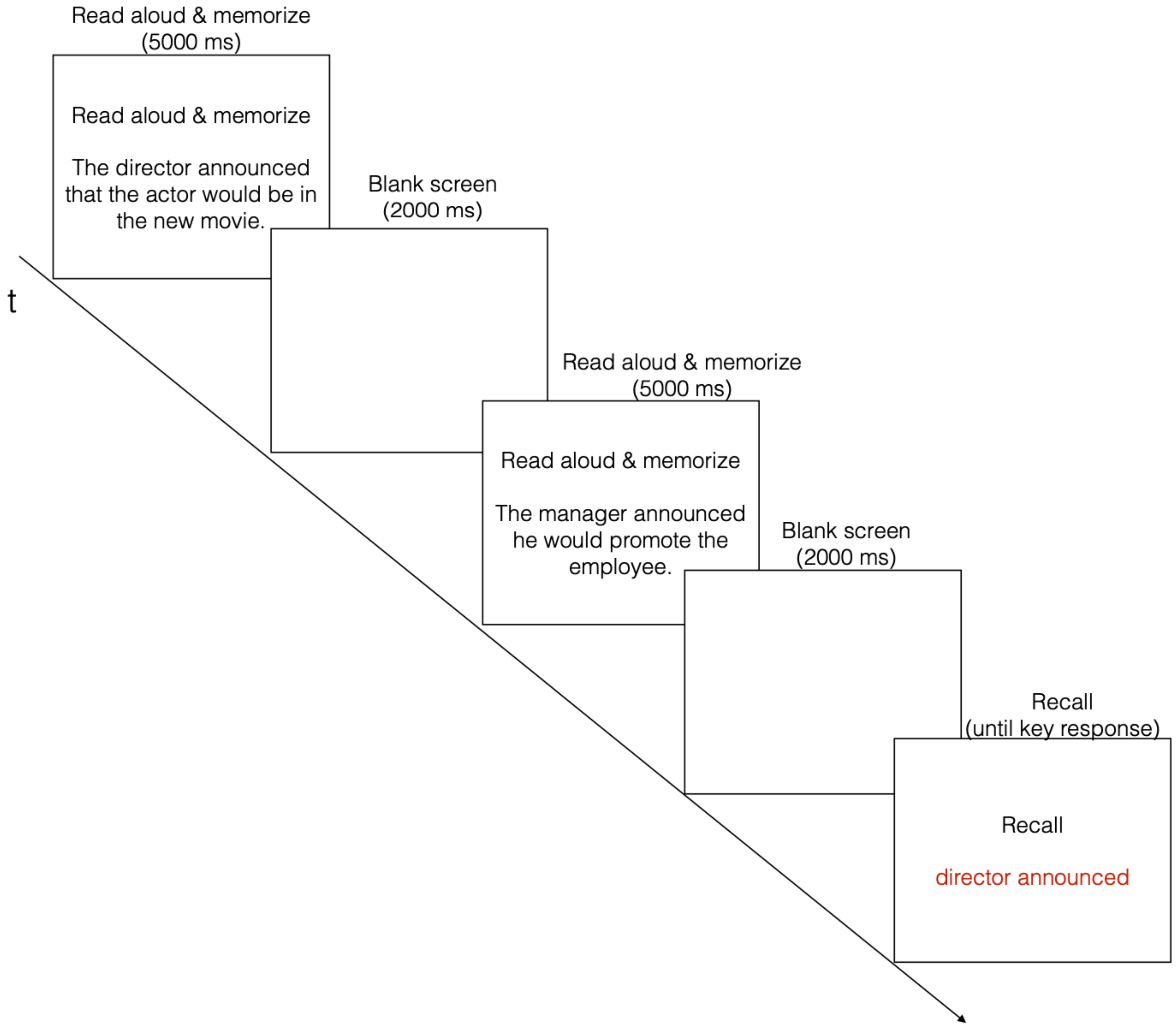
Using dative priming as an example:

If target sentences contain ‘give’…

Prime	Prime Type	Verb Type
The girl gave the boy the book.	DO	Same
The girl showed the boy the book.	DO	Different
The girl gave the book to the boy.	PD	Same
The girl showed the book to the boy	PD	Different



Task


Read aloud & memorize
(5000 ms)
Read aloud & memorize
The director announced that the actor would be in the new movie.
Blank screen
(2000 ms)
t
Read aloud & memorize
(5000 ms)
Read aloud & memorize
The manager announced he would promote the employee.
Blank screen
(2000 ms)
Recall
(until key response)
Recall
director announced

Design & Prediction

Exp. #	Prime	Target
Exp. 1	The manager {announced | implied} {that | Ø} he would promote the employee.	The director announced that he would nominate the actor.
Exp. 2	Who$_i$ did the manager {announce | imply} {that | Ø} he would promote t_i?	The director announced that he would nominate the actor.
Exp. 3	Who$_i$ did the manager {announce | imply} {that | Ø} he would promote t_i?	Who did the director announce that he would nominate t_i?
Exp. 4	Who$_i$ t_i {announced | implied} {that | Ø} the manager would promote the employee?	Who did the director announce that he would nominate t_i?
Exp. 5	The manager {announced | implied} {that | Ø} he would promote the employee.	I wonder who$_i$ the director announced that he would nominate t_i?

Cross-clausal FG dependencies?

Neither

Only prime

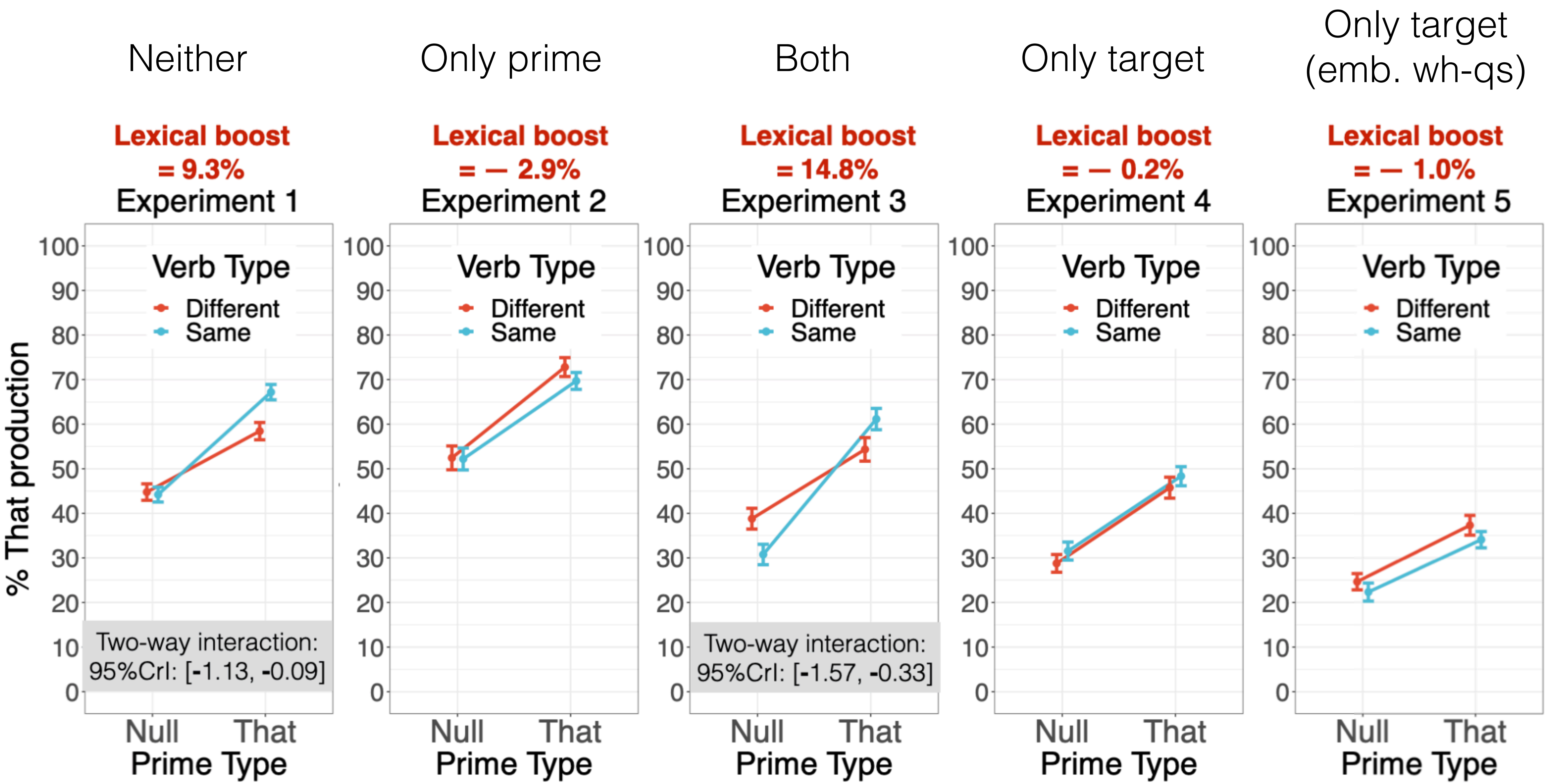
Both

Only target

Only target (emb. wh-q)

Prediction: the lexical boost effect should be observed only when both prime and target contains a cross-clausal filler-gap dependencies (Exp. 3), or when neither does (Exp. 1)

Results



Neither
Only prime
Both
Only target
Only target (emb. wh-qs)
Lexical boost = 9.3%
Lexical boost = — 2.9%
Lexical boost = 14.8%
Lexical boost = — 0.2%
Lexical boost = — 1.0%
Experiment 1
Experiment 2
Experiment 3
Experiment 4
Experiment 5
% That production
Verb Type
Different
Same
Two-way interaction: 95%CrI: [-1.13, -0.09]
Two-way interaction: 95%CrI: [-1.57, -0.33]
Null
That
Prime Type