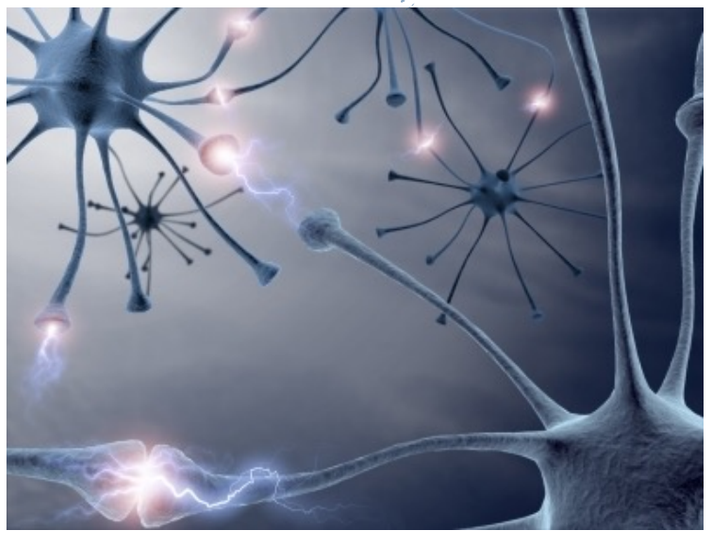
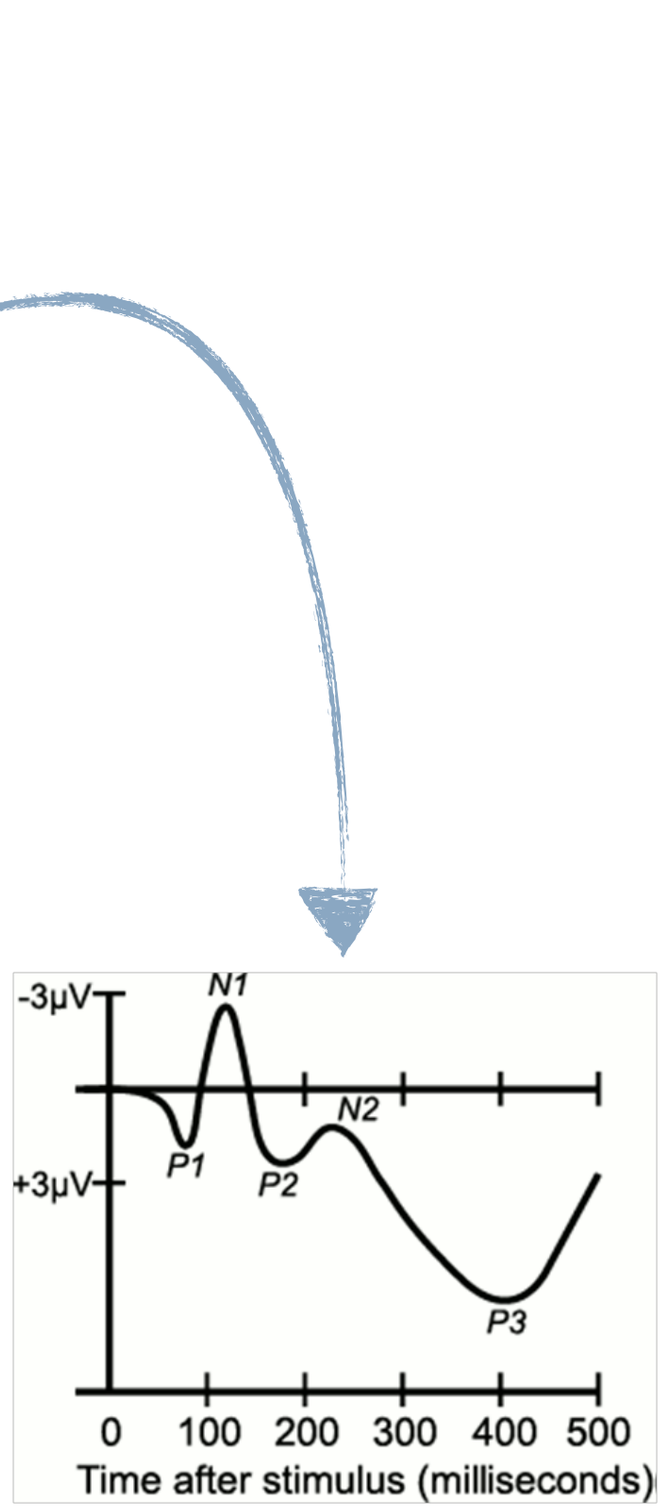
Connectionist Language Processing

Lecture 9: **Modeling the Electrophysiology of Language Comprehension**

Matthew W. Crocker
crocker@coli.uni-sb.de
Harm Brouwer
brouwer@coli.uni-sb.de

Event-Related Potentials (ERPs)

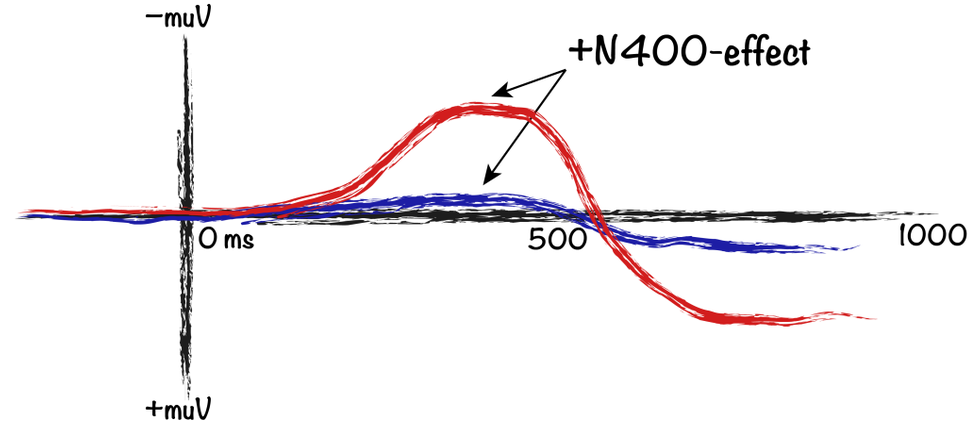




N400

He spread the warm bread with <u>socks</u>

He spread the warm bread with <u>butter</u>

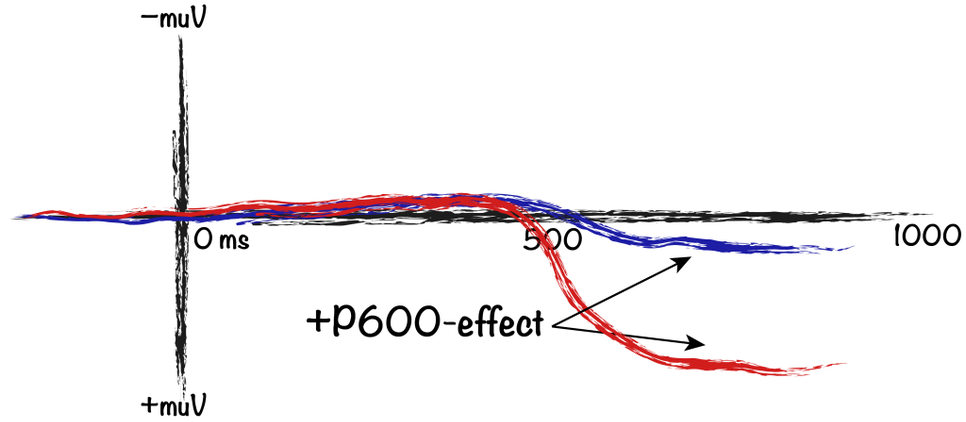


Kutas & Hillyard (1980)
Science

P600

The spoilt child <u>throw</u> the toys on the floor

The spoilt child <u>throws</u> the toys on the floor



Hagoort et al. (1993)
Lang. Cognitive Proc.

Theories of the N400 and P600

Problem: The N400 and P600 are the most salient language-sensitive ERP components, but their functional roles are not agreed upon

N400:

semantic integration

(e.g., Osterhout & Holcomb, 1992)

retrieval/activation of meaning

(e.g., Kutas & Federmeier 2000; Lau et al., 2008)

pre-activation and unification

(Baggio & Hagoort, 2011)

semantic inhibition

(Debruille, 2007)

P600:

syntactic repair/reanalysis

(e.g., Hagoort et al., 1993)

syntactic integration difficulty

(e.g., Kaan et al., 2000; Kaan and Swaab,2003)

conflict resolution

(e.g., Kolk et al., 2003; Kuperberg, 2007)

semantic integration

(Brouwer et al., 2012)

Why is it difficult to decide? Processing models are typically conceptual models, lacking the detail required for empirical (in)validation

Solution: Explicit computational models —> **quantitative predictions**

REQ: An integrated theory of the N400/P600 in language processing

From the standard view

N400 —> semantic integration

P600 —> syntactic processing

To the Retrieval-Integration account

N400 —> ~~semantic integration~~ —> lexical retrieval

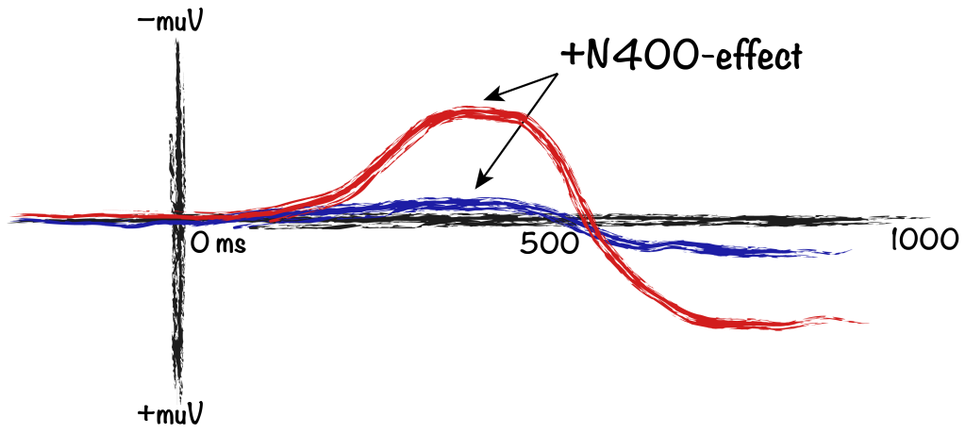
P600 —> ~~syntactic processing~~ —> semantic integration

Next: Derive an explicit neurocomputational model of this account

N400 as Semantic Integration

He spread the warm bread with <u>socks</u>

He spread the warm bread with <u>butter</u>



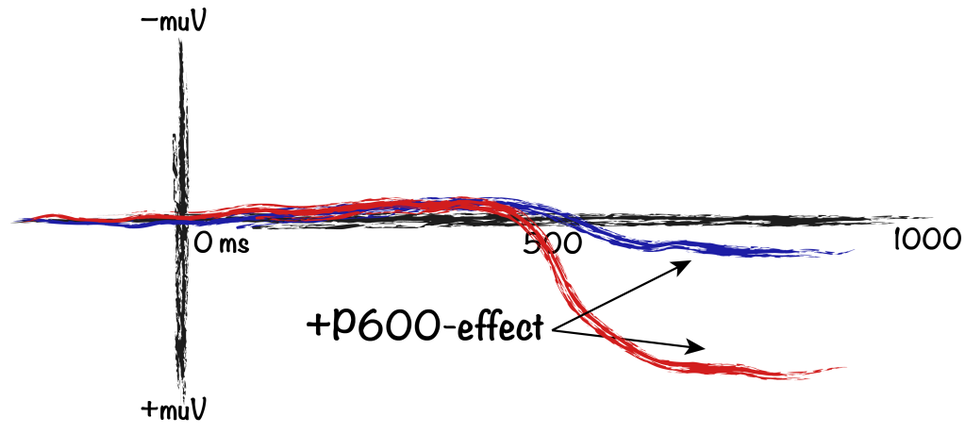
↑N400: effort involved in updating utterance representation with meaning of ‘socks’ relative to ‘butter’

Kutas & Hillyard (1980)
Science

P600 as Syntactic Repair/Reanalysis

The spoilt child <u>throw</u> the toys on the floor

The spoilt child <u>throws</u> the toys on the floor



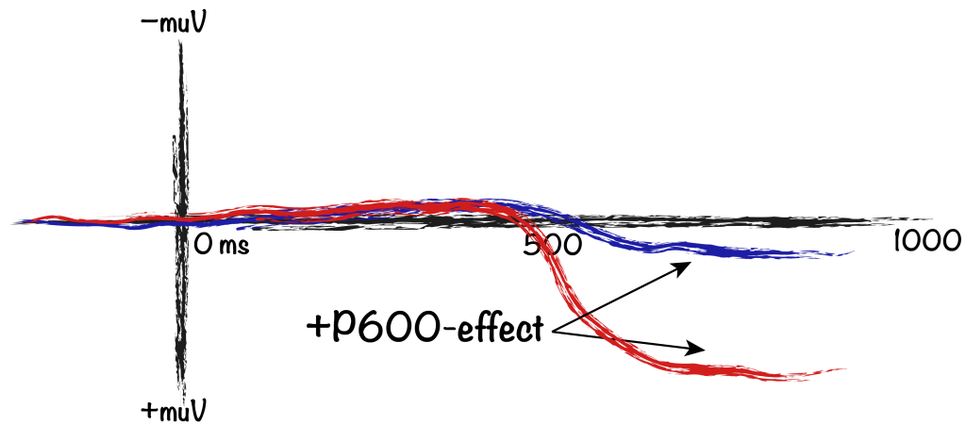
↑P600: effort involved in repairing the inflection of ‘throw’ relative to the felicitous inflection ‘throws’

Hagoort et al. (1993)
Lang. Cognitive Proc.

Puzzle: ‘Semantic P600’-effects

De speer heeft de atleten geworpen
‘The javelin has the athletes thrown’

De speer werd door de atleten geworpen
‘The javelin was by the athletes thrown’



Expected:	+ N400-effect	– P600-effect
Observed:	– N400-effect	+ P600-effect

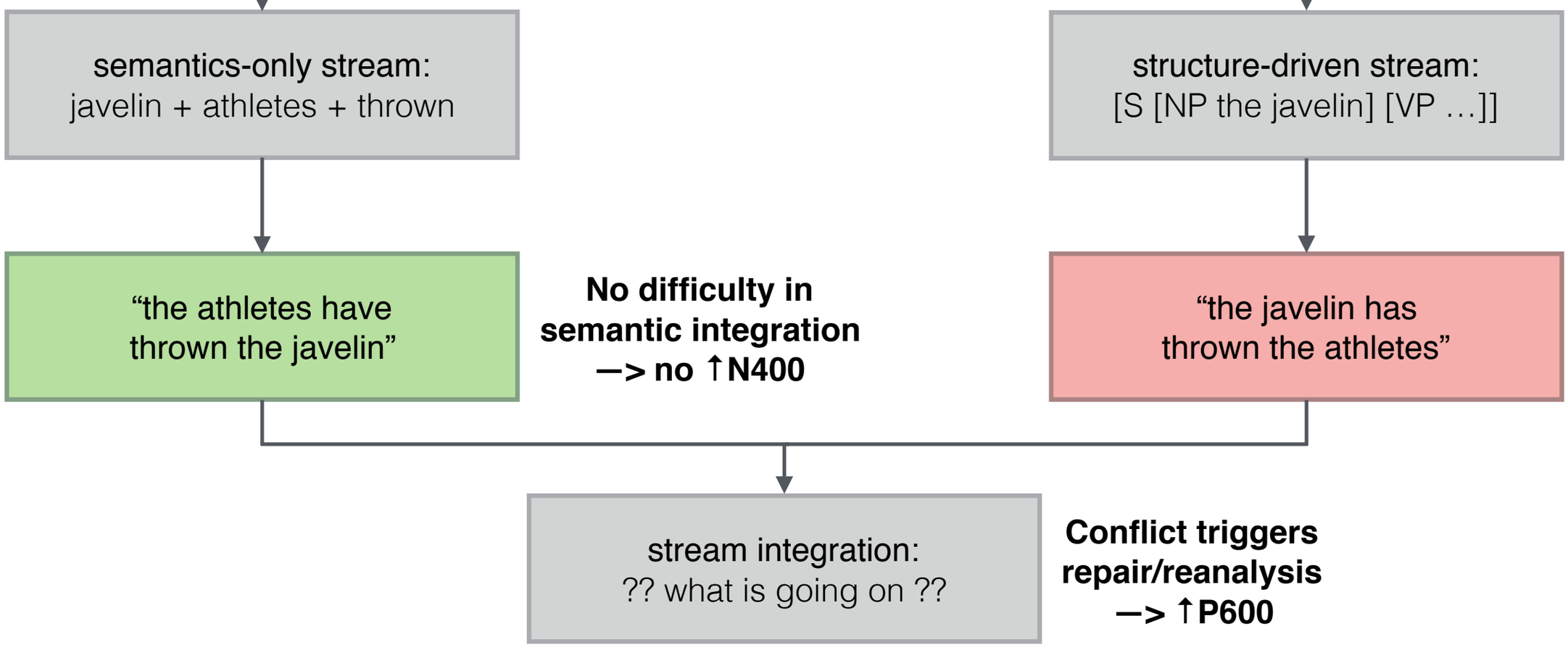
Solution: Comprehension system is tricked into a `**Semantic Illusion**’

Implication: Structure-independent semantic analysis stream

Hoeks et al. (2004)
Cogn. Brain Res.

A multi-stream explanation

De speer heeft de atleten [geworpen]
'The javelin has the athletes [thrown]'

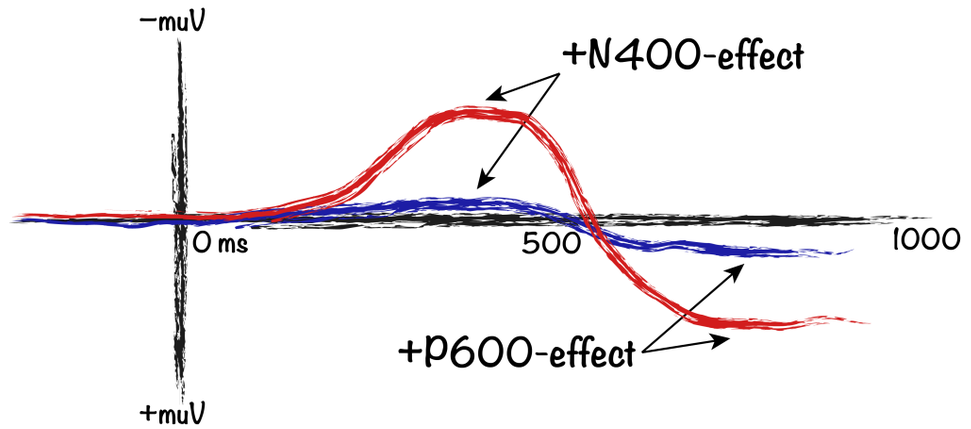


e.g., Kim & Osterhout (2005) *J. Mem. Lang.*

Problem: Biphasic N400/P600-effects

De speer werd door de atleten opgesomd
'The javelin was by the athletes summarized'

De speer werd door de atleten geworpen
'The javelin was by the athletes thrown'



Observed: + N400-effect + P600-effect

Multi-stream models predict an N400-effect only:

> Semantics-only stream: [no plausible analysis] —> ↑N400
> Structure-driven stream: [no plausible analysis]
> Integration of streams: [no conflict] —> no ↑P600

Q: Architectural deficit? Or wrong interpretations of N400 and P600?

Hoeks et al. (2004)
Cogn. Brain Res.

Brouwer et al. (2012)
Brain Res.

N400 as Lexical Retrieval

The "N400 ~ Retrieval" hypothesis

N400 reflects the retrieval of word meaning from long-term memory, a process that is facilitated if (part of) this meaning is already pre-activated due to lexical or contextual priming

He spread his warm bread with [socks]
He spread his warm bread with [butter]
(Kutas & Hillyard, 1980)

+ N400-effect

'The javelin has the athletes [thrown]'
'The javelin was by the athletes [thrown]'
(Hoeks et al., 2004)

– N400-effect

Kutas and Federmeier (2000, 2011)
Trends Cogn. Sci.; Annu. Rev. of Psychol.

Van Berkum (2009)
In Sauerland, U. and Yatsushiro, K. (eds.)

Lau et al. (2008)
Nat. Rev. Neurosci.

Q: But then what about **Semantic Integration**?

P600 as Semantic Integration

The "P600 ~ Integration" hypothesis

P600 is a family of late positivities that reflect the word-by-word construction, reorganization, or updating of an utterance meaning representation (with the meaning of an incoming word)

Utterance (re)composition is effortful when e.g.:

> New discourse entities require accommodation [**referent introduction**]

> Entity relations need to be established/revised [**thematic role assignment**]

> The current interpretation needs to be reorganized [**garden-paths**]

> Syntactic violations render the interpretation unclear [**agreement errors**]

> The constructed interpretation is not straightforwardly meaningful [**irony**]

> The interpretation conflicts with world knowledge [**'Semantic Illusions'**]

Implication: Biphasic N400/P600 "Retrieval-Integration" cycles

Brouwer et al. (2012) *Brain Res.*

Brouwer et al. (2017) *Cognitive Sci.*

The Retrieval-Integration account

Retrieval-Integration account of the N400 and P600

~N400: Every word modulates N400 amplitude, reflecting **retrieval** of its associated conceptual knowledge from long-term memory

~P600: Every word modulates P600 amplitude, reflecting **integration** of its retrieved meaning into the unfolding utterance representation

The Retrieval-Integration account as a processing 'model':

> **Single-stream model:** No need for a semantics-only processing stream

> **Reverberating dynamics:** Integration and Retrieval are interdependent

> **Qualitative predictions:** Broadest empirical coverage of extant models

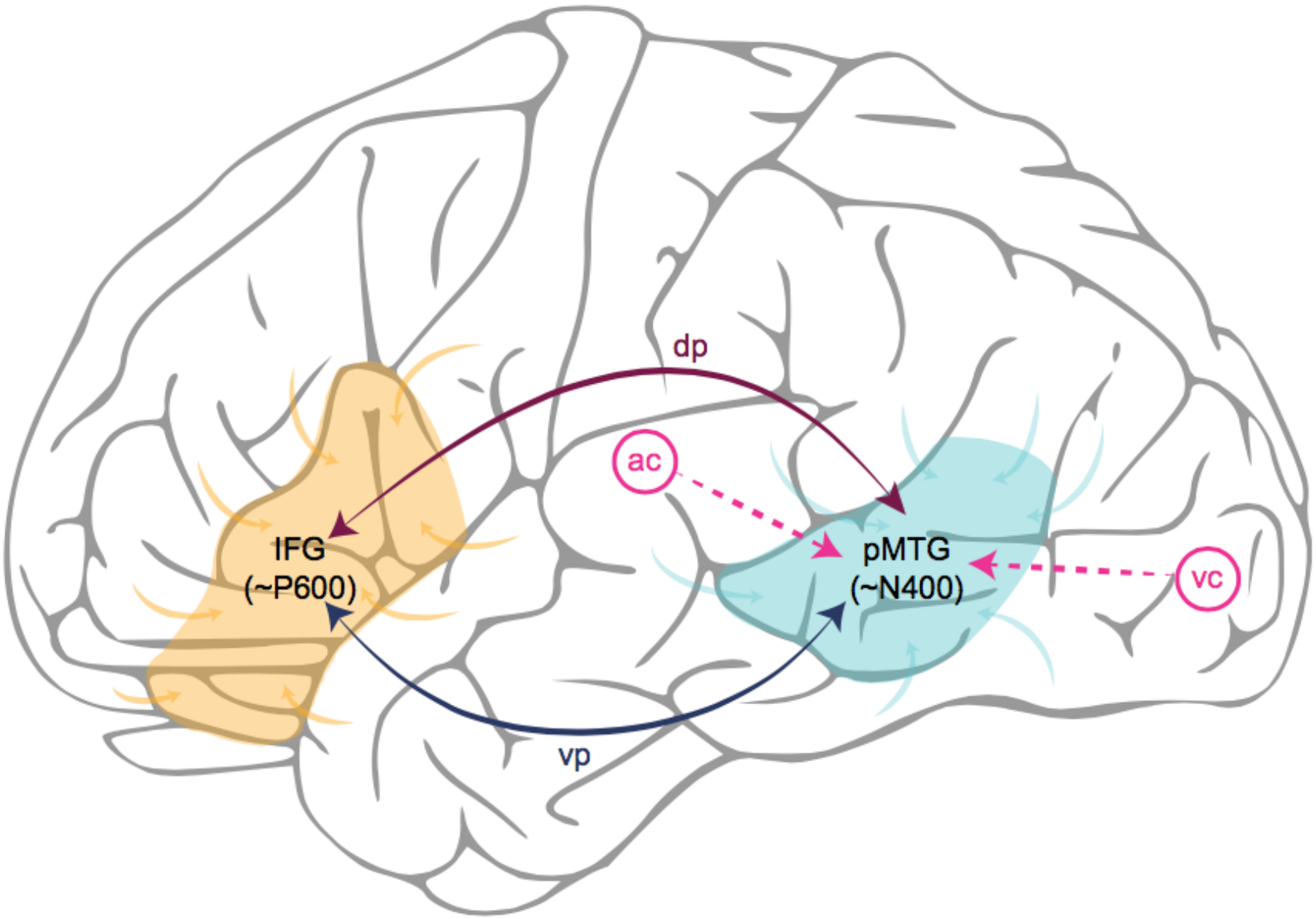
> **Architectural precision:** A conceptual 'box-and-arrow' model

Next: Instantiate the account as a neurocomputational model

Brouwer et al. (2012) *Brain Res.*

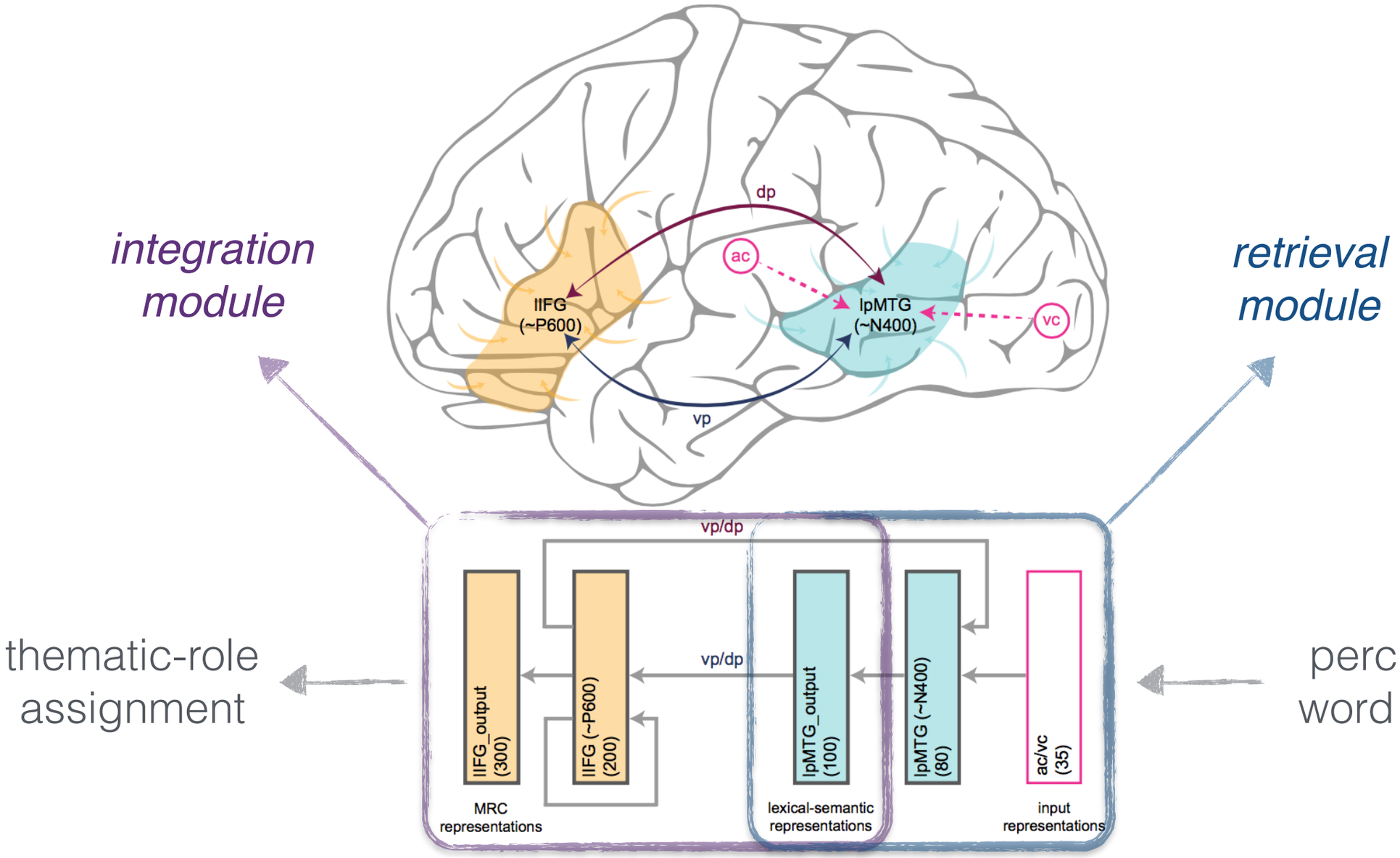
Brouwer et al. (2017) *Cognitive Sci.*

Aligning Electrophysiology and Neuroanatomy



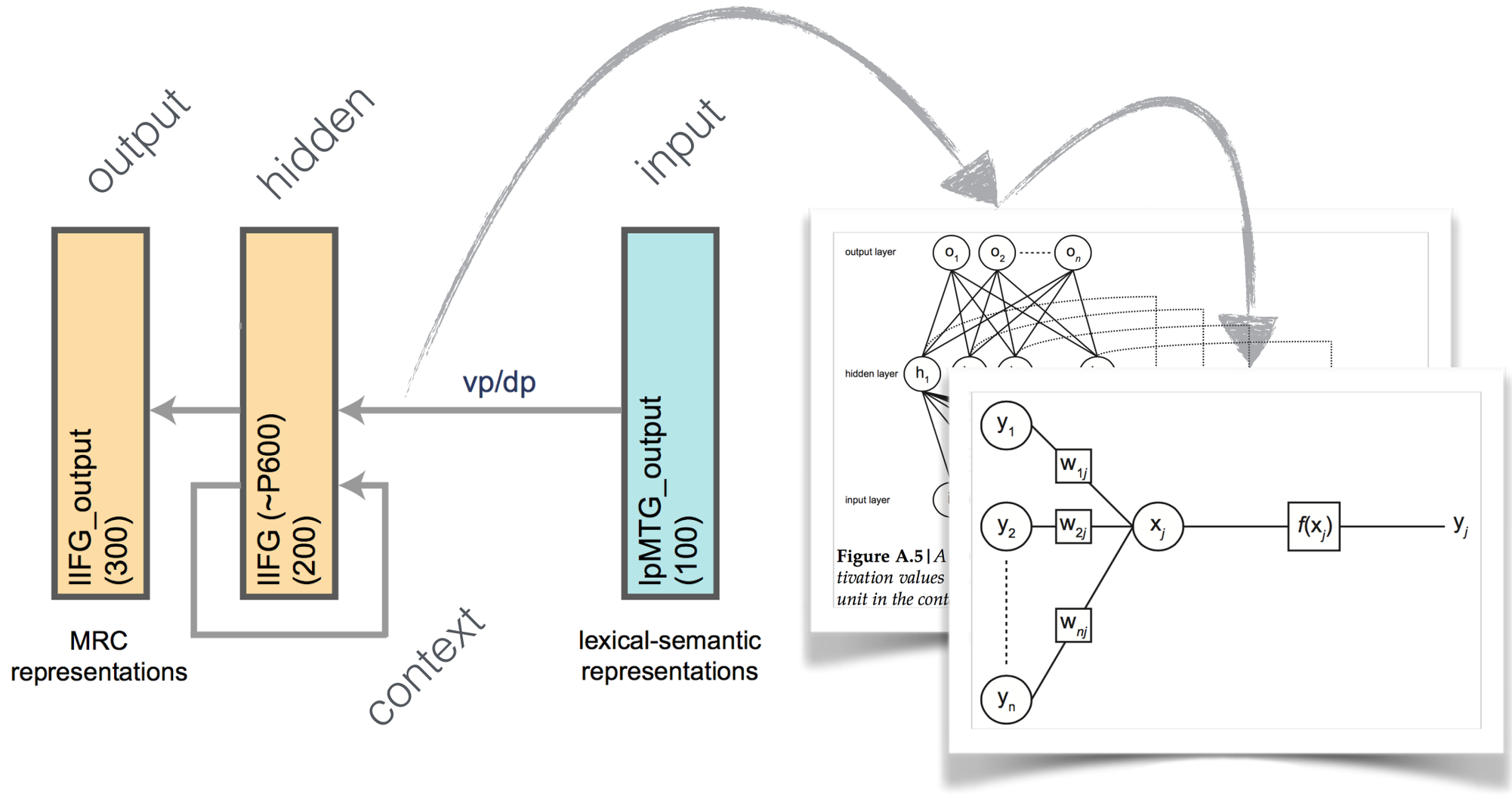
Brouwer and Hoeks (2013)
Front. Hum. Neurosci.

A Neurocomputational Model



Brouwer et al. (2017)
Cognitive Sci.

(De)constructing the Integration Module



> IM is an SRN that transforms sequences of **lexical-semantic representations** (word meanings) into an **utterance interpretation** (thematic-role assignment)

Brouwer et al. (2017)
Cognitive Sci.

Lexical-semantic representations (word meanings)

"In many of the **most influential theories of word meaning** and of concepts and categorization, **semantic features** have been used as their **representational currency**. For example, classical, prototype, and exemplar theories of categorization and conceptual representation all make use of features (Medin & Schaffer, 1978; Minda & Smith, 2002; Smith & Medin, 1981), as do network models of semantic memory and language processing (Collins & Loftus, 1975)."

> **lexical-semantic representations** will be modeled as **semantic feature vectors**

McRae et al. (2005)
Beh. Res. Meth. Instr. and Comp.

Integration System—Representations

Lexical-semantic representations (word meaning)

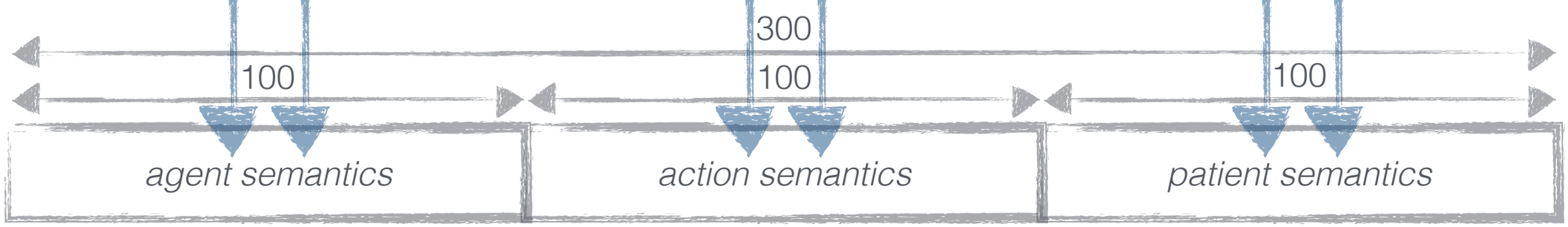
> corpus-derived 100-dimensional, binary feature vectors (using COALS)

cat	1100110011100001011100010001001010110100010101011101101011001010111111001110001010001010110111100101
dog	1100000001100001011110110001010100110000011111110101110000011010110000011101011101100101100101111101
walk	1101110100000110000100101111010101111101111101001101111010110010111011010111110010010000100111101100
eat	1101010000010111100000001111001000100011101100000111111111100110111000010010111011000001110110011111
food	1101110000111011110010011110001111100110111111000100011011111100100100100000001010001101000011000111

Rohde et al. (under revision)

Utterance interpretations (thematic role assignments)

> 300-dimensional thematic role assignment (*agent-action-patient*) vectors



cf. Mayberry et al. (2009). *Cogn. Sci.*

Brouwer et al. (2017)
Cognitive Sci.

Zooming in: The COALS model

Correlated Occurrence Analogue to Lexical Semantics (COALS)

Step 1—construct a co-occurrence matrix, using a 4-word ramped window: **1 2 3 4 [word] 4 3 2 1**

Step 2—convert weighted co-occurrence frequencies to pairwise correlations:

$$w'_{a,b} = \frac{T \cdot w_{a,b} - \sum_j w_{a,j} \cdot \sum_i w_{i,b}}{(\sum_j w_{a,j} \cdot (T - \sum_j w_{a,j}) \cdot \sum_i w_{i,b} \cdot (T - \sum_i w_{i,b}))^{\frac{1}{2}}} \quad \text{where} \quad T = \sum_i \sum_j w_{i,j}$$

Step 3—"normalize" correlations:

$$norm(w'_{a,b}) = \begin{cases} 0 & \text{if } w'_{a,b} < 0 \\ \sqrt{w'_{a,b}} & \text{otherwise} \end{cases}$$

(reduces distance between small and large correlations)

Step 4—reduce dimensionality with Singular Value Decomposition (SVD):

$$\hat{X}_{15000\times14000} = \hat{U}_{15000\times100}\hat{S}_{100\times100}\hat{V}^T_{100\times14000}$$

Step 5—Extract COALS vector for each word:

$$V_c = X_c\hat{V}\hat{S}^{-1}$$

(and set positive units to 1 and negative values to 0 to obtain binary vectors)

Rohde et al. (under revision)

Integration Module—Training

> The IM is trained to comprehend Dutch sentences with the following structure:

Active sentences:

De	[AGENT]	heeft	het/de	[PATIENT]	[ACTION]
The	[AGENT]	has	$\text{the}_{(+/-\text{NEUTER})}$	[PATIENT]	[ACTION]

Passive sentences:

De	[PATIENT]	werd	door	het/de	[AGENT]	[ACTION]
The	[PATIENT]	was	by	$\text{the}_{(+/-\text{NEUTER})}$	[AGENT]	[ACTION]

and it learns that:

every NP can be an Agent or a Patient: people can construct an interpretation for “The bread bakes the baker” (think about a typical Disney film, for instance)

minimal world knowledge: certain Agent-Action-Patient configurations are more likely than others (a baker is more likely to bake a bread than a ball; cf. Mayberry et al. 2009)

> after training, the comprehension accuracy of the IM is perfect

Brouwer et al. (2017)
Cognitive Sci.

Zooming in: Training the model

The model was trained using backpropagation and bounded gradient descent

The sum squared error of the model was minimized:

$$E_c = \frac{1}{2}\sum_j (y_j - d_j)^2$$

by iteratively adjusting weights on the basis of “bounded” weight delta’s:

$$\Delta w_{ij}(t) = -\varepsilon\rho\frac{\partial E}{\partial w_{ij}} + \alpha\Delta w_{ij}(t-1)$$

where ε = learning rate; $\rho = \begin{cases} \frac{1}{||\partial E/\partial w||} & \text{if } ||\partial E/\partial w|| > 1 \\ 1 & \text{otherwise} \end{cases}$ = a scaling factor;

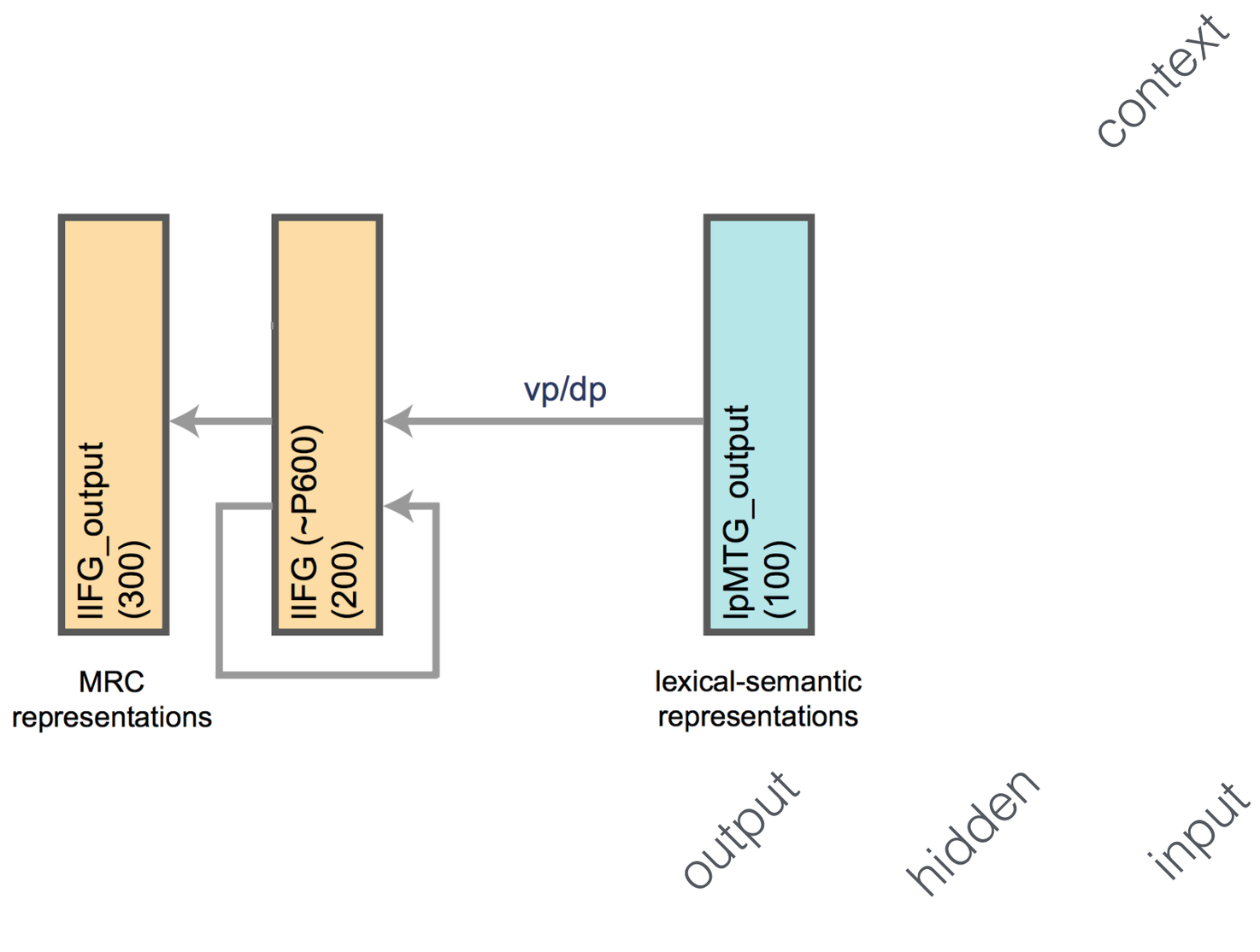
$\frac{\partial E}{\partial w_{ij}} = \delta_j y_i$ = weight gradient where for output units: $\delta_j = (y_j - d_j)(y_j(1-y_j)+0.1)$

for hidden units: $\delta_j = (y_j(1-y_j)+0.1)\sum_k \delta_k w_{jk}$

and α = momentum coefficient;

Rohde (2002), *PhD thesis*
Rumelhart et al. (1986) *Nature*

(De)constructing the Retrieval Module



> RM activates the lexical-semantic representations (word meanings) corresponding to incoming **acoustic/orthographic representations** (perceived word forms), while taking the unfolding utterance representation (context) into account

Brouwer et al. (2017)
Cognitive Sci.

Retrieval Module—Representations and Training

Acoustic/orthographic representations

> 35-dimensional, localist vectors (each neuron encodes a single word)

painting	00000000000000000**1**00000000000000
the	000000**1**00000000000000000000000000
walked	0000000000001000000000000000000000

(this scheme rules out N400-effects due to orthographic neighbourhood size; see, e.g., Laszlo & Federmeier, 2011)

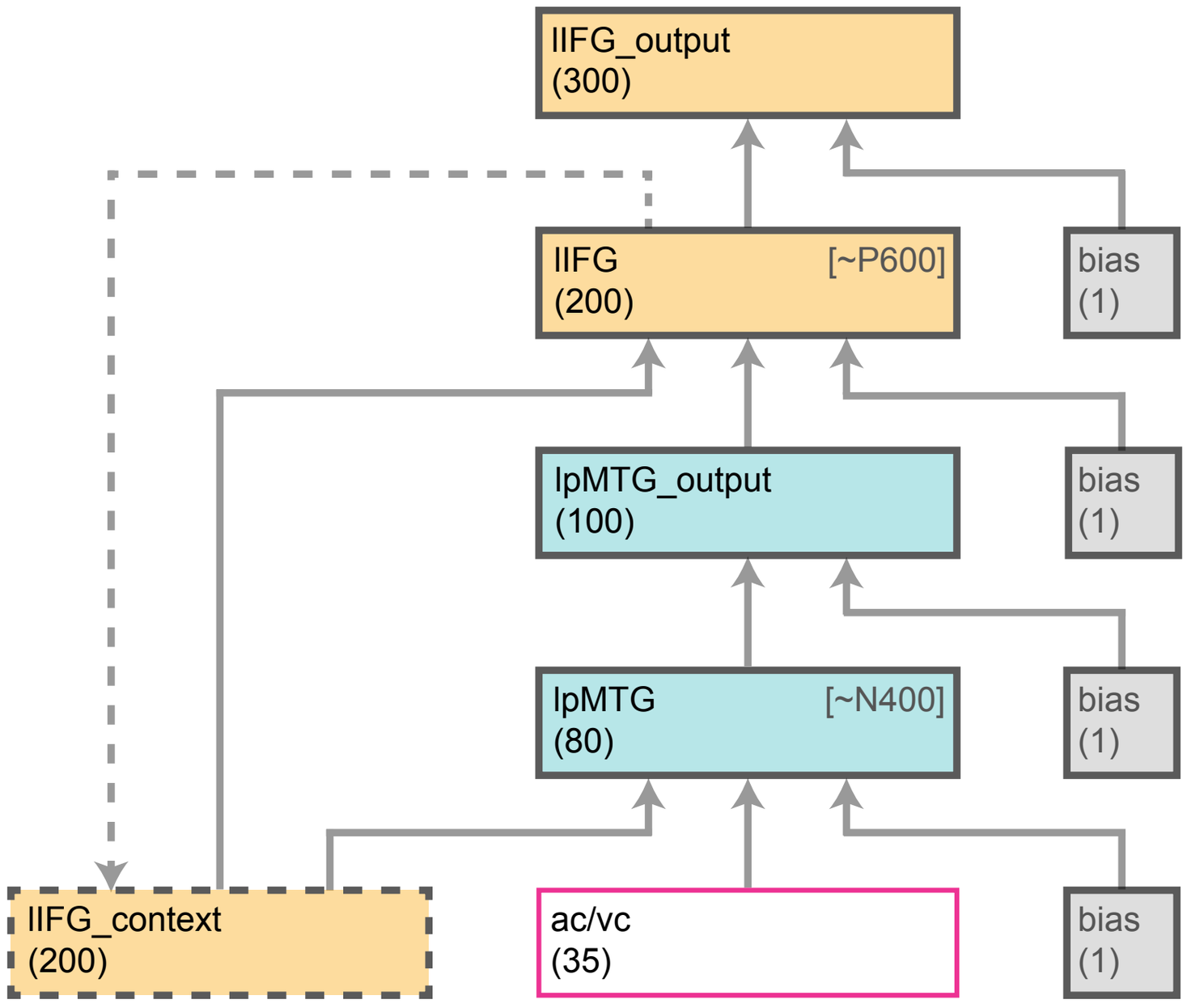
> the full model (RM+IM) is trained on the same sentences as the IM, but the inputs are now acoustic/orthographic (rather than lexical-semantic) representations

> critically, the **weights in the IM are frozen**, such that the RM is forced to engage in context-sensitive retrieval (i.e., take the current utterance representation into account)

> after training, the comprehension accuracy of the entire model (RM+IM) is perfect

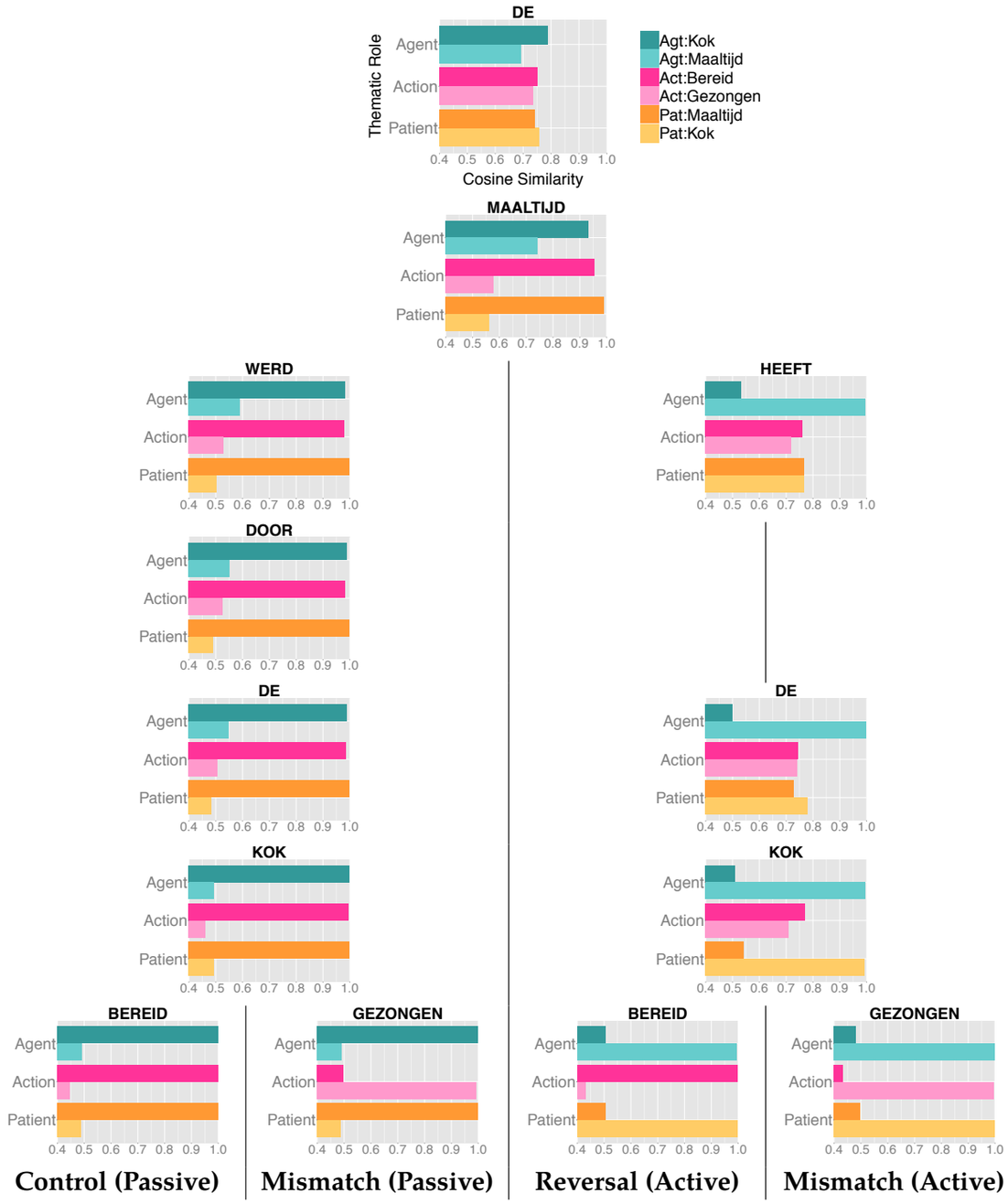
Brouwer et al. (2017)
Cognitive Sci.

Zooming out: Full model architecture



Brouwer et al. (2017)
Cognitive Sci.

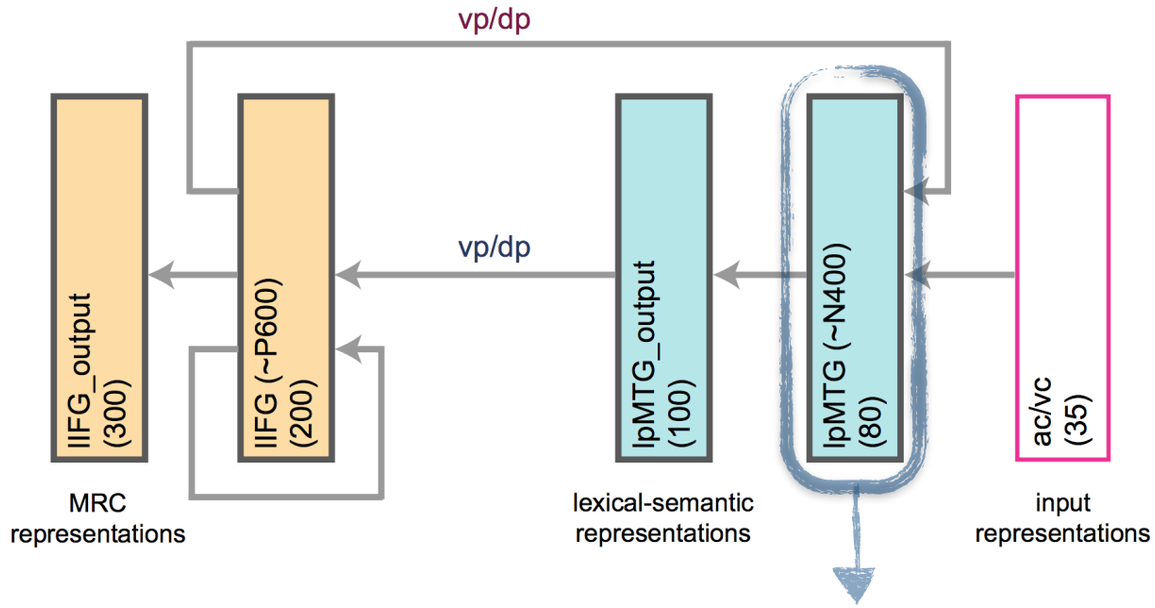
Processing in the model



Brouwer et al. (2017)
Cognitive Sci.

Linking Hypothesis—N400 component

"N400 amplitude is a measure of 'unpreparedness'. If no features relevant to an incoming word are pre-activated, N400 amplitude will be maximal; if the lexical-semantic features of an incoming word are consistent with those pre-activated in memory, N400 amplitude will be reduced. **Hence, N400 amplitude is a measure of how much the activation pattern in memory changes due to the processing of an incoming word. As such, we compute the correlates of N400 amplitude at the lpMTG layer, where the activation of lexical-semantic features takes place (~memory retrieval), as the degree to which the pattern of activity induced by the current word, and that induced by the previous word are different.**"



$$\mathrm{N400} = 1 - cos(lpMTG_t, lpMTG_{t-1})$$

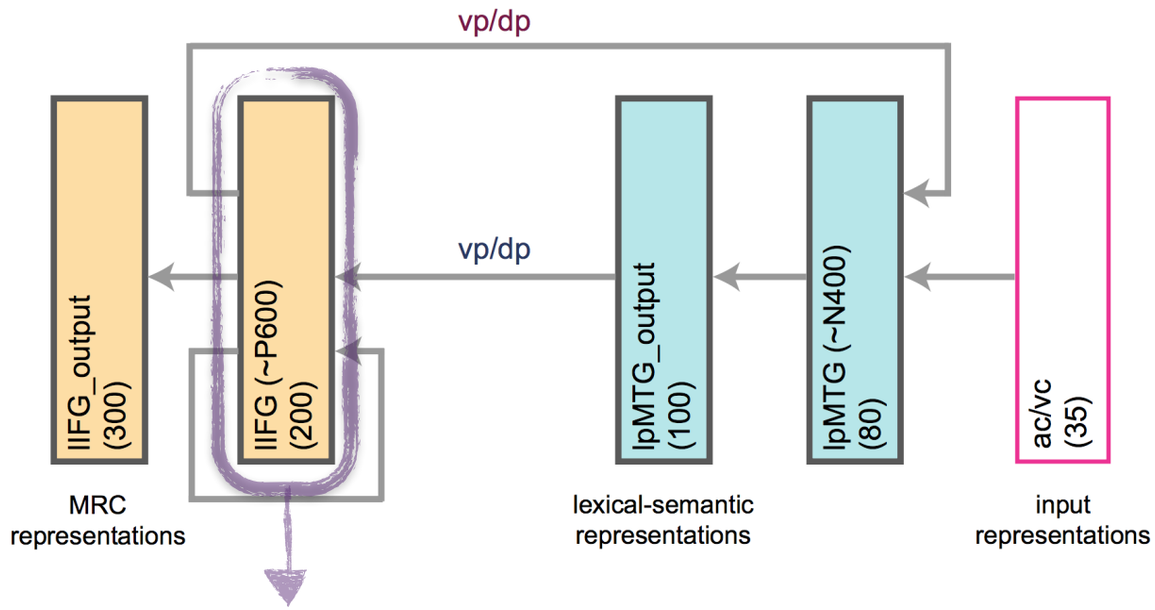
(if no difference cos(x,y) = 1; otherwise: 0 > cos(x,y) <1)

Brouwer (2014)
PhD thesis

Brouwer et al. (2017)
Cognitive Sci.

Linking Hypothesis—P600 component

"P600 amplitude, in turn, reflects the difficulty of establishing coherence. **The more the current [utterance interpretation] needs to be reorganized or augmented in order to become coherent, the higher P600 amplitude.** Hence, P600 amplitude is effectively a measure of how much the representation of the unfolding state of affairs changes due to the integration of an incoming word. As such, **we compute the correlates of P600 amplitude as the difference between the previous and the current state of affairs at the lIFG layer, where the (re)construction of an [utterance interpretation]—in terms of thematic-role assignment—takes place** (see also Crocker et al., 2010)."



$$\text{P600} = 1 - cos(lIFG_t, lIFG_{t-1})$$

(if no difference cos(x,y) = 1; otherwise: 0 > cos(x,y) <1)

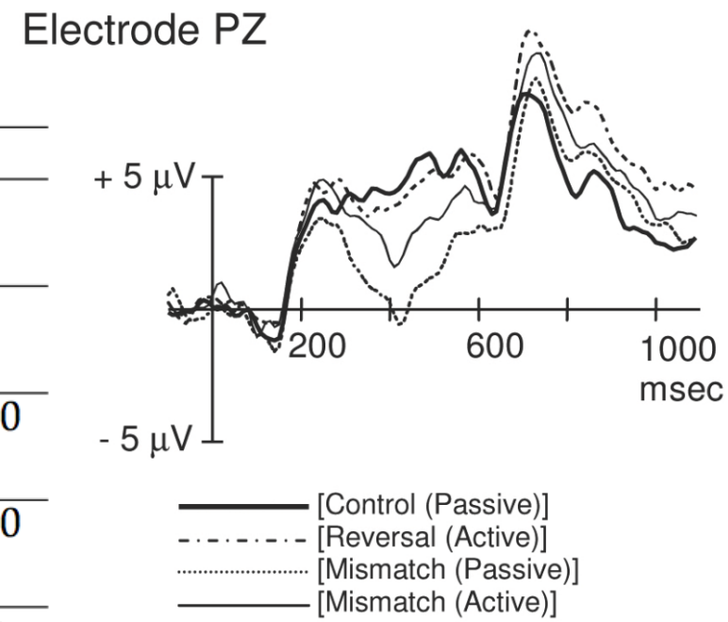
Brouwer (2014)
PhD thesis

Brouwer et al. (2017)
Cognitive Sci.

Simulating an ERP experiment

Item	Condition	Effect
De speer werd door de atleten geworpen *The javelin was by the athletes thrown*	Control (Passive)	—
De speer heeft de atleten geworpen *The javelin has the athletes thrown*	Reversal (Active)	P600
De speer werd door de atleten opgesomd *The javelin was by the athletes summarized*	Mismatch (Passive)	N400/P600
De speer heeft de atleten opgesomd *The javelin has the athletes summarized*	Mismatch (Active)	N400/P600

Hoeks et al. (2004), *Cogn. Brain. Res.*



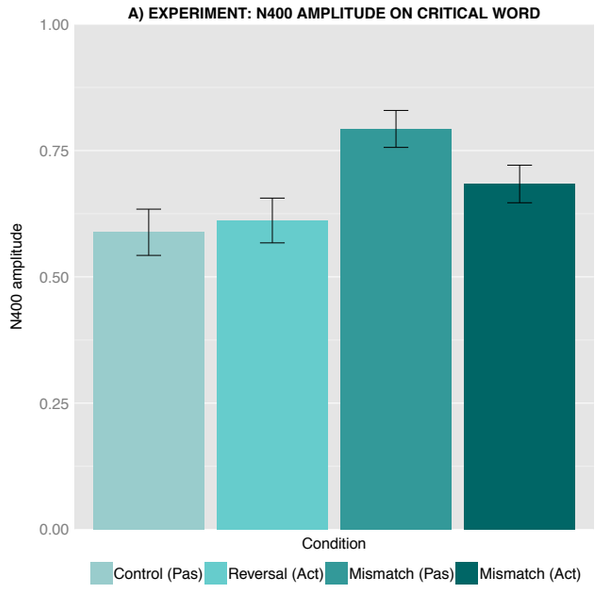
> Two simulation experiments, each with a different set of lexical items, and 10 sentences per condition

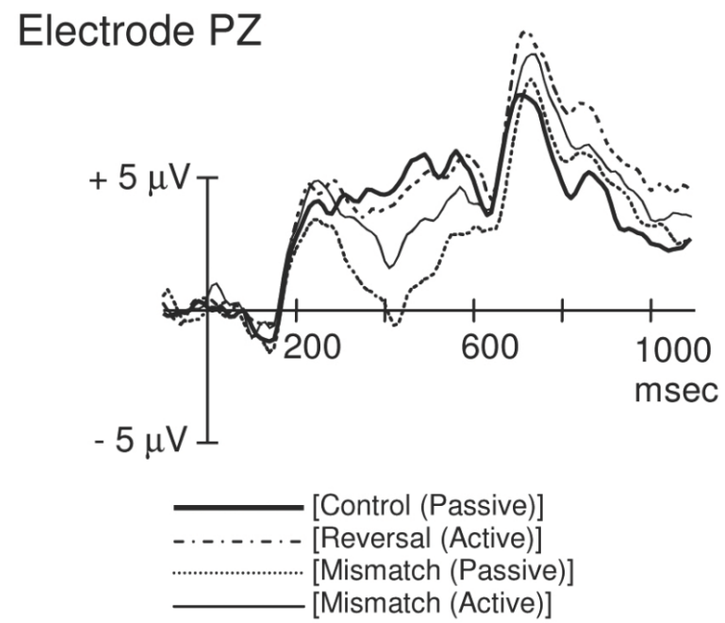
Desired N400-effects: only for mismatches relative control

Desired P600-effects: for reversal and mismatches relative control

Brouwer et al. (2017)
Cognitive Sci.

Simulation results—N400

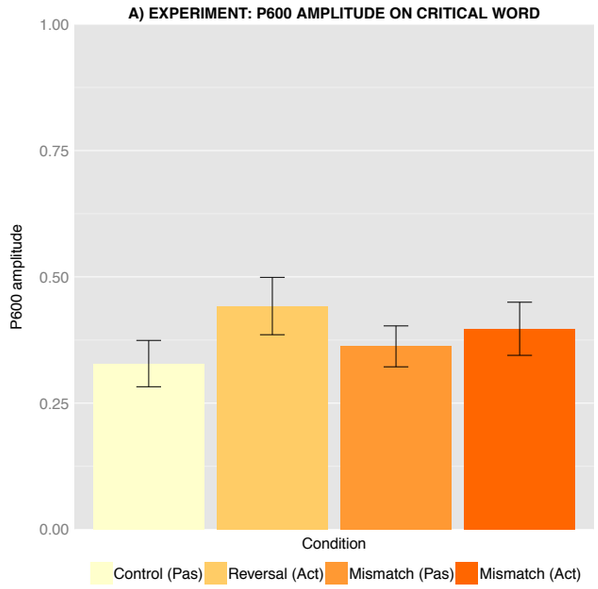


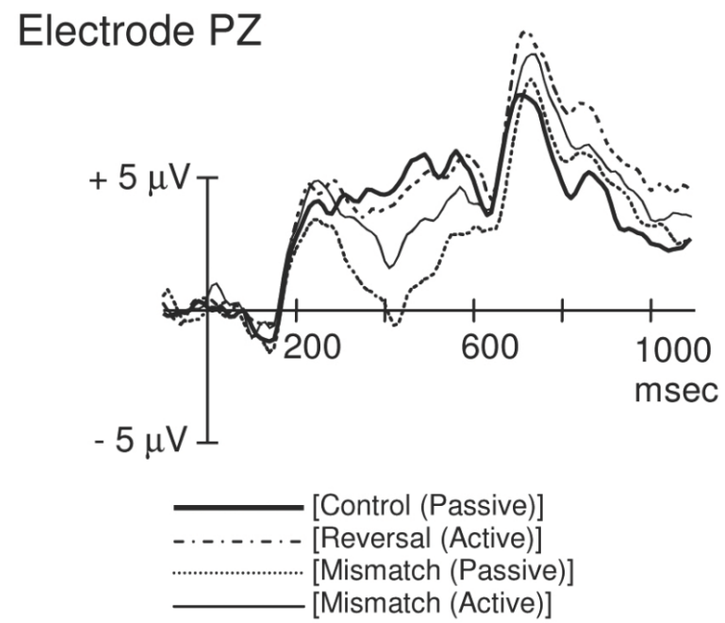


Main effect of Condition (Exp 1: F(3,27)=45.1; p<.001; Exp 2: F(3,27)=12.3; p<.001); pairwise comparisons (Bonferroni corrected): no N400-effect for reversals (Exp 1: p=.47; Exp 2: p=.91), and a significant N400-effect for the two other anomalous conditions (Exp 1: p-values<.005; Exp 2: p-values<.01).

Brouwer (2014), *PhD thesis*
Brouwer et al. (under review)

Simulation results—P600





Main effect of Condition (Exp 1: F(3,27)=136.5; p<.001; Exp 2: F(3,27)=70.1; p<.001); pairwise comparisons (Bonferroni corrected): significant P600-effect for all three anomalous conditions (Exp 1: all three p-values<.001; Exp 2: all three p-values<.001).

Brouwer (2014), *PhD thesis*
Brouwer et al. (under review)

Conclusions

> We have derived the **Retrieval-Integration account of the N400 and the P600**

> Instantiated it as a **neurocomputational model of language comprehension**

> Proposed **explicit and scalable linking hypotheses** to electrophysiology:

N400 —> Retrieval

P600 —> Integration

> The model accounts for signature semantically induced N400 and P600 modulation patterns

Simulation Materials

Exp.	Agent	Patient	NEUTER	Action	Mismatch
1	voetballer *soccer player*	doelpunt *goal*	+	gescoord *scored*	gediend *served*
1	militair *soldier*	land *country*	+	gediend *served*	gescoord *scored*
1	kok *cook*	maaltijd *meal*	-	bereid *prepared*	gezongen *sung*
1	zanger *singer*	lied *song*	+	gezongen *sung*	bereid *prepared*
1	advocaat *lawyer*	bedrijf *company*	+	aangeklaagd *sued*	gelopen *ran*
1	atleet *athlete*	marathon *marathon*	-	gelopen *ran*	aangeklaagd *sued*
1	politicus *politician*	debat *debate*	+	gevoerd *engaged*	uitgegeven *published*
1	uitgever *publisher*	roman *novel*	-	uitgegeven *published*	gevoerd *engaged*
1	arts *doctor*	diagnose *diagnosis*	-	gesteld *made*	geschilderd *painted*
1	schilder *painter*	schilderij *painting*	+	geschilderd *painted*	gesteld *made*

Exp.	Agent	Patient	NEUTER	Action	Mismatch
2	rechercheur *detective*	moord *murder case*	-	opgelost *solved*	verhoogd *raised*
2	werkgever *employer*	salaris *salary*	+	verhoogd *raised*	opgelost *solved*
2	dief *thief*	museum *museum*	+	beroofd *robbed*	getrokken *pulled*
2	tandarts *dentist*	tand *tooth*	-	getrokken *pulled*	beroofd *robbed*
2	schipper *sailor*	schip *ship*	+	aangelegd *berthed*	geregisseerd *directed*
2	regisseur *director*	film *movie*	-	geregiseerd *directed*	aangelegd *berthed*
2	piloot *pilot*	vliegtuig *airplane*	+	bestuurd *steered*	afgelegd *taken*
2	student *student*	tentamen *examen*	+	afgelegd *taken*	bestuurd *steered*
2	verzekeraar *insurer*	verzekering *insurance*	-	uitgekeerd *paid*	gereden *rode*
2	wielrenner *cyclist*	etappe *stage*	+	gereden *rode*	uitgekeerd *paid*

Active sentences:

De [AGENT] heeft het/de [PATIENT] [ACTION]

The [AGENT] has the$_{(+/-\text{NEUTER})}$ [PATIENT] [ACTION]

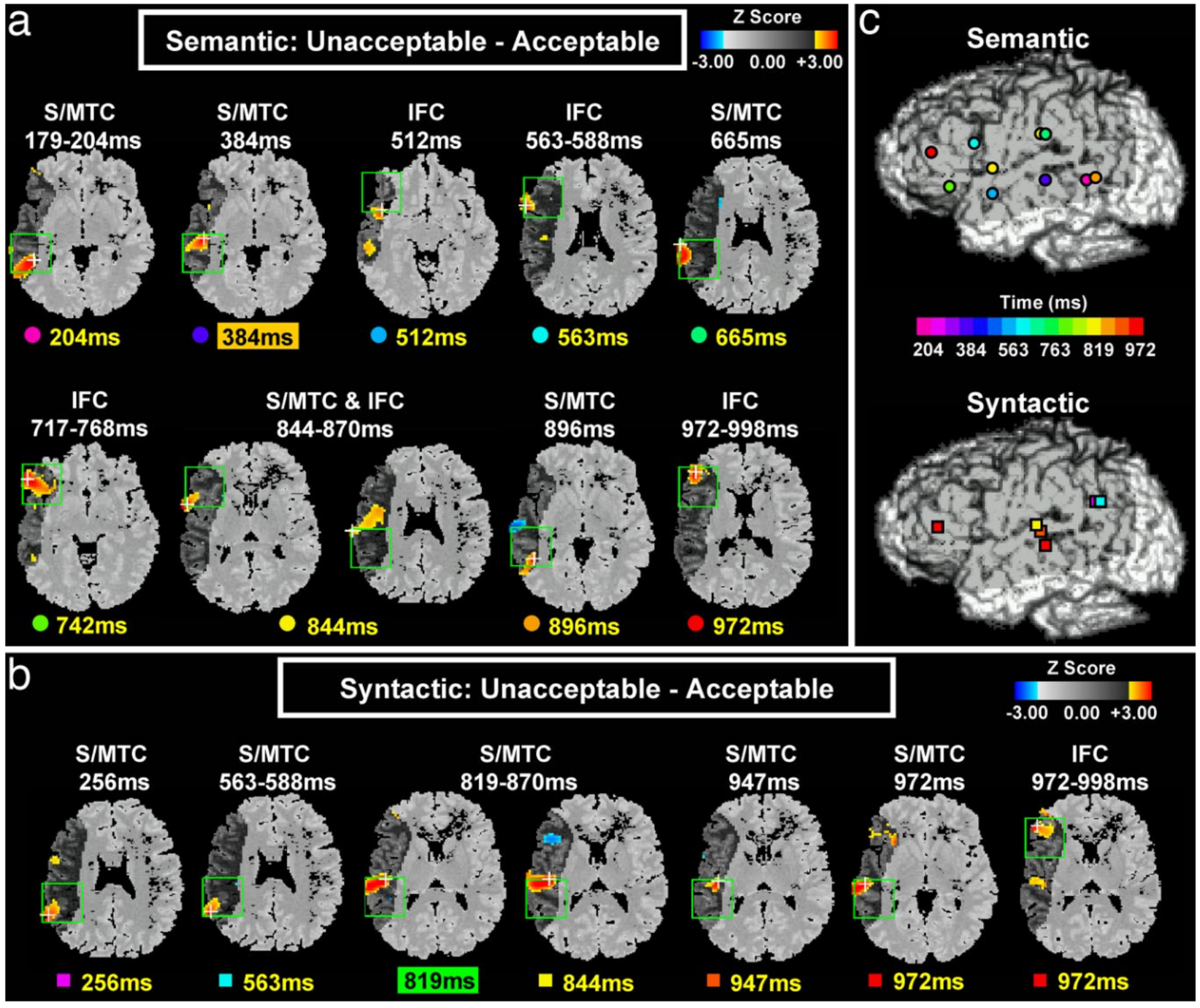
Passive sentences:

De [PATIENT] werd door het/de [AGENT] [ACTION]

The [PATIENT] was by the$_{(+/-\text{NEUTER})}$ [AGENT] [ACTION]

Brouwer (2014), *PhD thesis*
Brouwer et al. (under review)

lIFG/lpMTG communication



Tse et al. (2007), *PNAS*