LISA

Linguistically-Informed Self-Attention for Semantic Role Labeling

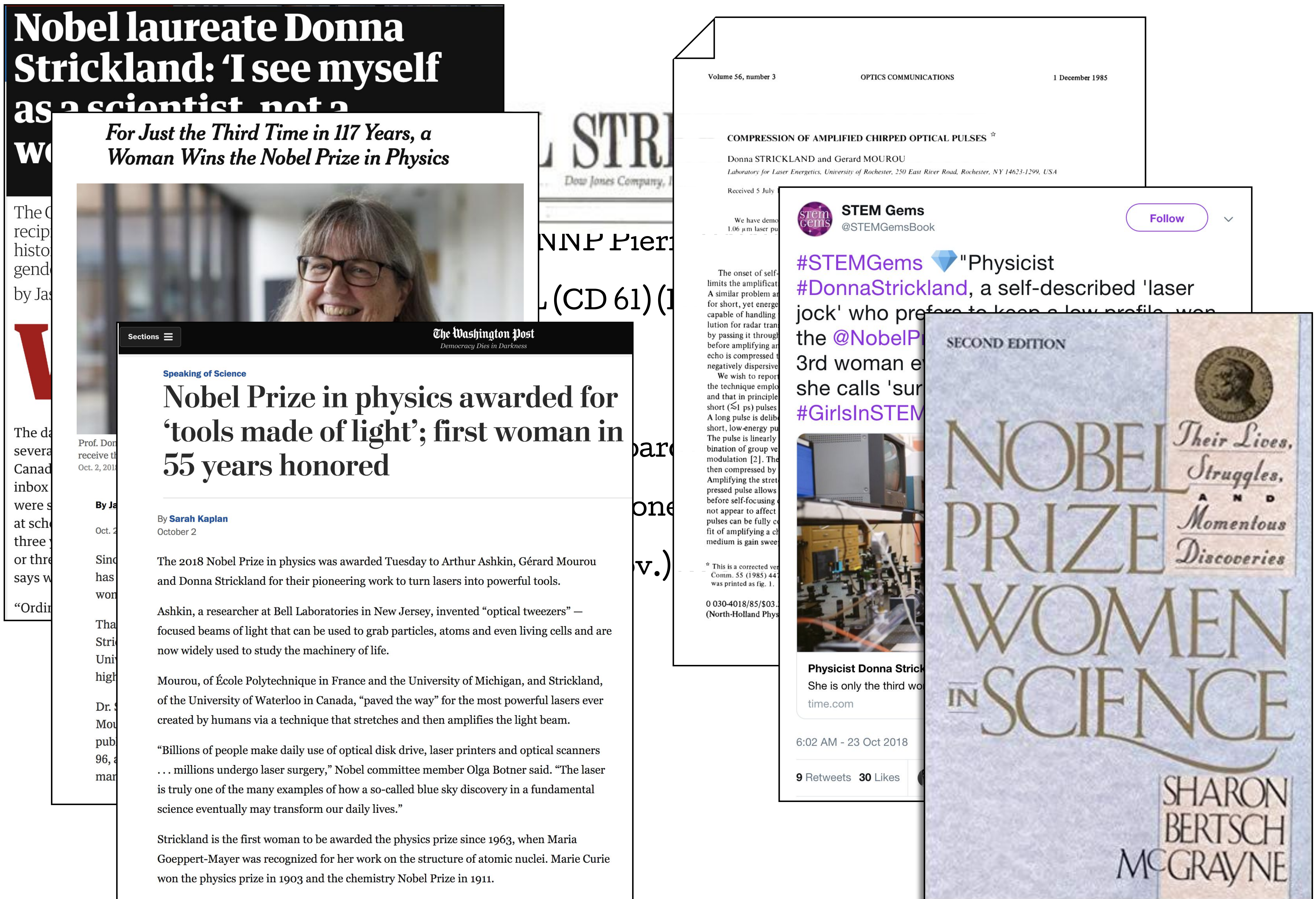
    

Emma Strubell[1] Patrick Verga[1] Daniel Andor[2] David Weiss[2] Andrew McCallum[1]

[1] UMassAmherst College of Information and Computer Sciences

[2] Google AI

Want fast, accurate, robust NLP



Speaking of Science

Nobel Prize in physics awarded for 'tools made of light'; first woman in 55 years honored

By Sarah Kaplan
October 2

The 2018 Nobel Prize in physics was awarded Tuesday to Arthur Ashkin, Gérard Mourou and Donna Strickland for their pioneering work to turn lasers into powerful tools.

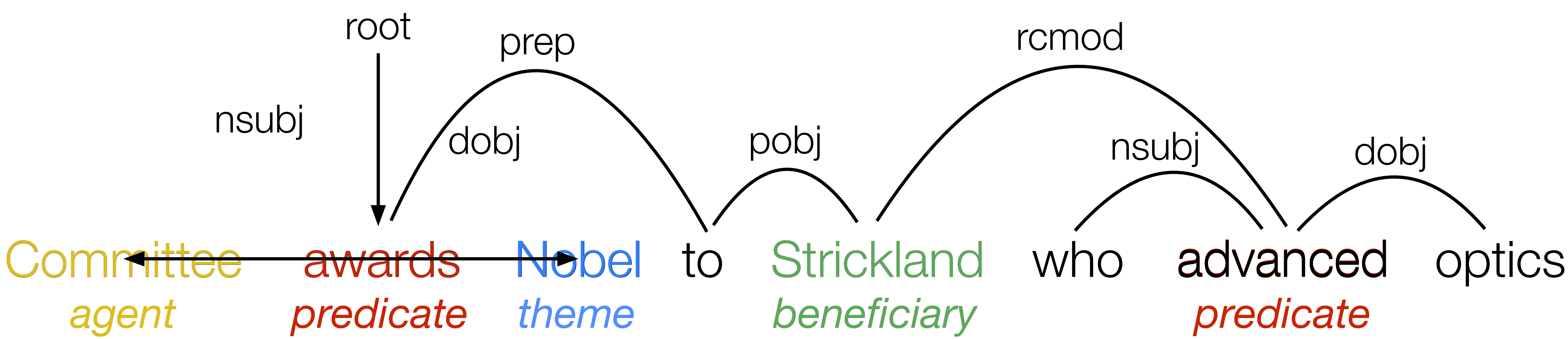
Ashkin, a researcher at Bell Laboratories in New Jersey, invented "optical tweezers" — focused beams of light that can be used to grab particles, atoms and even living cells and are now widely used to study the machinery of life.

Mourou, of École Polytechnique in France and the University of Michigan, and Strickland, of the University of Waterloo in Canada, "paved the way" for the most powerful lasers ever created by humans via a technique that stretches and then amplifies the light beam.

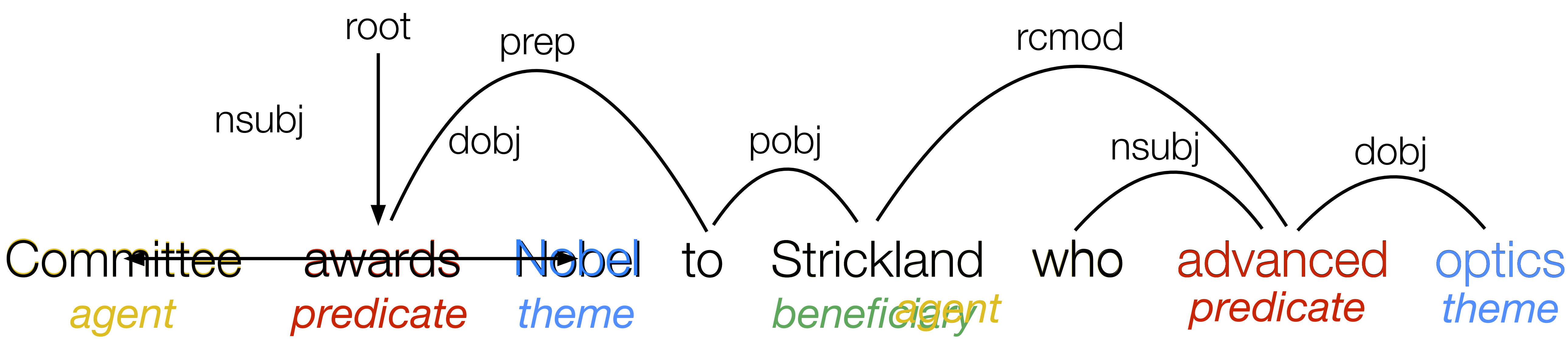
"Billions of people make daily use of optical disk drive, laser printers and optical scanners . . . millions undergo laser surgery," Nobel committee member Olga Botner said. "The laser is truly one of the many examples of how a so-called blue sky discovery in a fundamental science eventually may transform our daily lives."

Strickland is the first woman to be awarded the physics prize since 1963, when Maria Goeppert-Mayer was recognized for her work on the structure of atomic nuclei. Marie Curie won the physics prize in 1903 and the chemistry Nobel Prize in 1911.

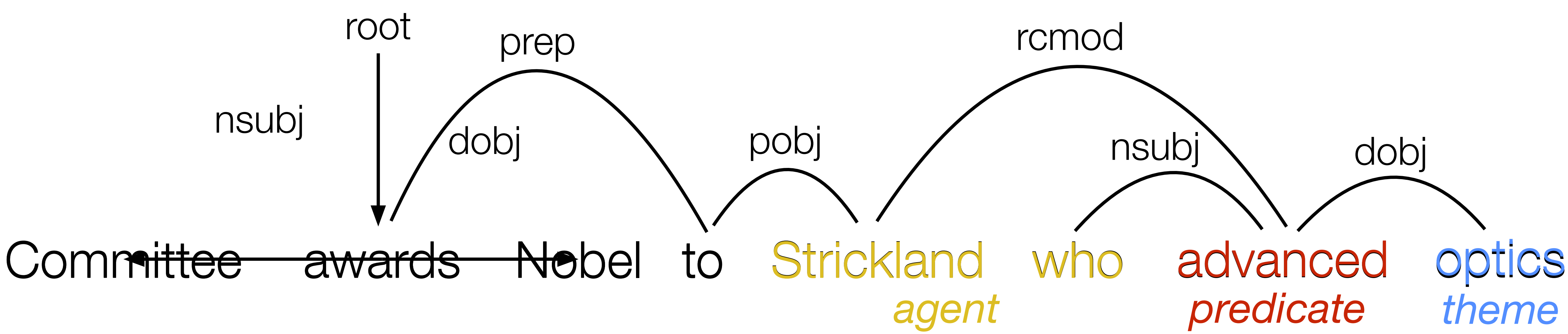
SRL: Who did what to whom?



SRL: Who did what to whom?

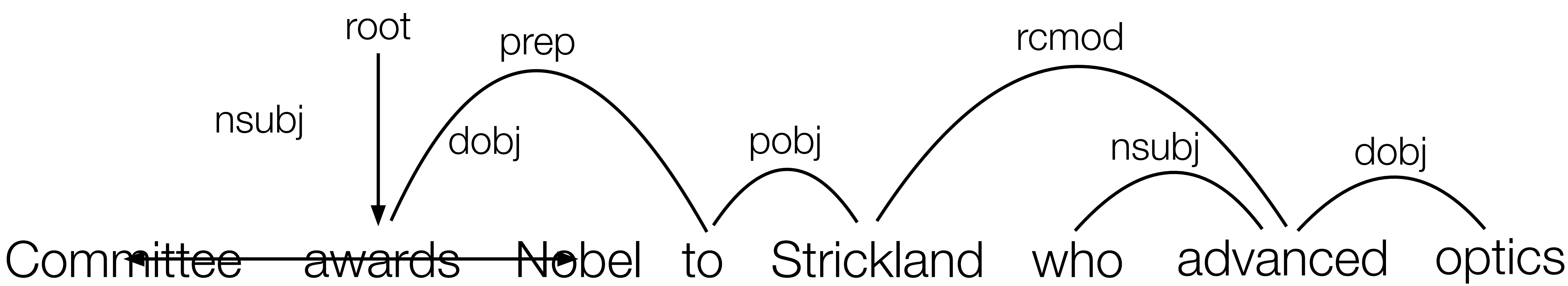


SRL: Who did what to whom?



Committee	awards	Nobel	to	Strickland	who	advanced	optics
agent	*predicate*	*theme*		*beneficiary*			

PropBank SRL: Who did what to whom?



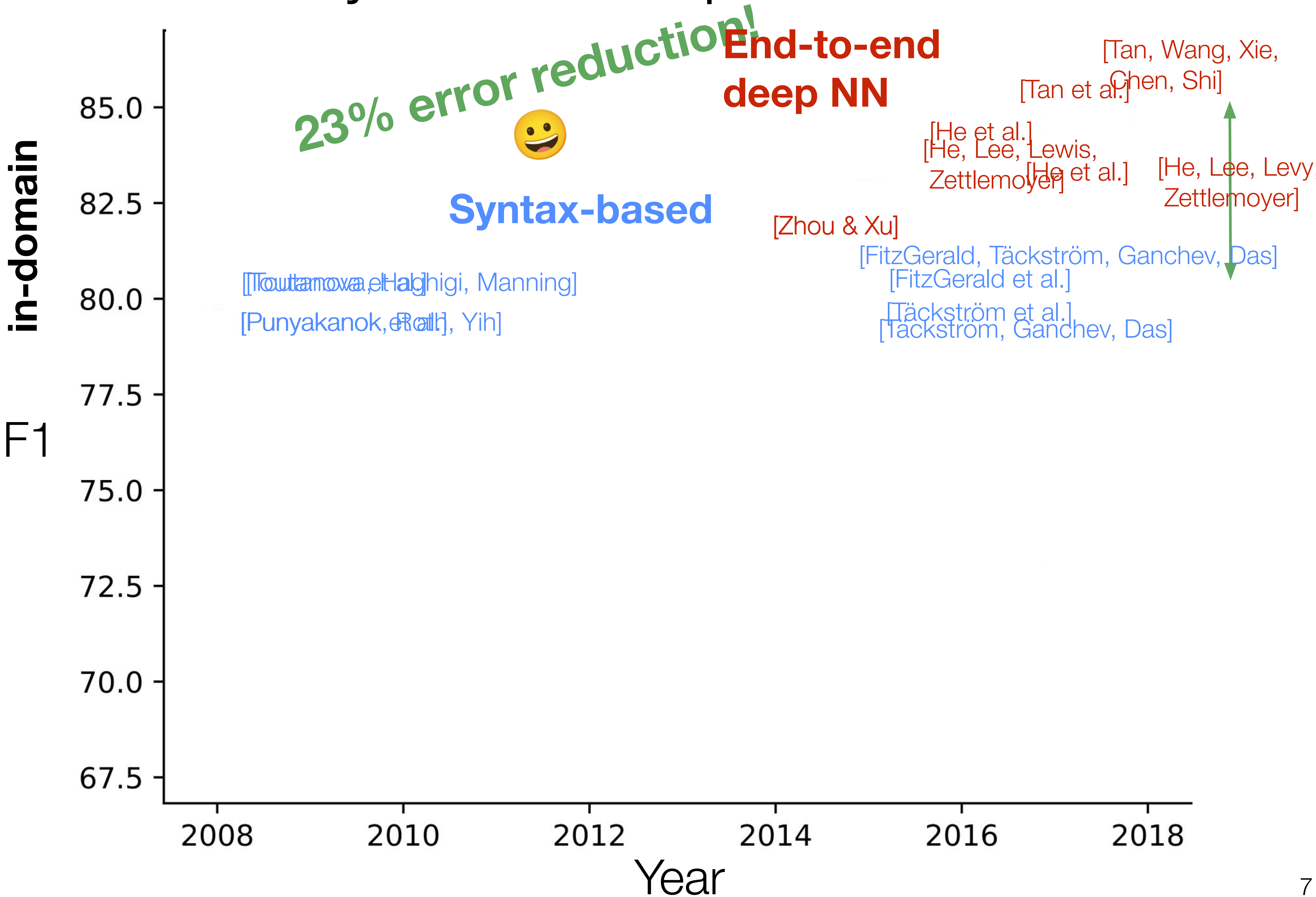
Committee awards Nobel to Strickland who advanced optics

Strickland — ARG_0 *agent*; who — $R\text{-}ARG_0$; advanced — V *predicate*; optics — ARG_1 *theme*

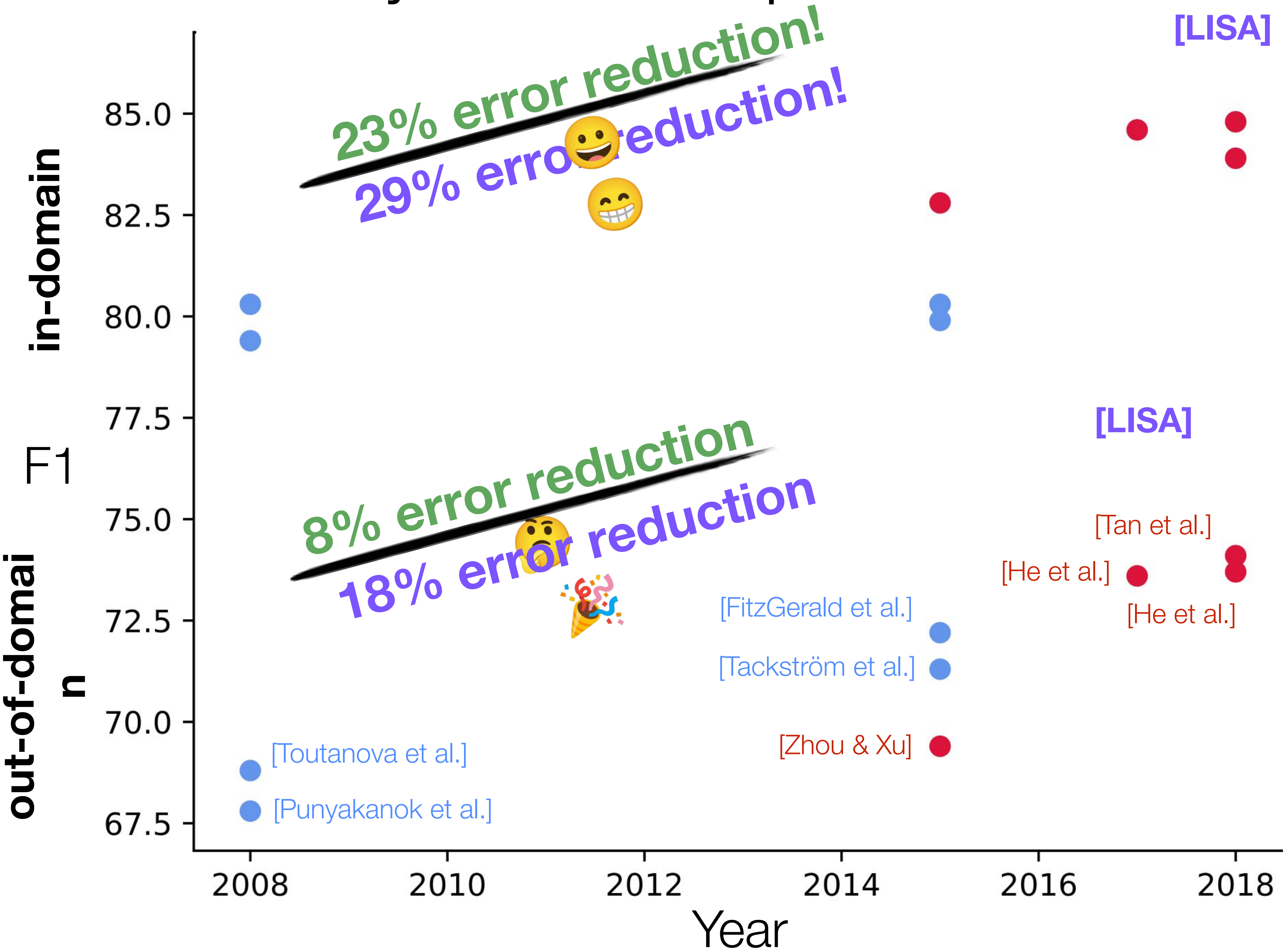
Committee awards Nobel to Strickland who advanced optics

Committee — ARG_0 *agent*; awards — V *predicate*; Nobel — ARG_1 *theme*; Strickland — ARG_2 *beneficiary*

10 years of PropBank SRL



10 years of PropBank SRL



Linguistically-Informed Self-Attention

- **Multi-task learning, single-pass inference**
  - Part-of-speech tag
  - Labeled dependency parsing
  - Predicate detection
  - Semantic role spans & labeling
- **Syntactically-informed self-attention**
  - Multi-head self-attention supervised by syntax

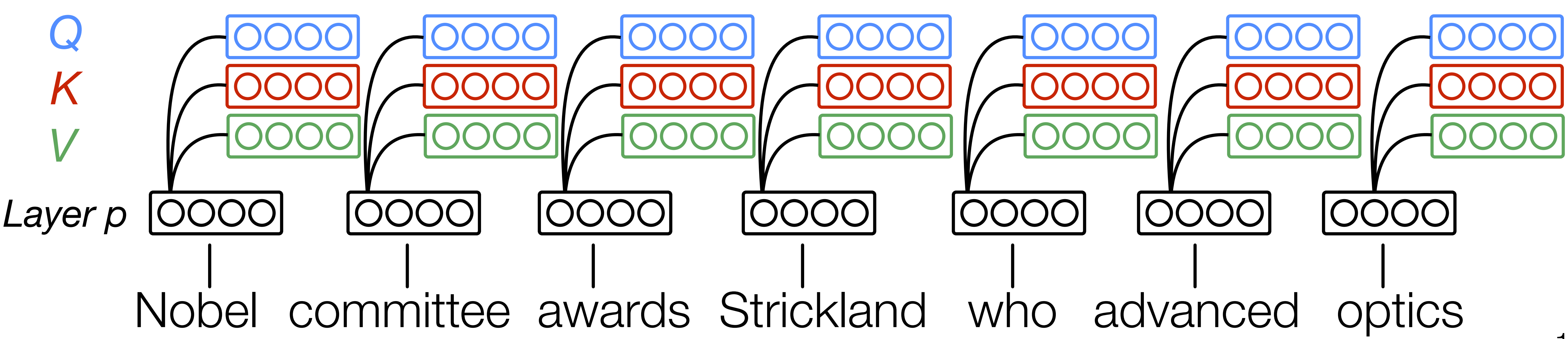
Multi-head self-attention supervised by syntax

Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
  - Multi-head self-attention [Vaswani et al. 2017]
  - Syntactically-informed self-attention
  - Multi-task learning, single-pass inference
- Experimental results & error analysis

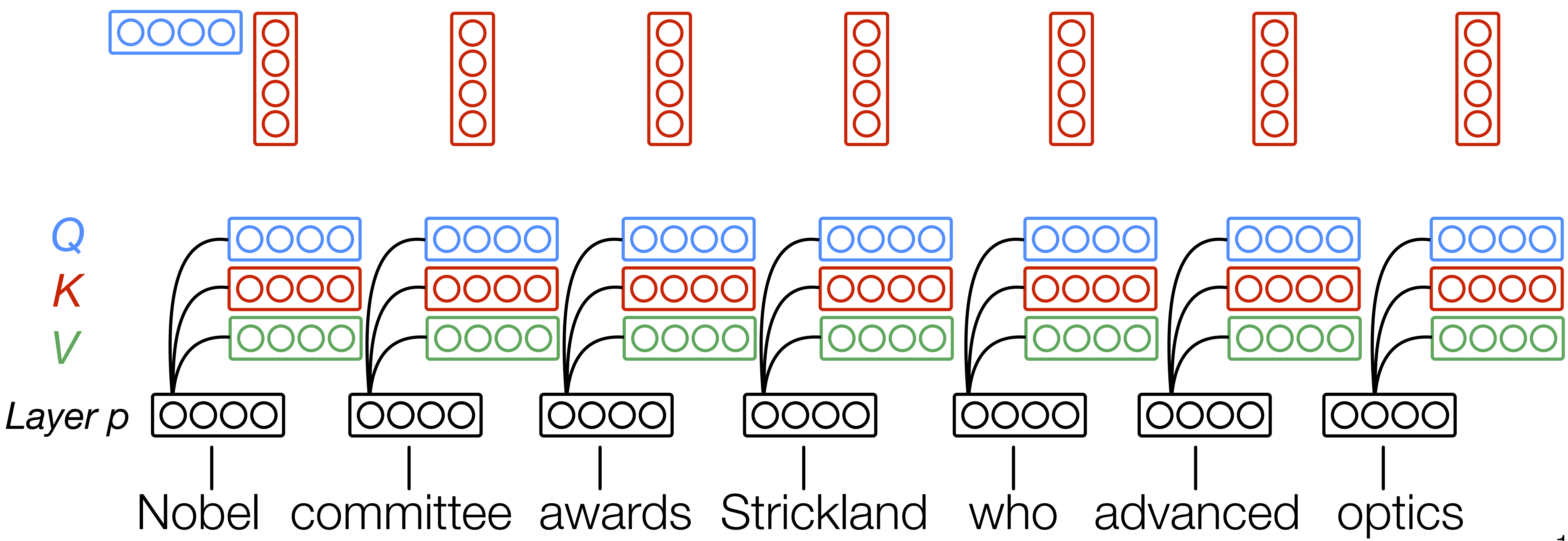
[Vaswani et al. 2017]

Self-attention

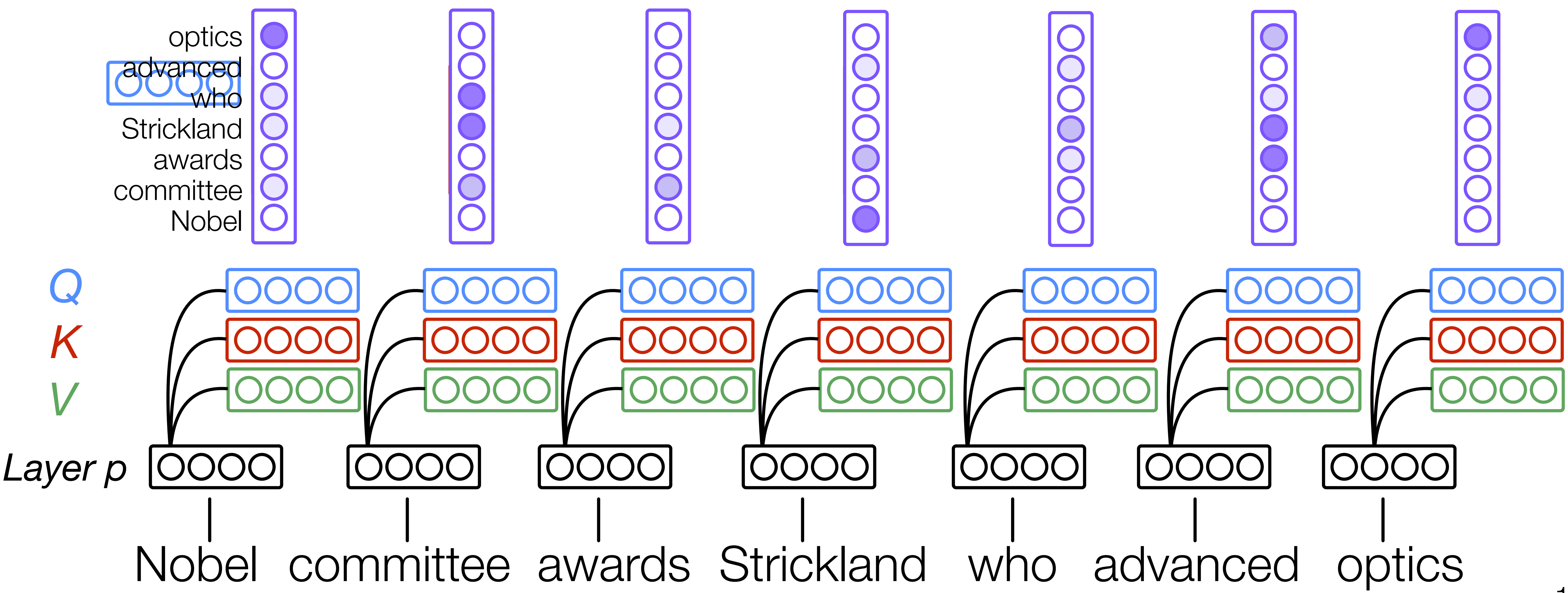


[Vaswani et al. 2017]

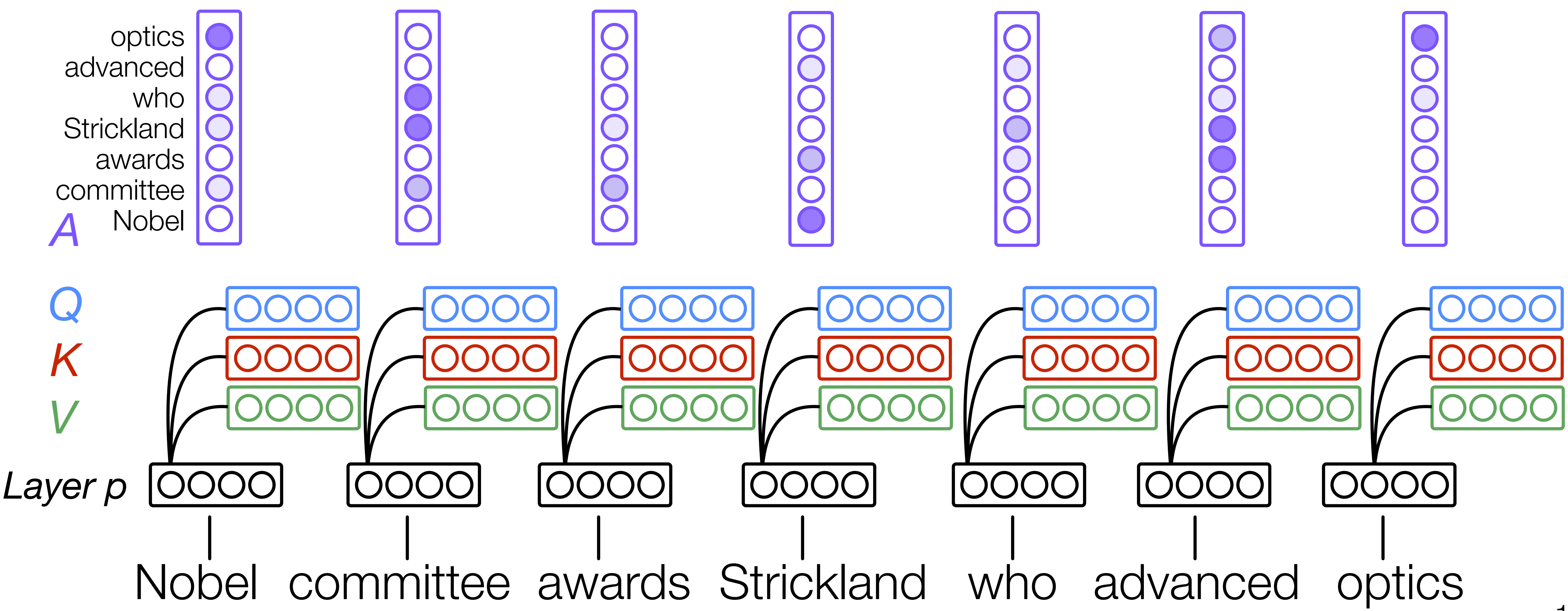
Self-attention



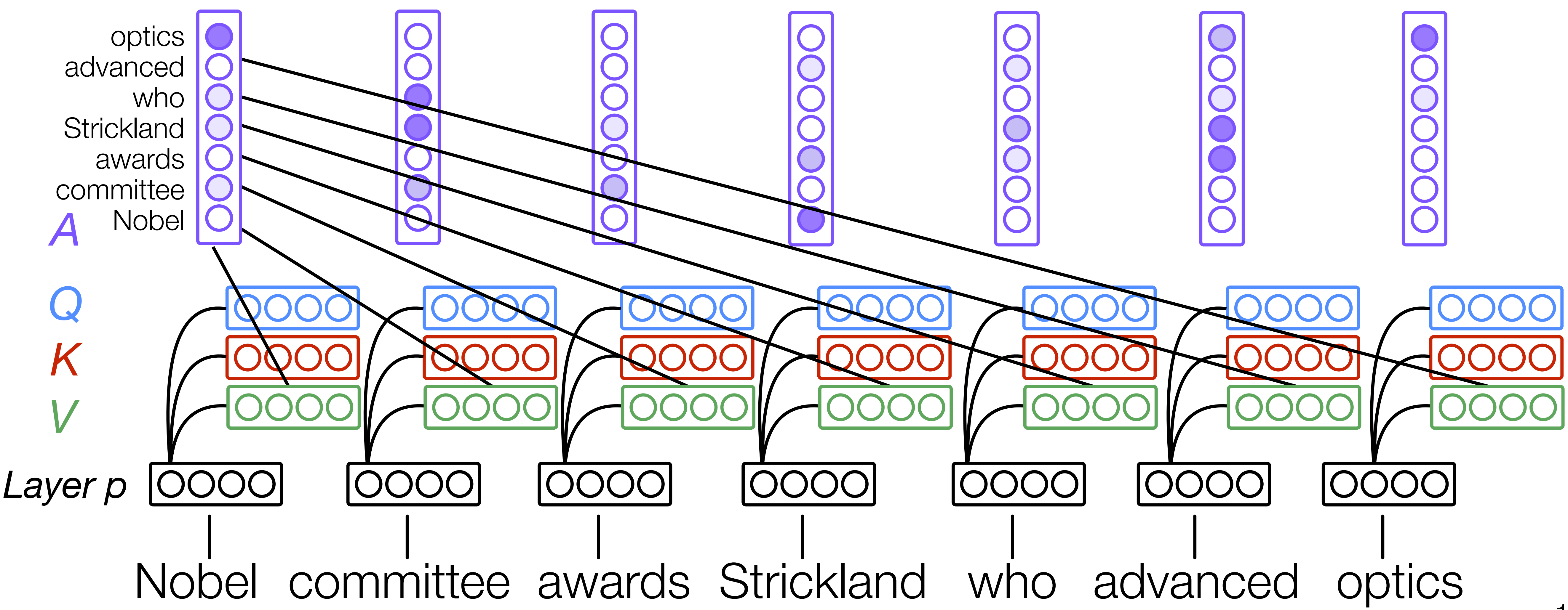
Self-attention



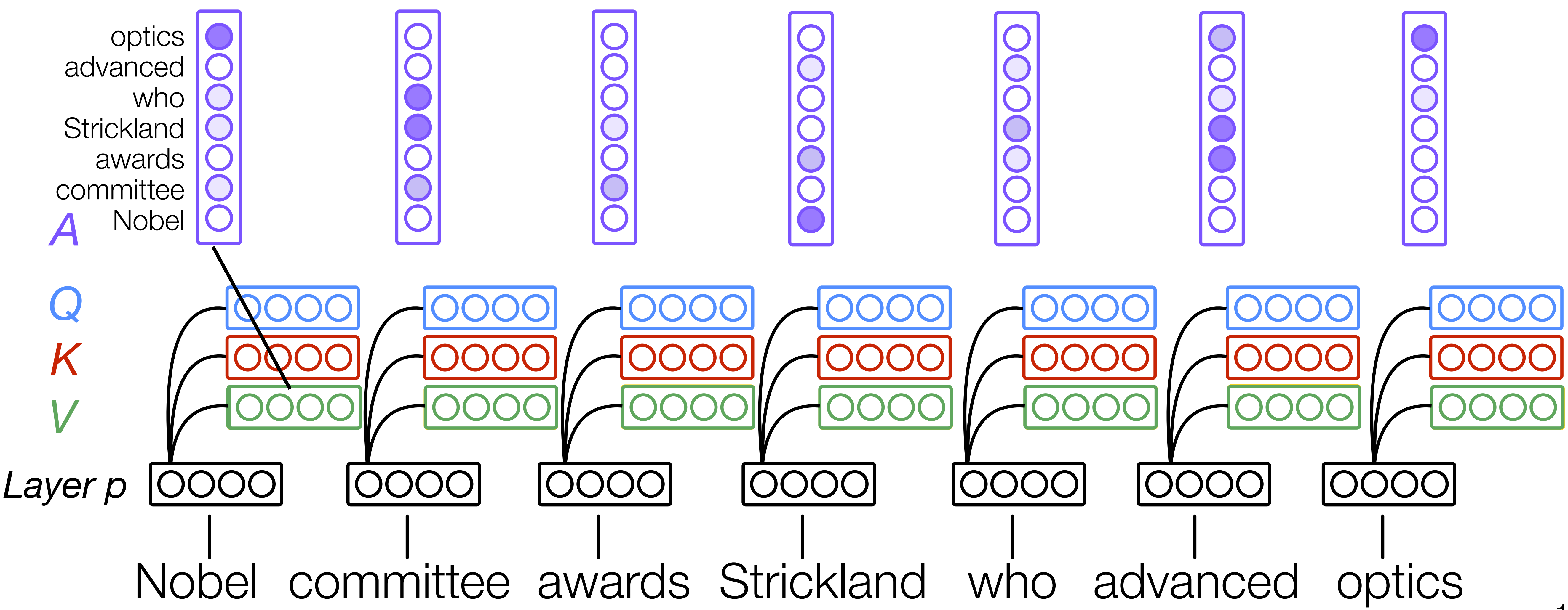
Self-attention



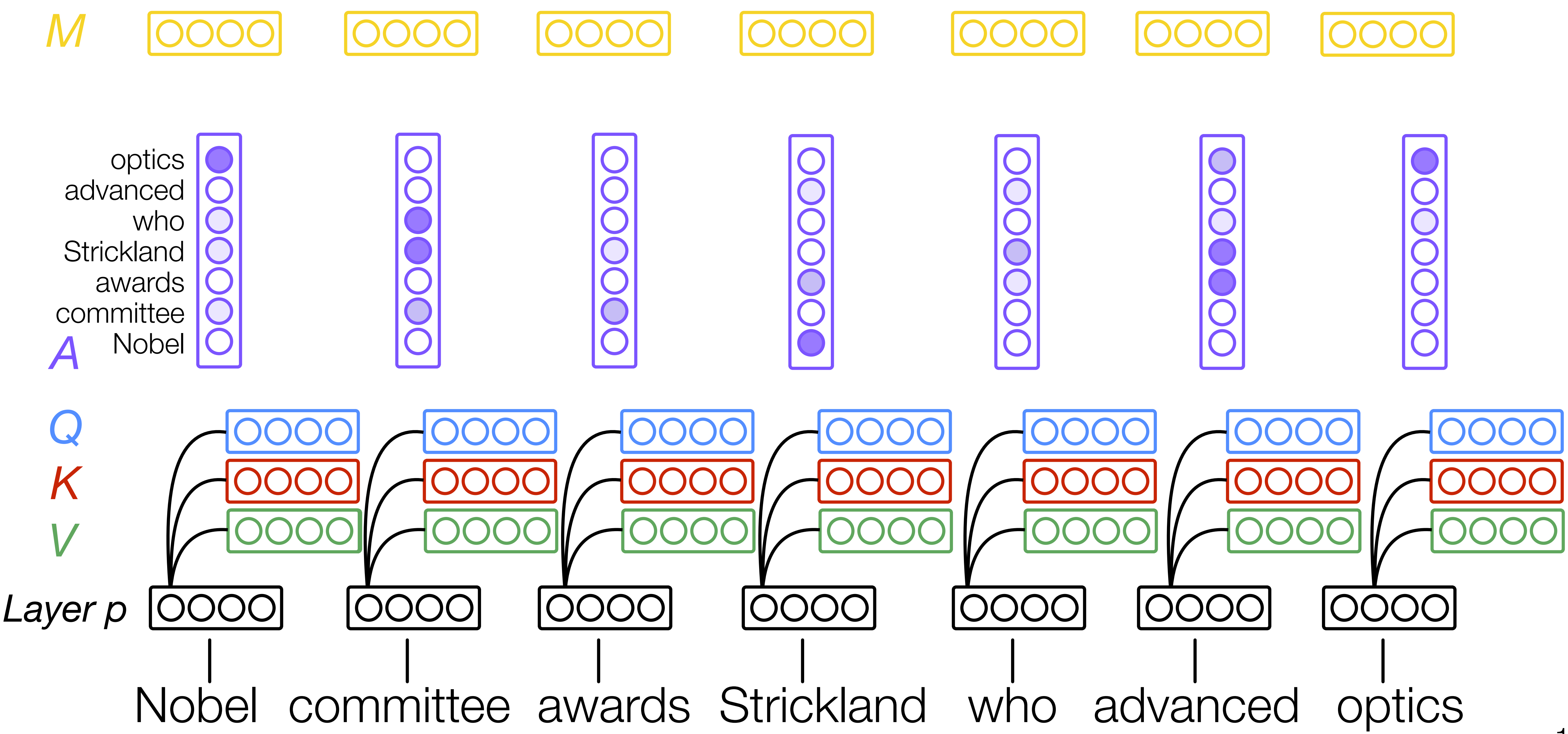
Self-attention



Self-attention

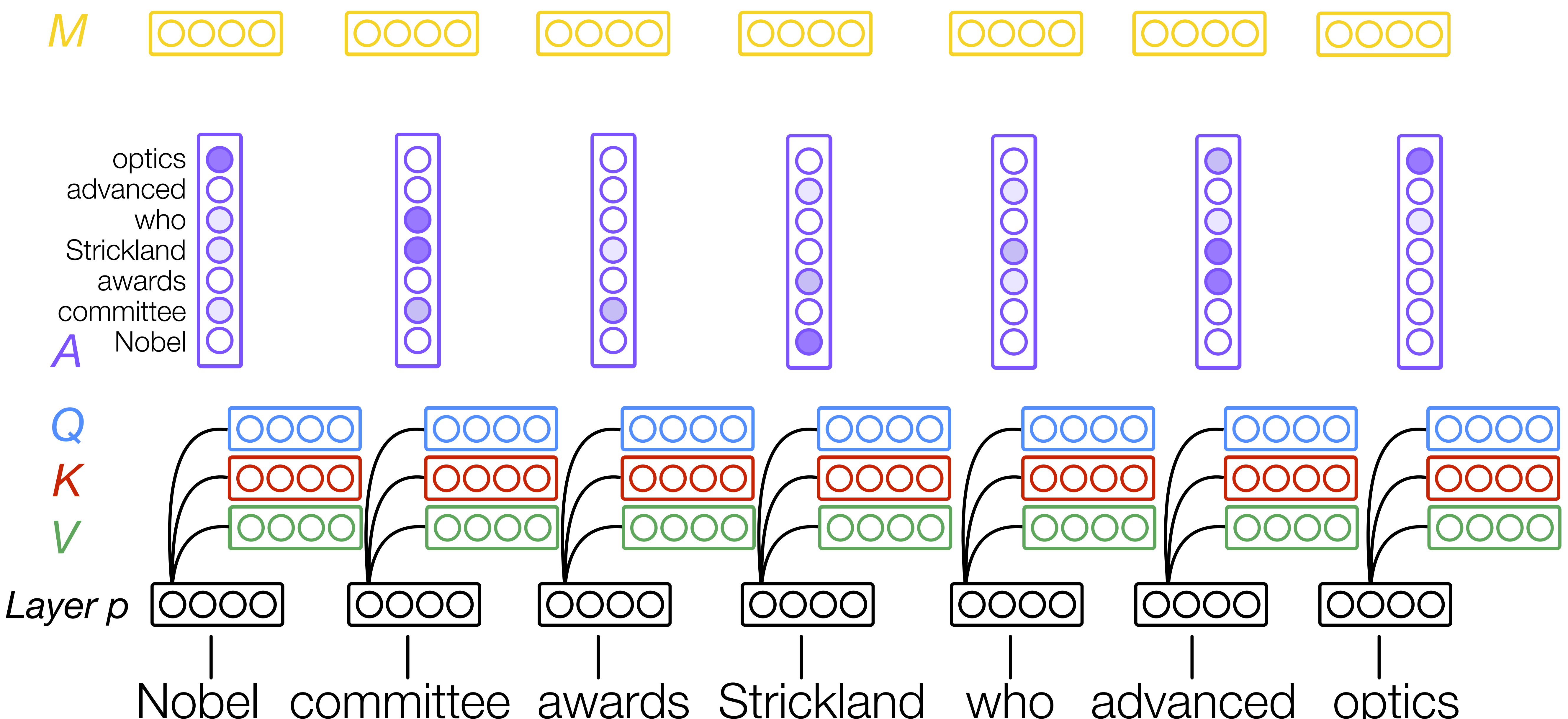


Self-attention

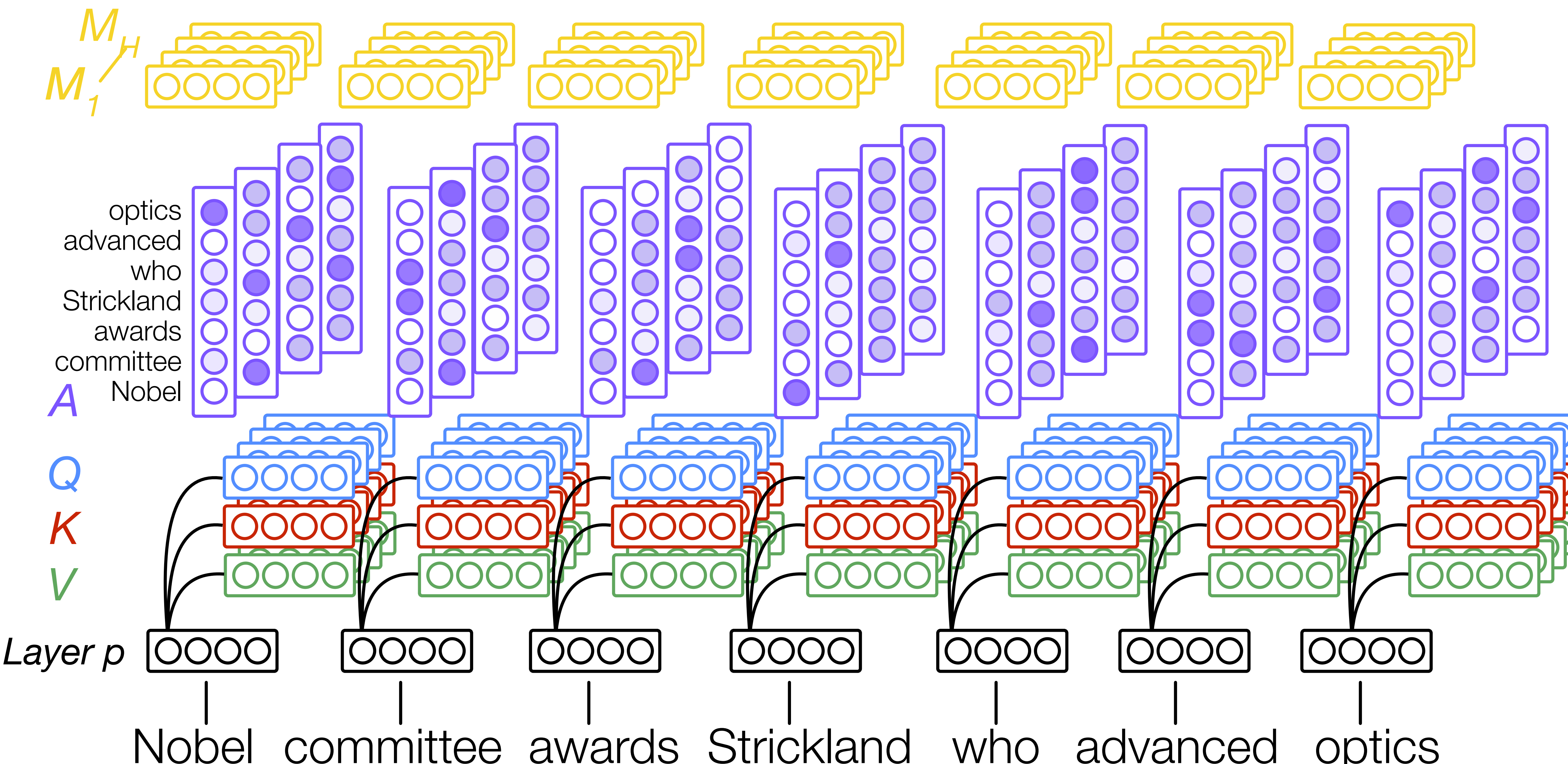


[Vaswani et al. 2017]

Self-attention

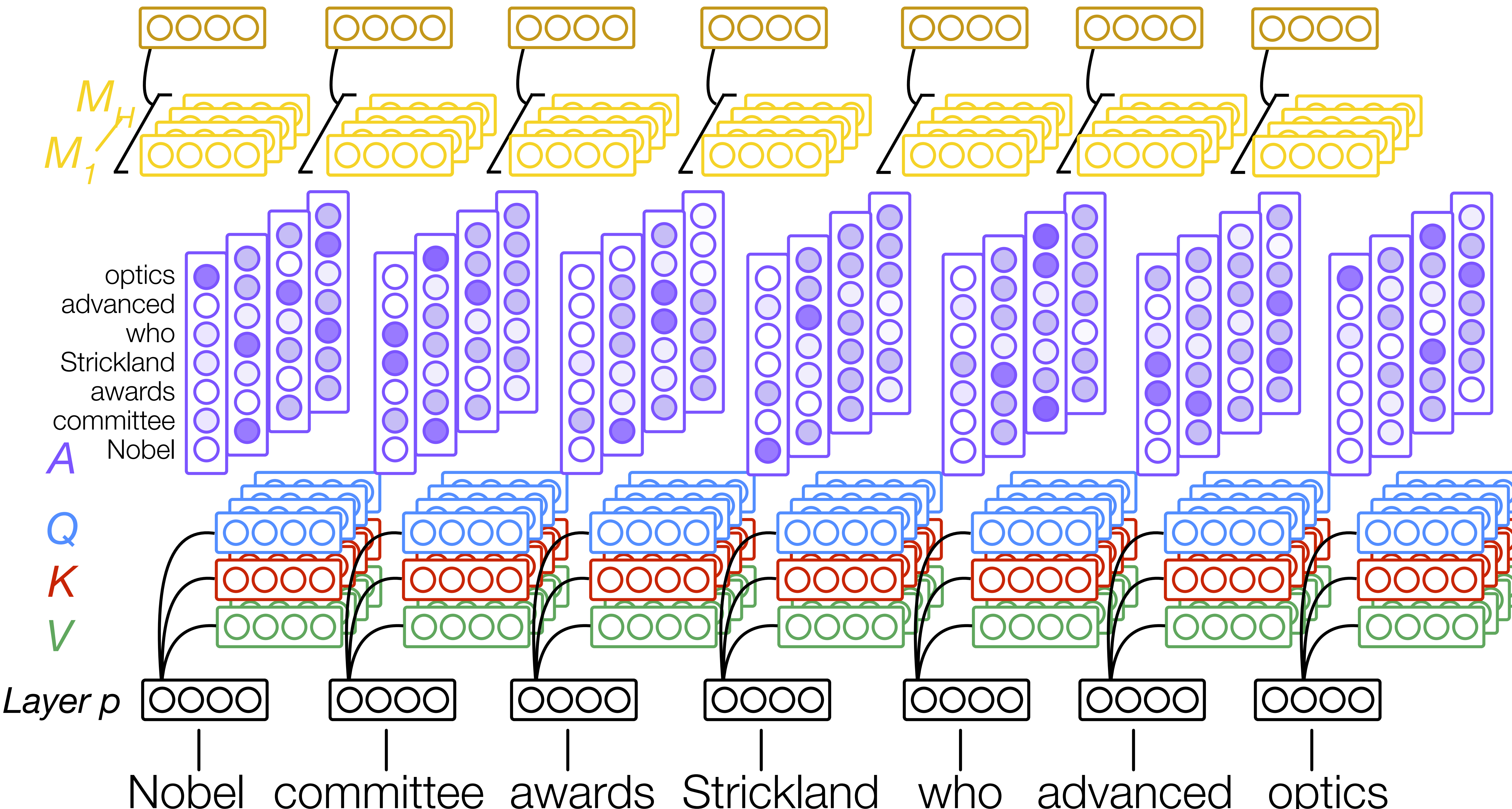


Multi-head self-attention



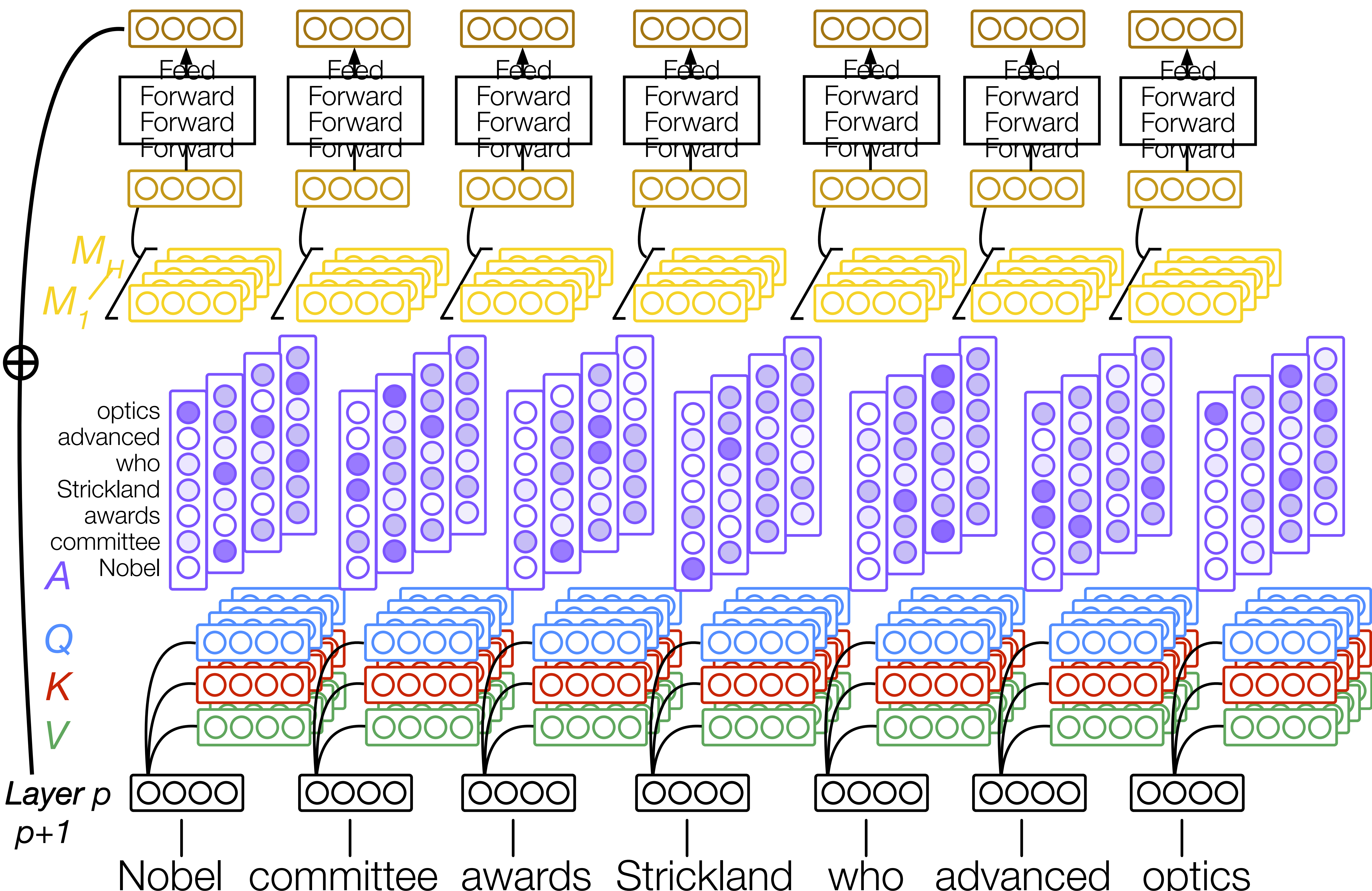
[Vaswani et al. 2017]

Multi-head self-attention



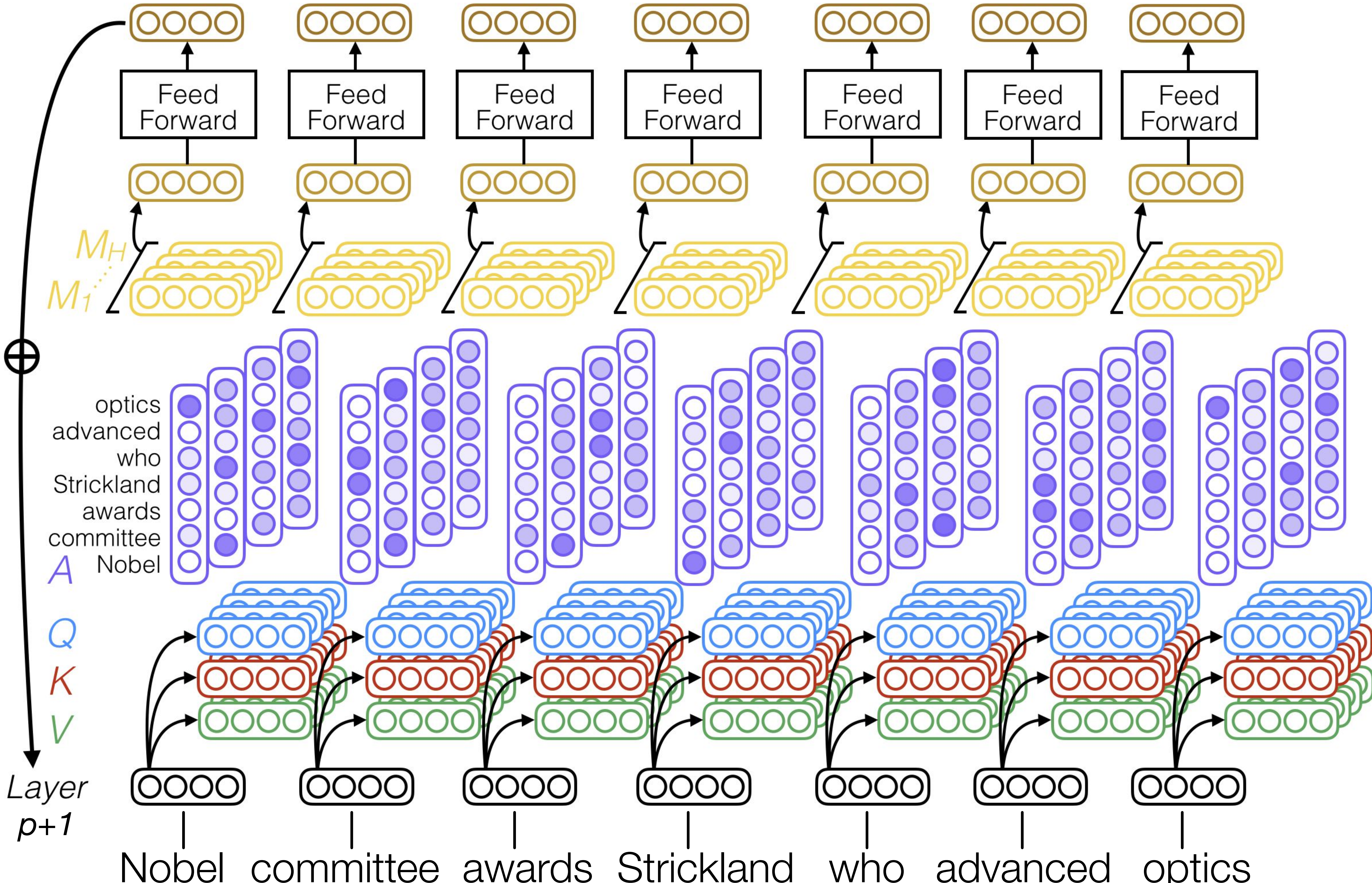
[Vaswani et al. 2017]

Multi-head self-attention



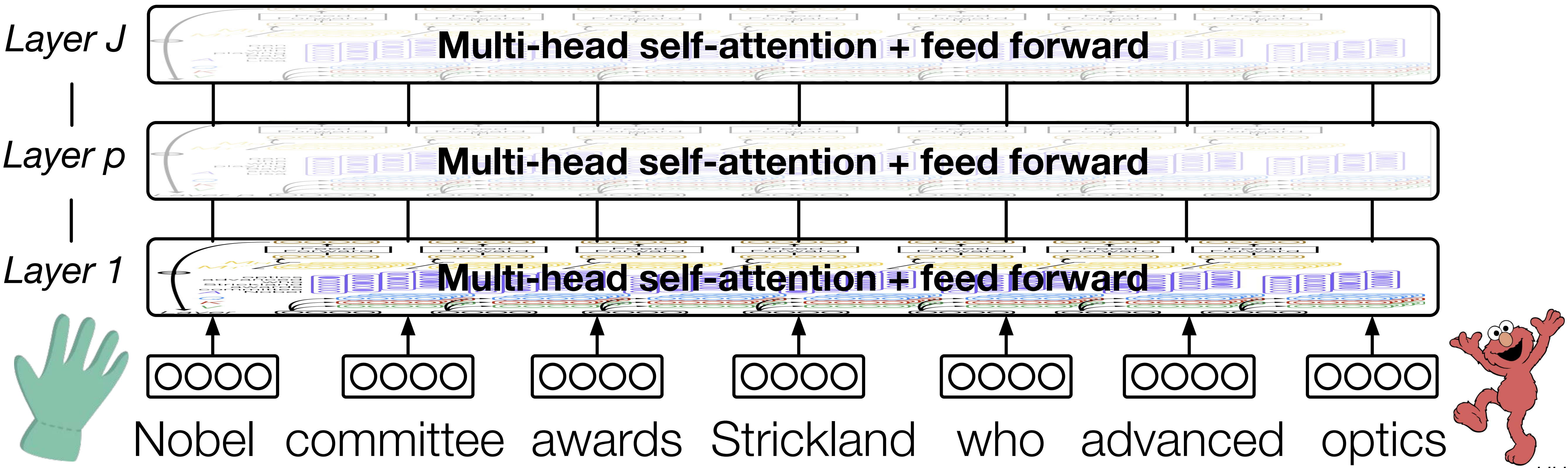
[Vaswani et al. 2017]

Multi-head self-attention



Multi-head self-attention

[Vaswani et al. 2017]



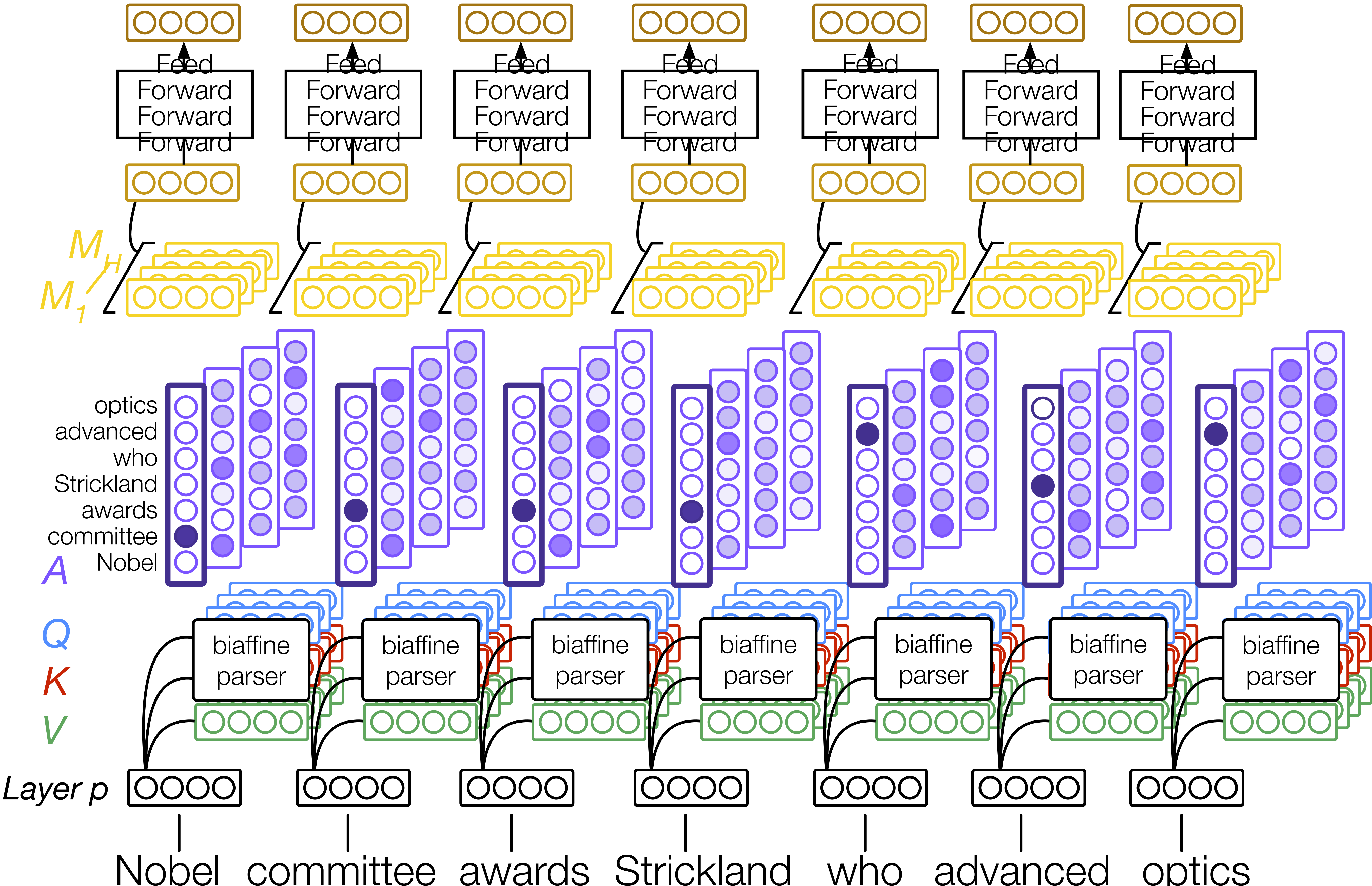
Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
  - Multi-head self-attention [Vaswani et al. 2017]
  - Syntactically-informed self-attention
  - Multi-task learning, single-pass inference
- Experimental results & error analysis

How to incorporate syntax?

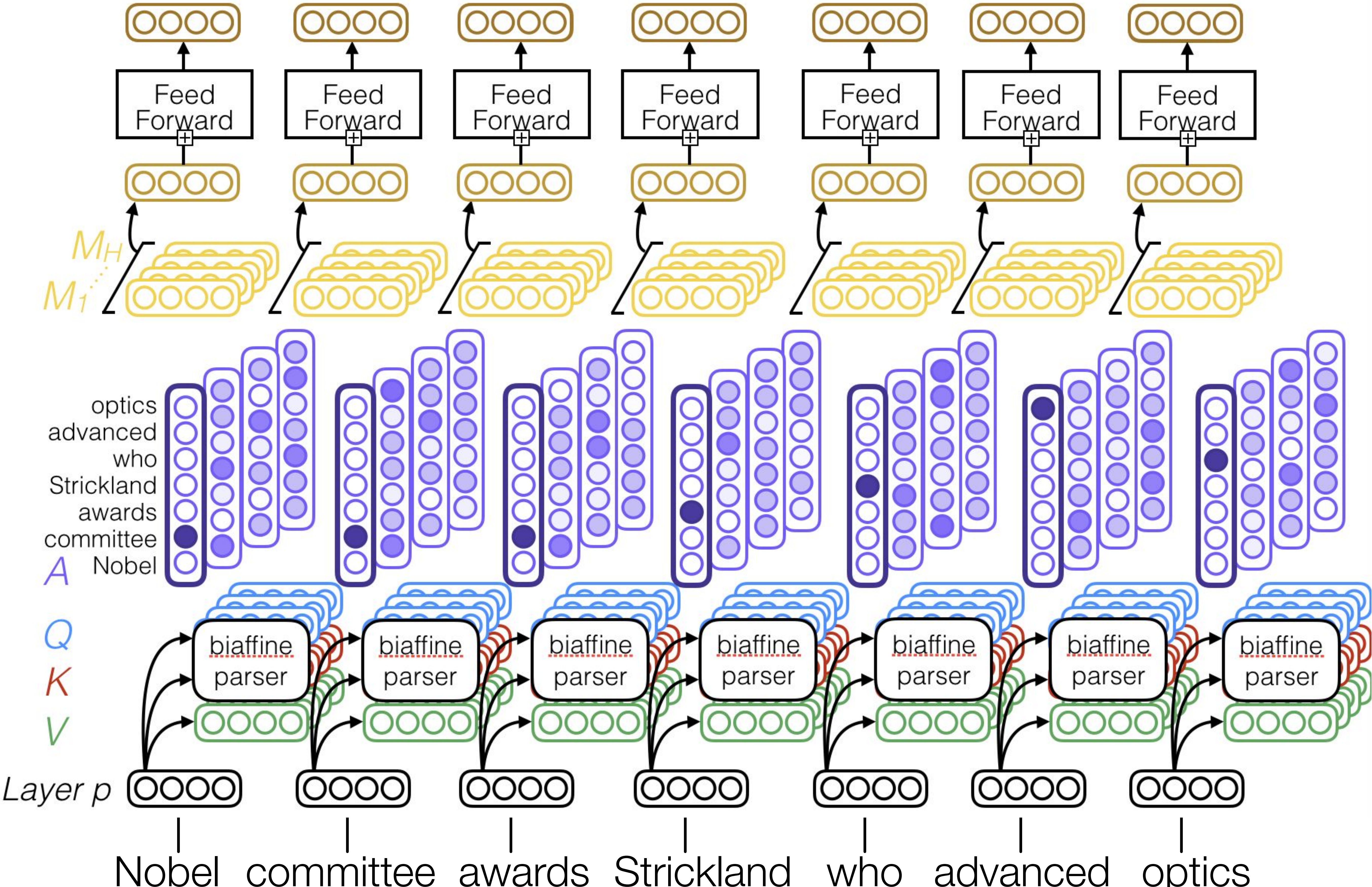
- Multi-task learning [Caruana 1993; Collobert et al. 2011]:
  - Overfits to training domain like single-task end-to-end NN.
  - Must re-train SRL model to leverage new (improved) syntax.
- Dependency path embeddings [Roth & Lapata 2016]; Graph CNN over parse [Marcheggiani & Titov 2017]
  - Restricted context: path to predicate or fixed-width window.
- Syntactically-informed self-attention
  - In one head, token attends to its likely syntactic parent(s).
  - Global context: In next layer, tokens observe all other parents.
  - At test time: can use own predicted parse, *OR* supply syntax to improve SRL model without re-training.

Syntactically-informed self-attention

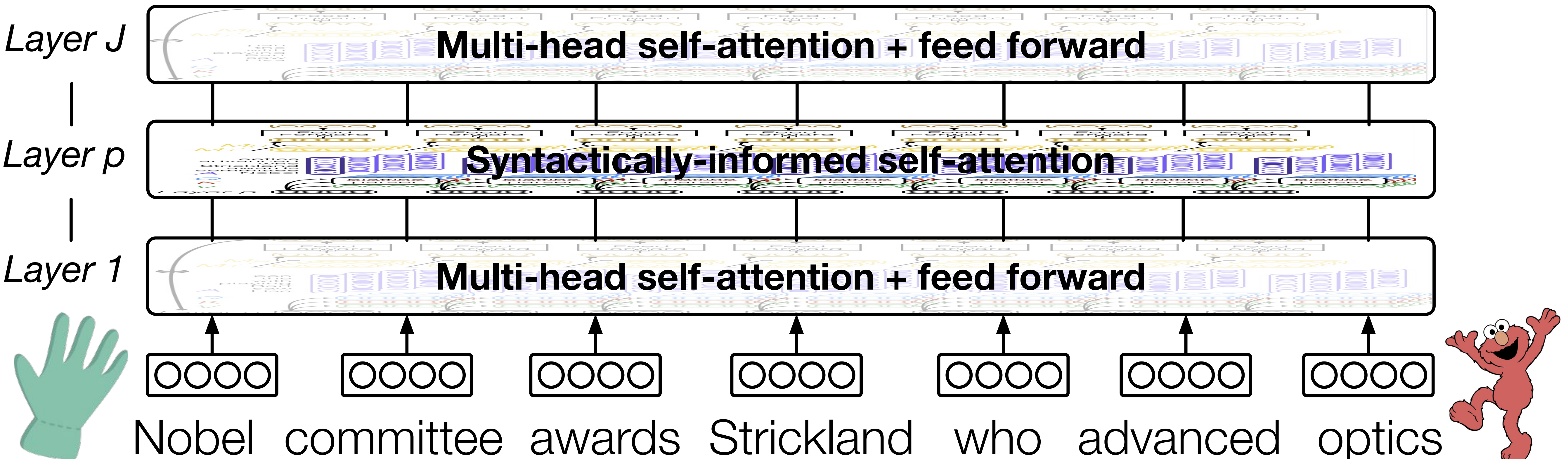


[Dozat and Manning 2017]

Syntactically-informed self-attention



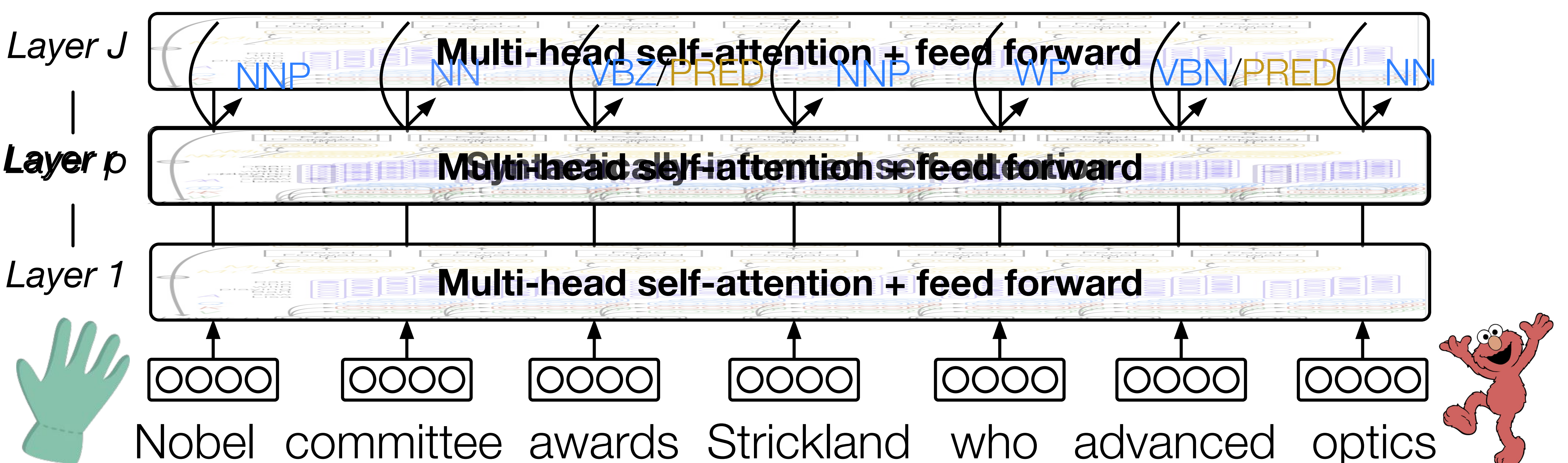
Syntactically-informed self-attention



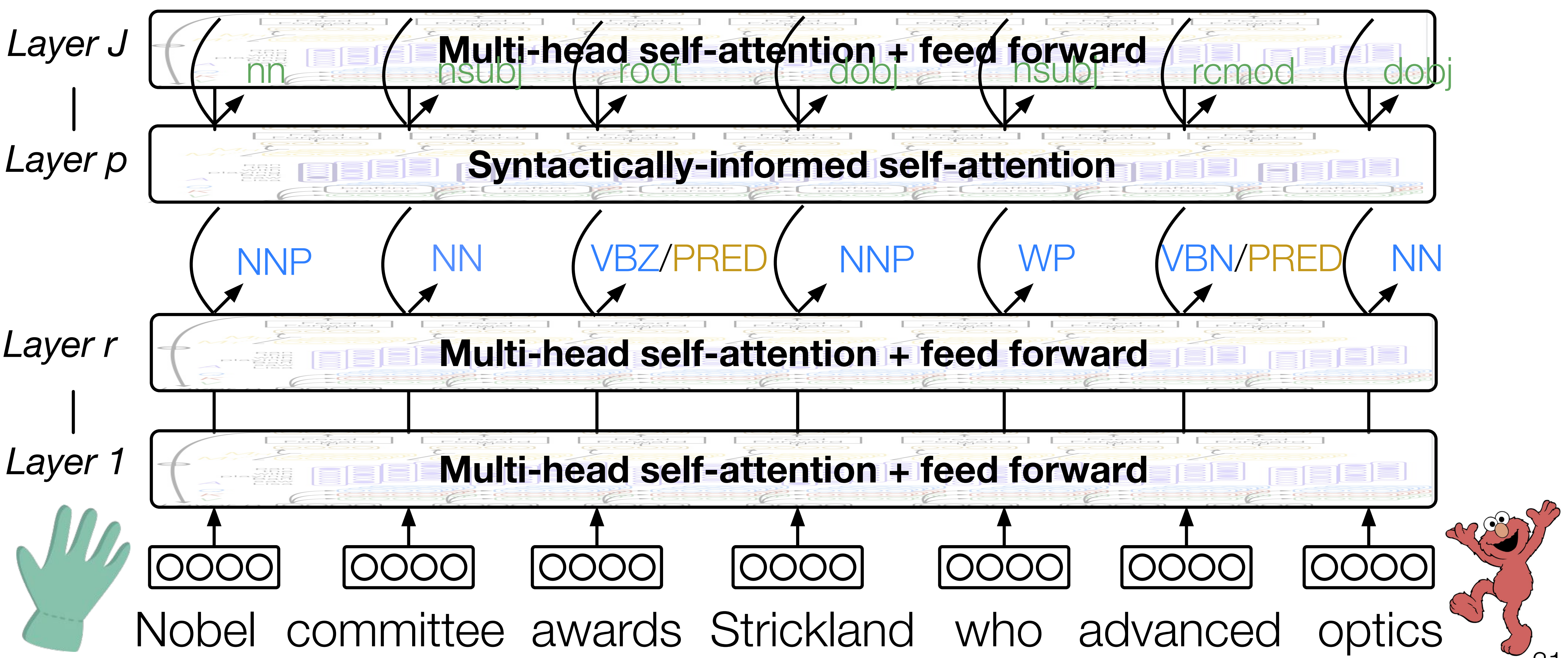
Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
    - Multi-head self-attention
    - Syntactically-informed self-attention
    - Multi-task learning, single-pass inference
- Experimental results & error analysis

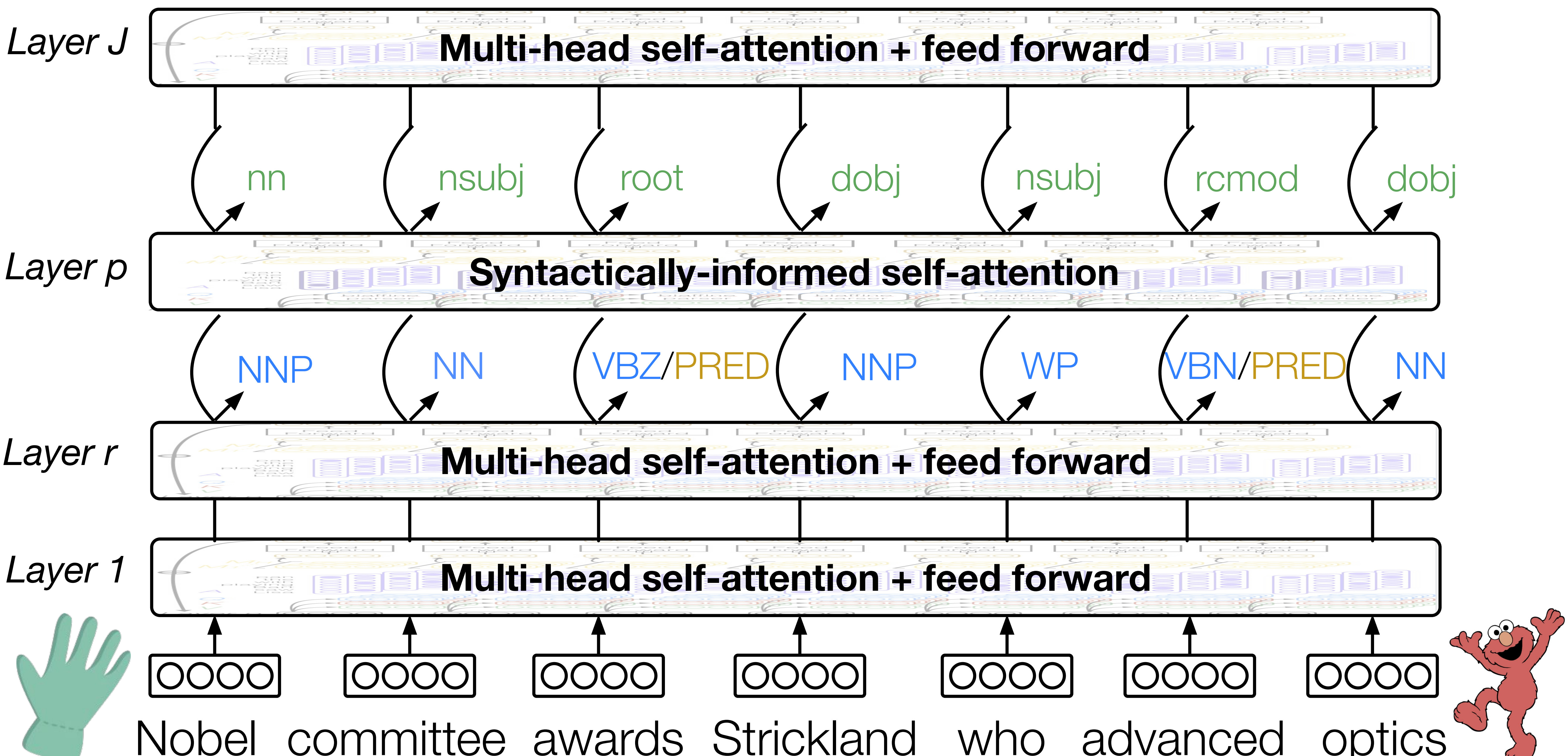
LISA: Linguistically-Informed Self-Attention



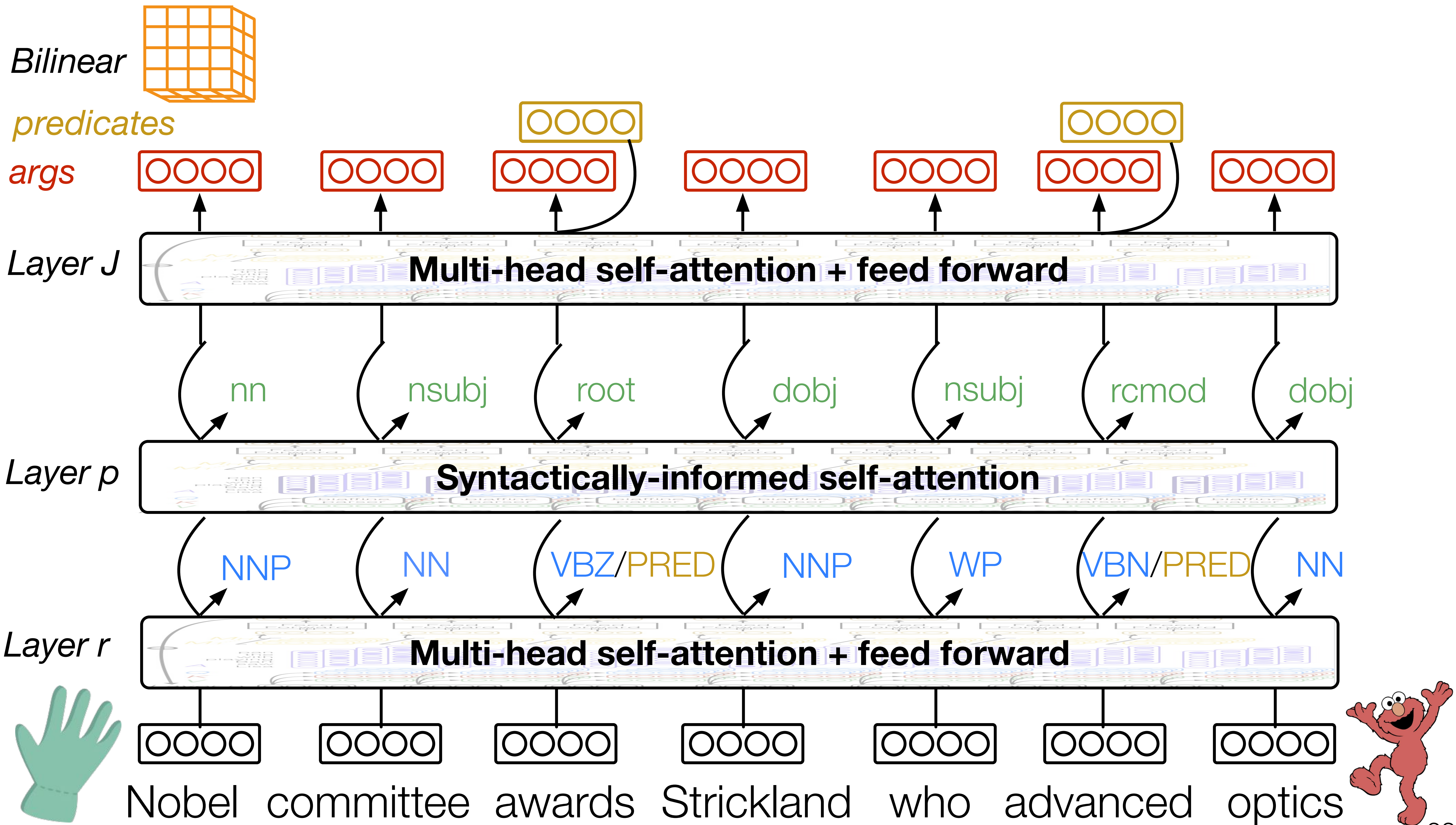
LISA: Linguistically-Informed Self-Attention



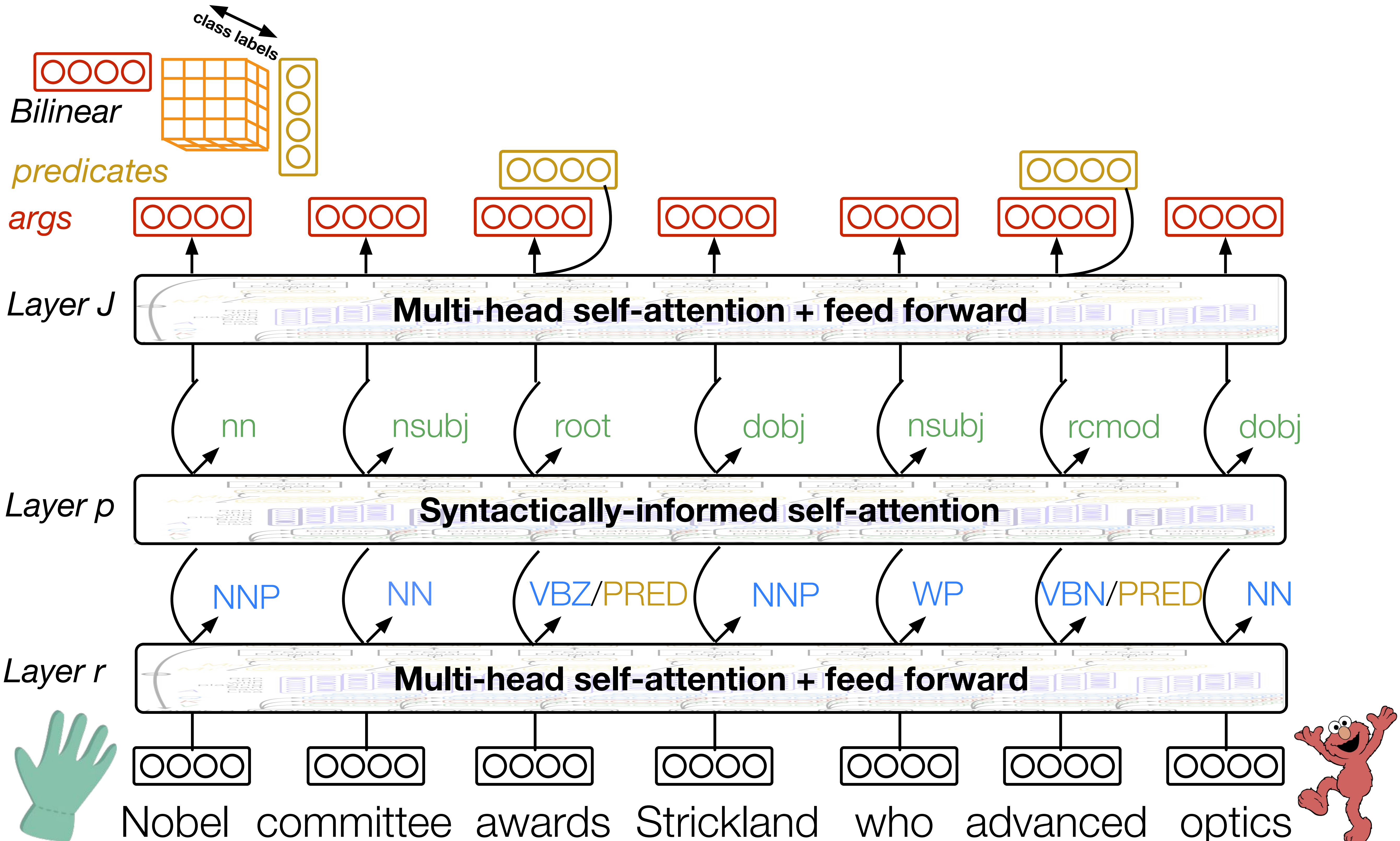
LISA: Linguistically-Informed Self-Attention



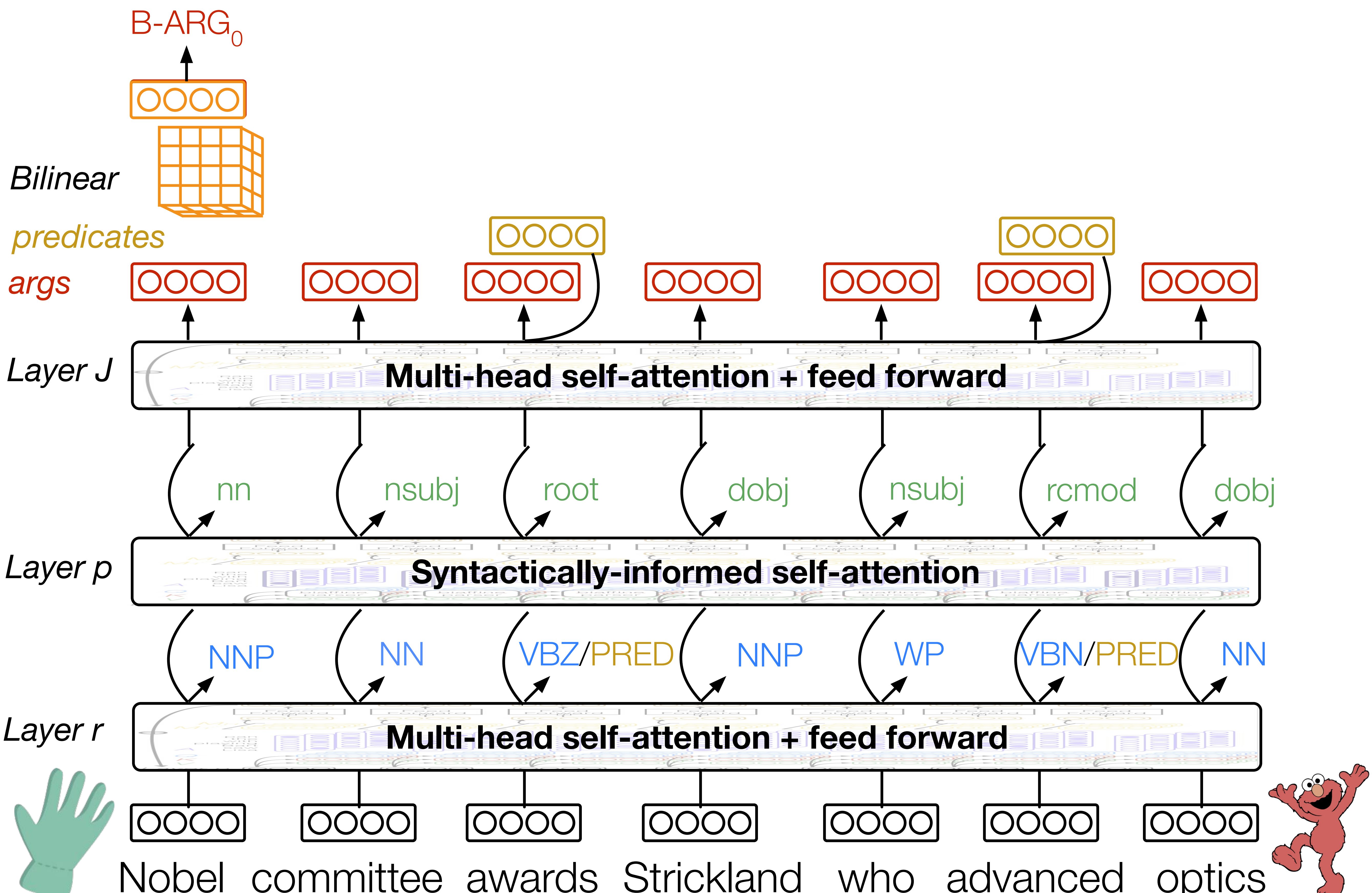
LISA: Linguistically-Informed Self-Attention



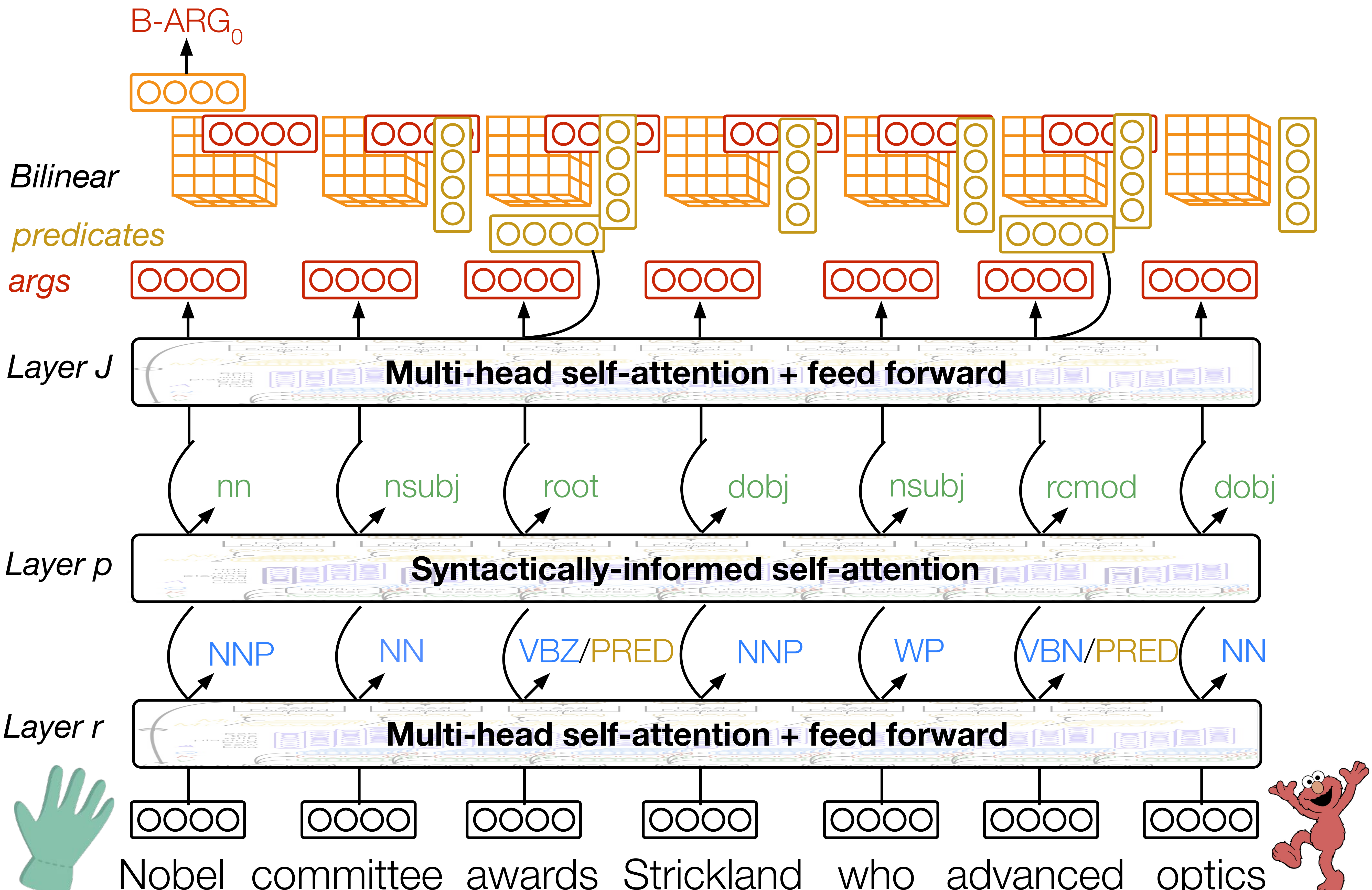
LISA: Linguistically-Informed Self-Attention



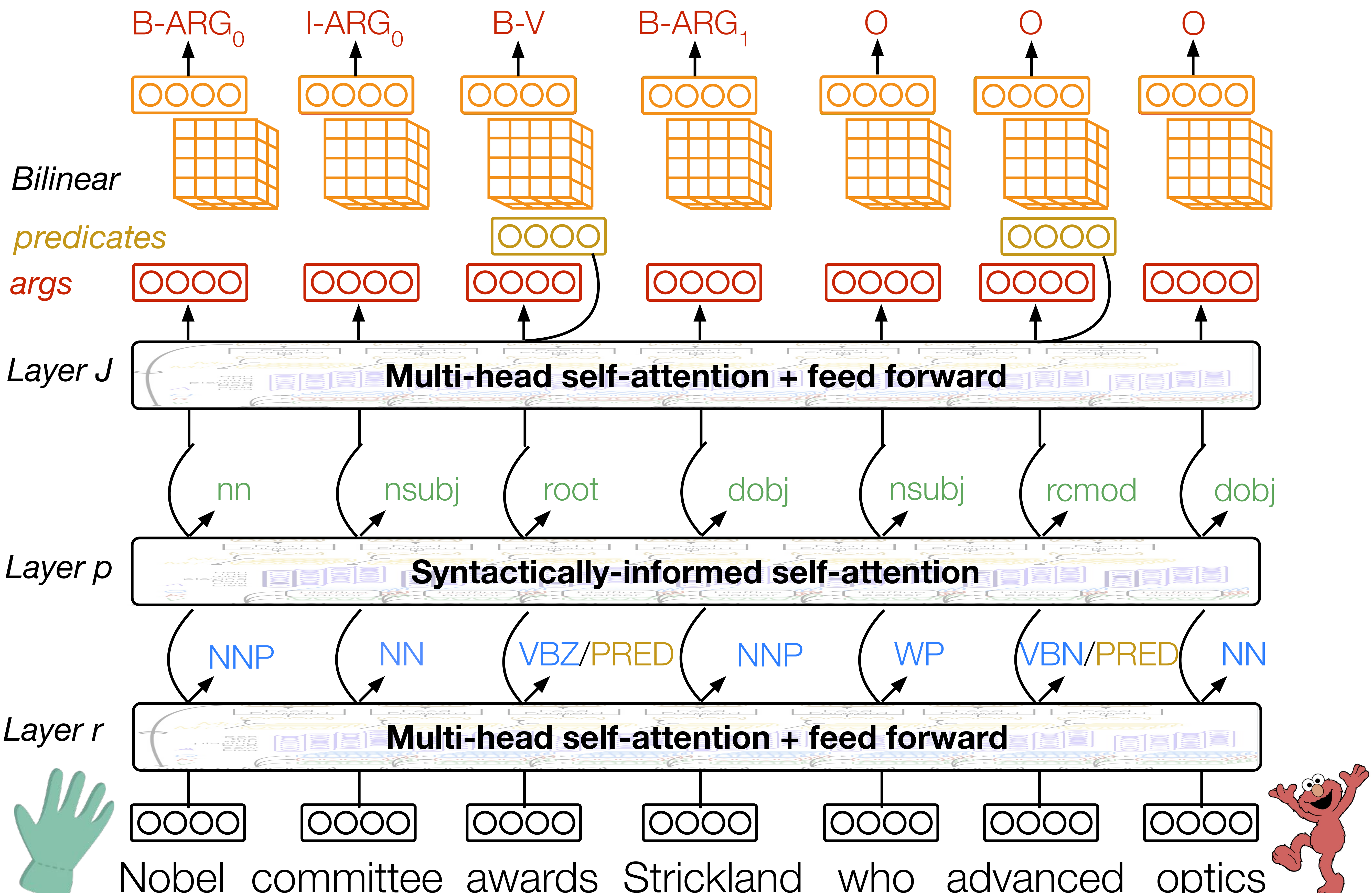
LISA: Linguistically-Informed Self-Attention



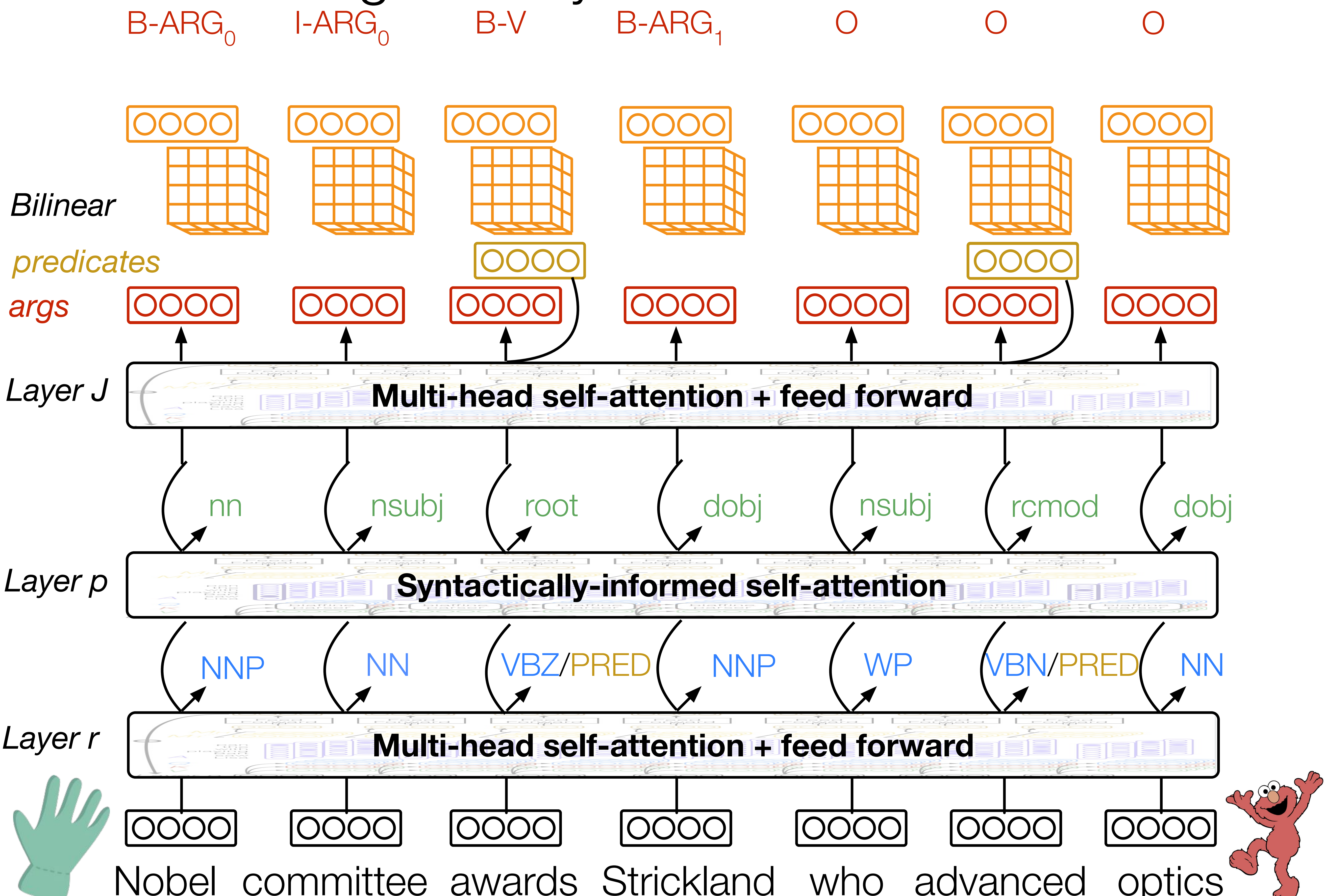
LISA: Linguistically-Informed Self-Attention



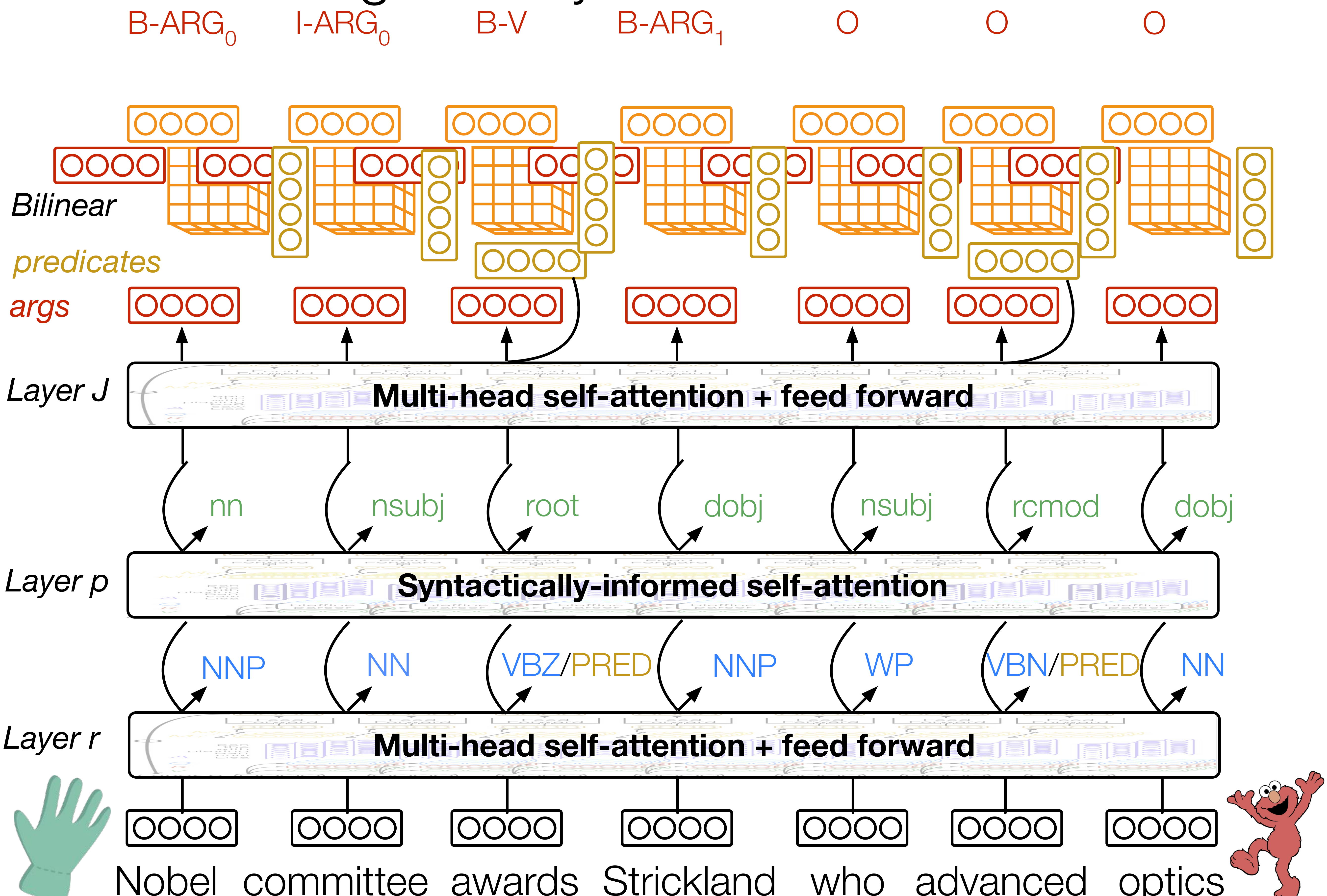
LISA: Linguistically-Informed Self-Attention



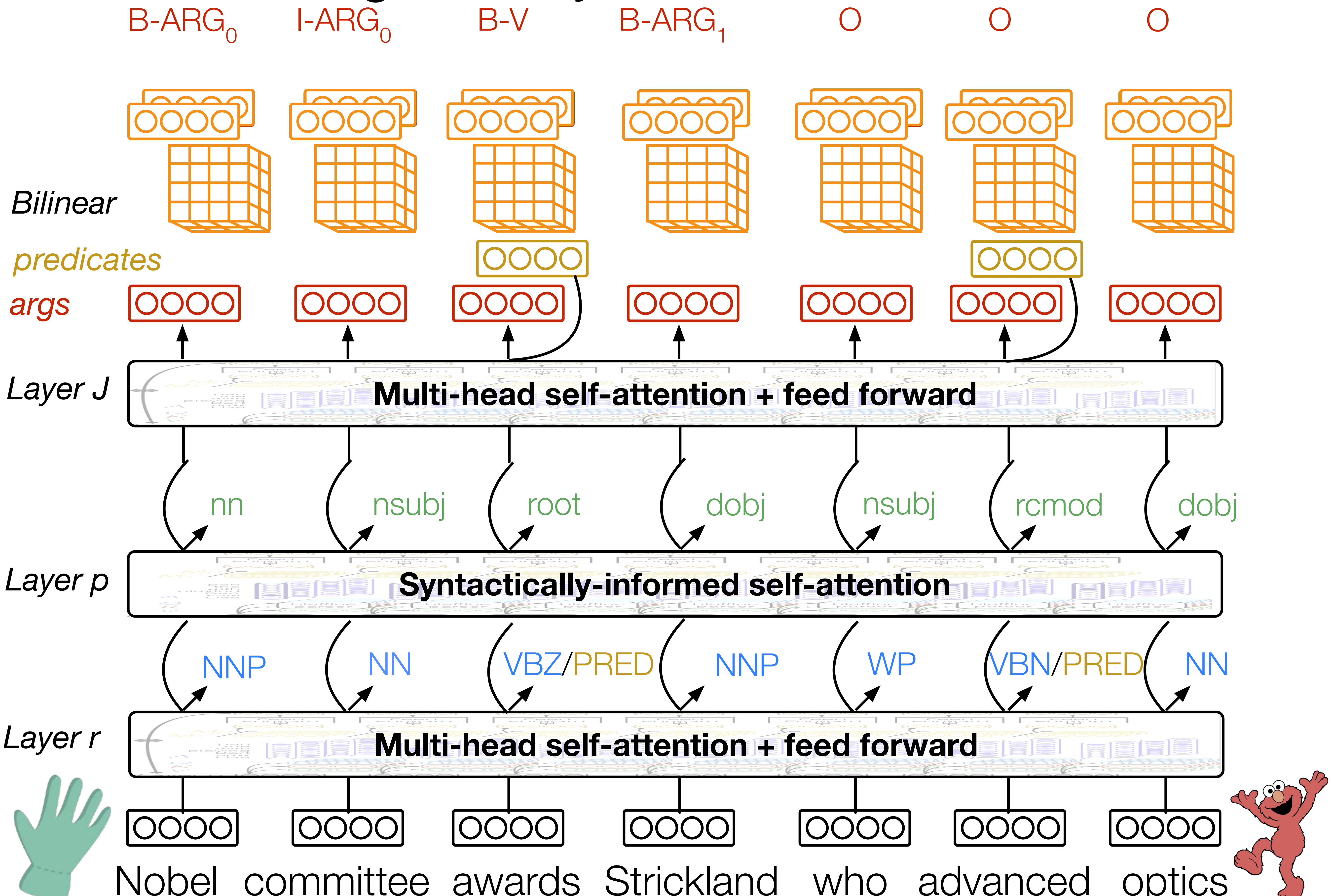
LISA: Linguistically-Informed Self-Attention



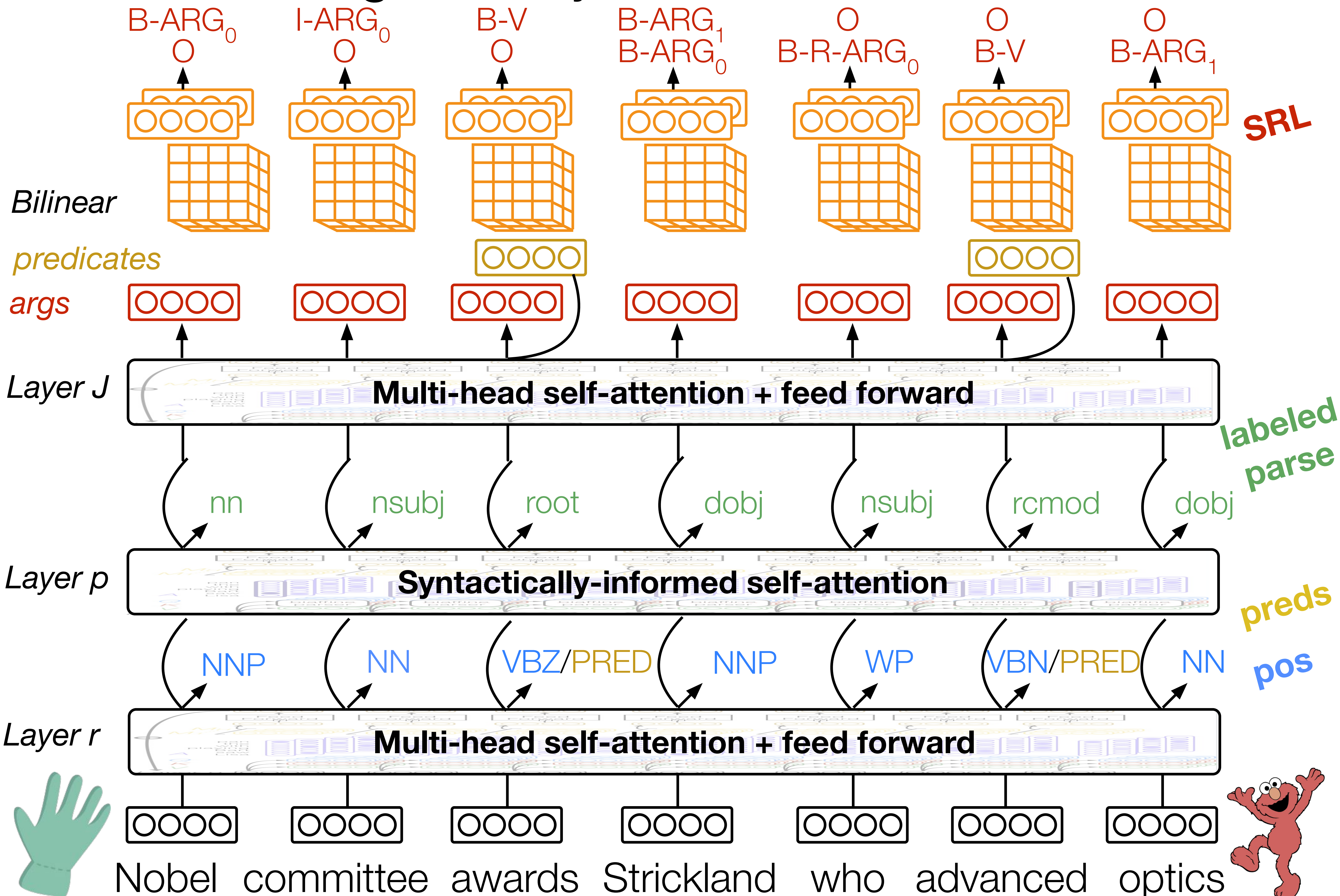
LISA: Linguistically-Informed Self-Attention



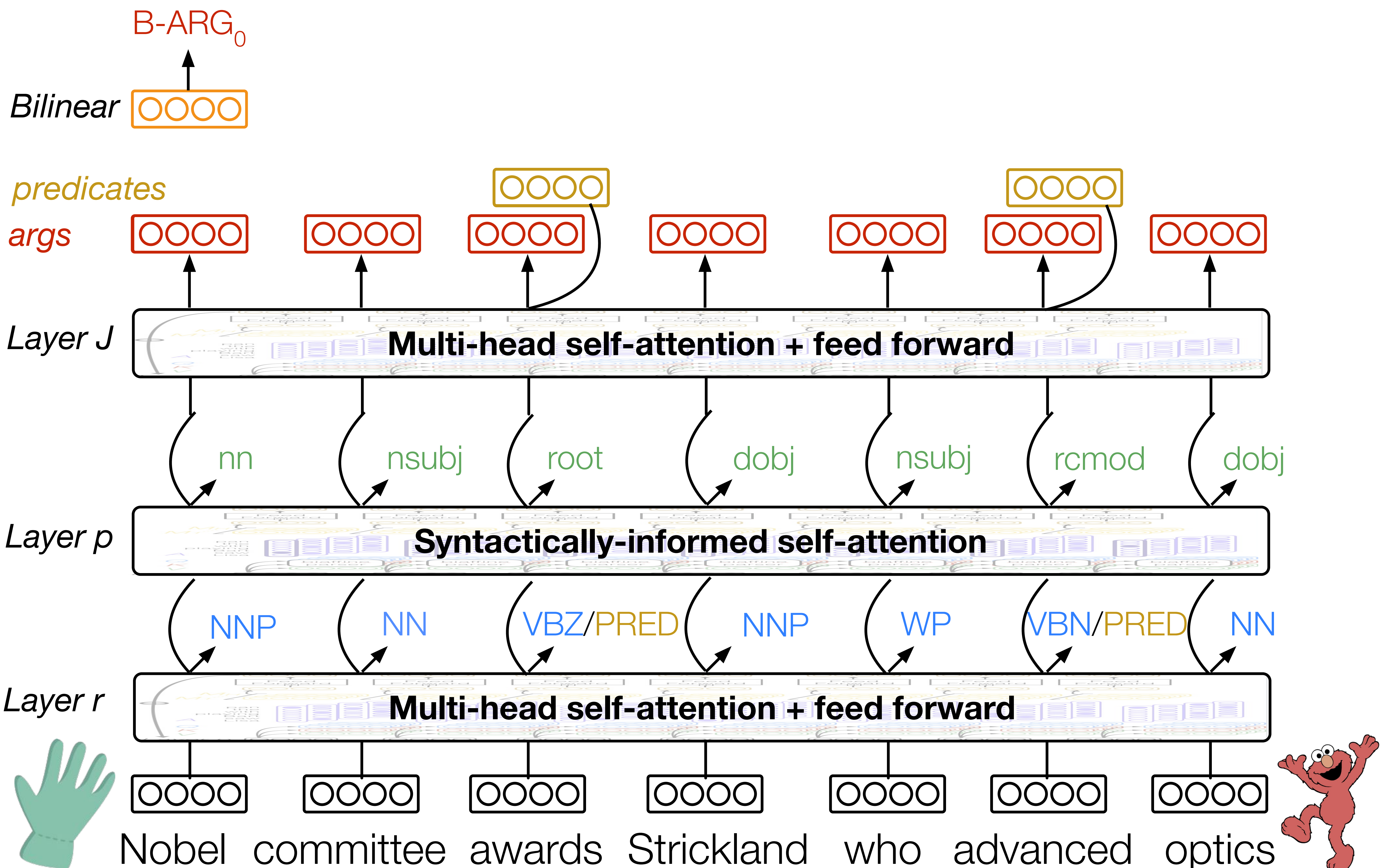
LISA: Linguistically-Informed Self-Attention



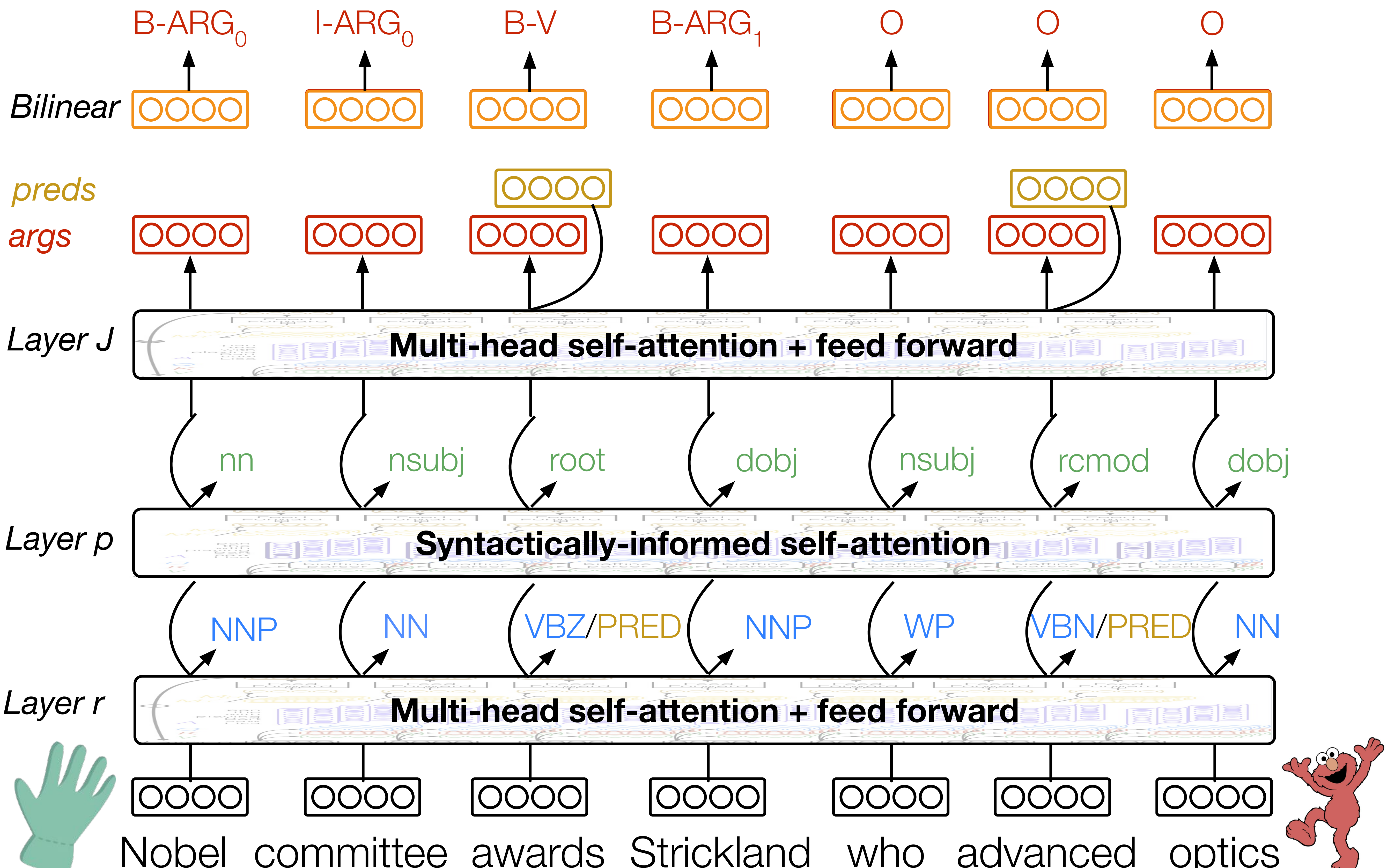
LISA: Linguistically-Informed Self-Attention



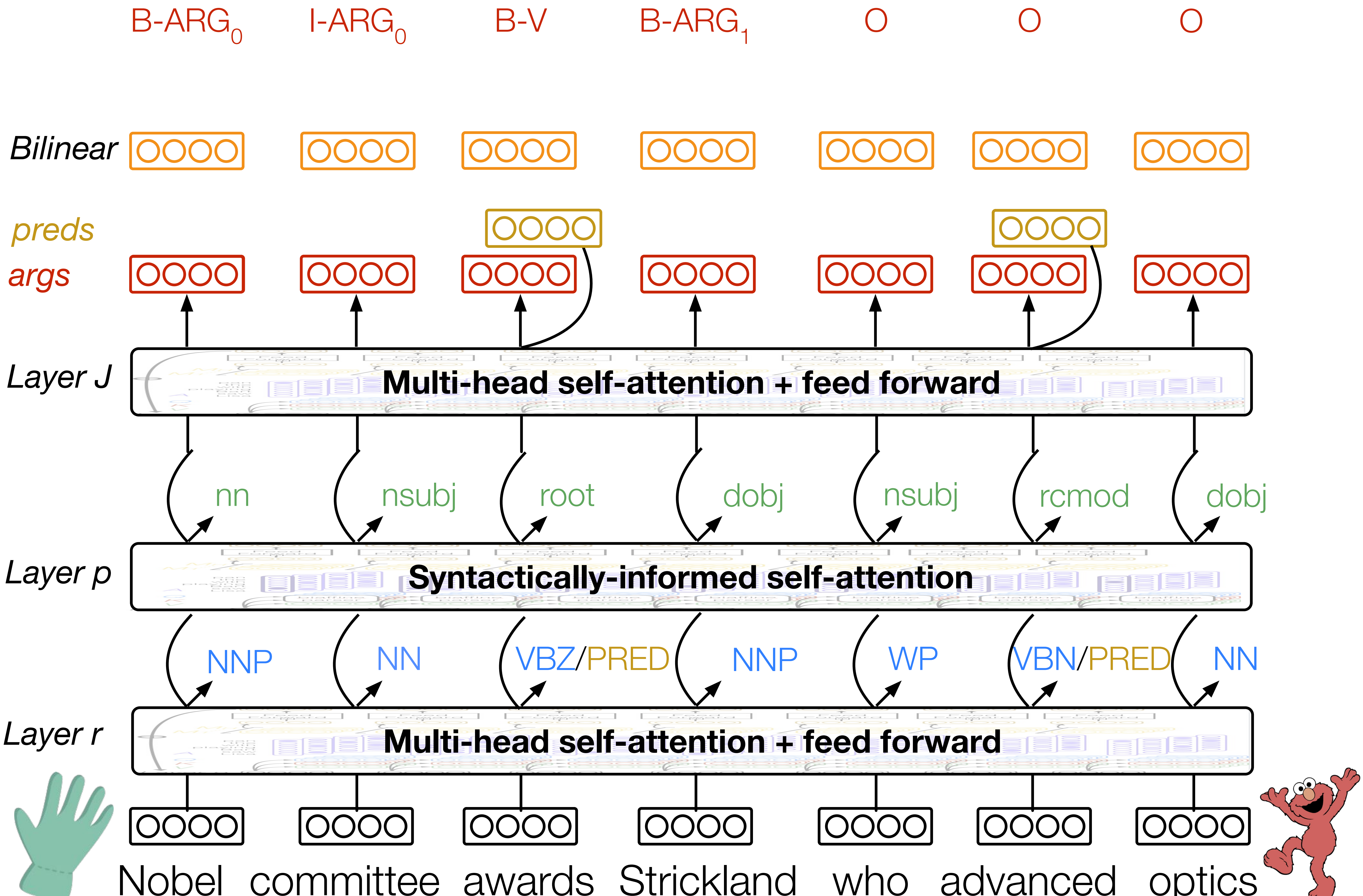
LISA: Linguistically-Informed Self-Attention



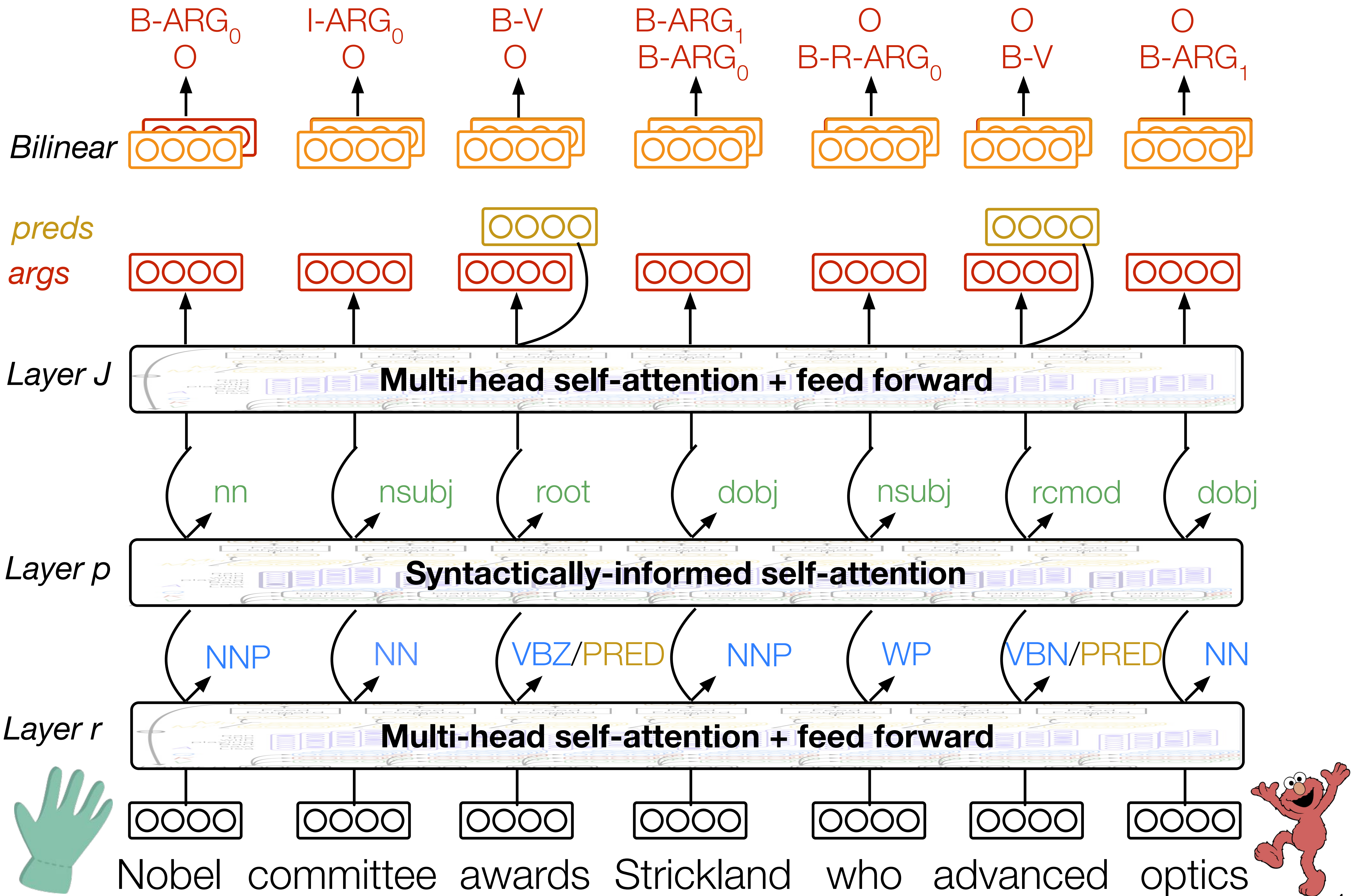
LISA: Linguistically-Informed Self-Attention



LISA: Linguistically-Informed Self-Attention



LISA: Linguistically-Informed Self-Attention



Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
  - Multi-head self-attention
  - Syntactically-informed self-attention
  - Multi-task learning, single-pass inference
- Experimental results & error analysis

Experimental results

	CoNLL-2005	CoNLL-2012
domains	Train, dev: news Test: news, novels	Train, dev, test: 7 domains (news, telephone, bible, ...)
word embeddings	GloVe [Pennington et al. 2014] ELMo [Peters et al. 2018]	GloVe [Pennington et al. 2014] ELMo [Peters et al. 2018]
predicates	predicted; gold	predicted
baselines	He et al. 2017 He et al. 2018 Tan et al. 2018	He et al. 2018
our models	SA LISA LISA+D&M, +Gold	SA LISA LISA+D&M, +Gold
	Lisa_Gold	Lisa_Gold

Experimental results: CoNLL-2005

	 GloVe		 ELMo	
	in-domain	**out-of-domain**	**in-domain**	**out-of-domain**
He et al. 2017	82.7	70.1	---	---
He et al. 2018	82.5	70.8	86.0	76.1
SA	83.72	71.51	86.09	76.35
LISA	83.61	71.91	86.55	78.05
+D&M	84.99 94.9 UAS	74.66 90.3 UAS	86.90 96.3 UAS	78.25 93.4 UAS
	+2.49 F1	+3.86 F1	+0.9 F1 ?	+2.15 F1

Experimental results: CoNLL-2005

	GloVe	ELMo 
	in-domain (dev)	in-domain (dev)

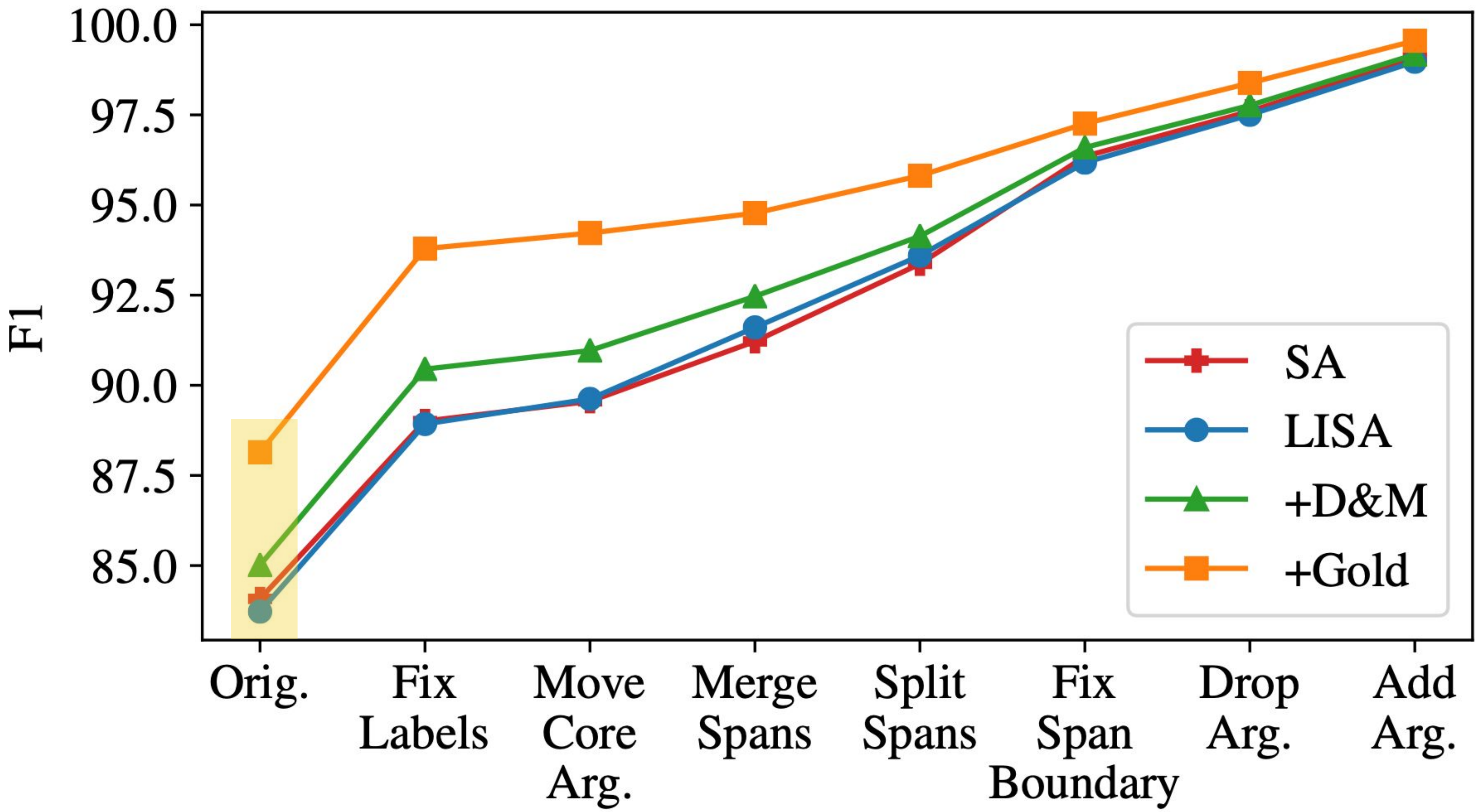
96.5 UAS!

Experimental results: Validation

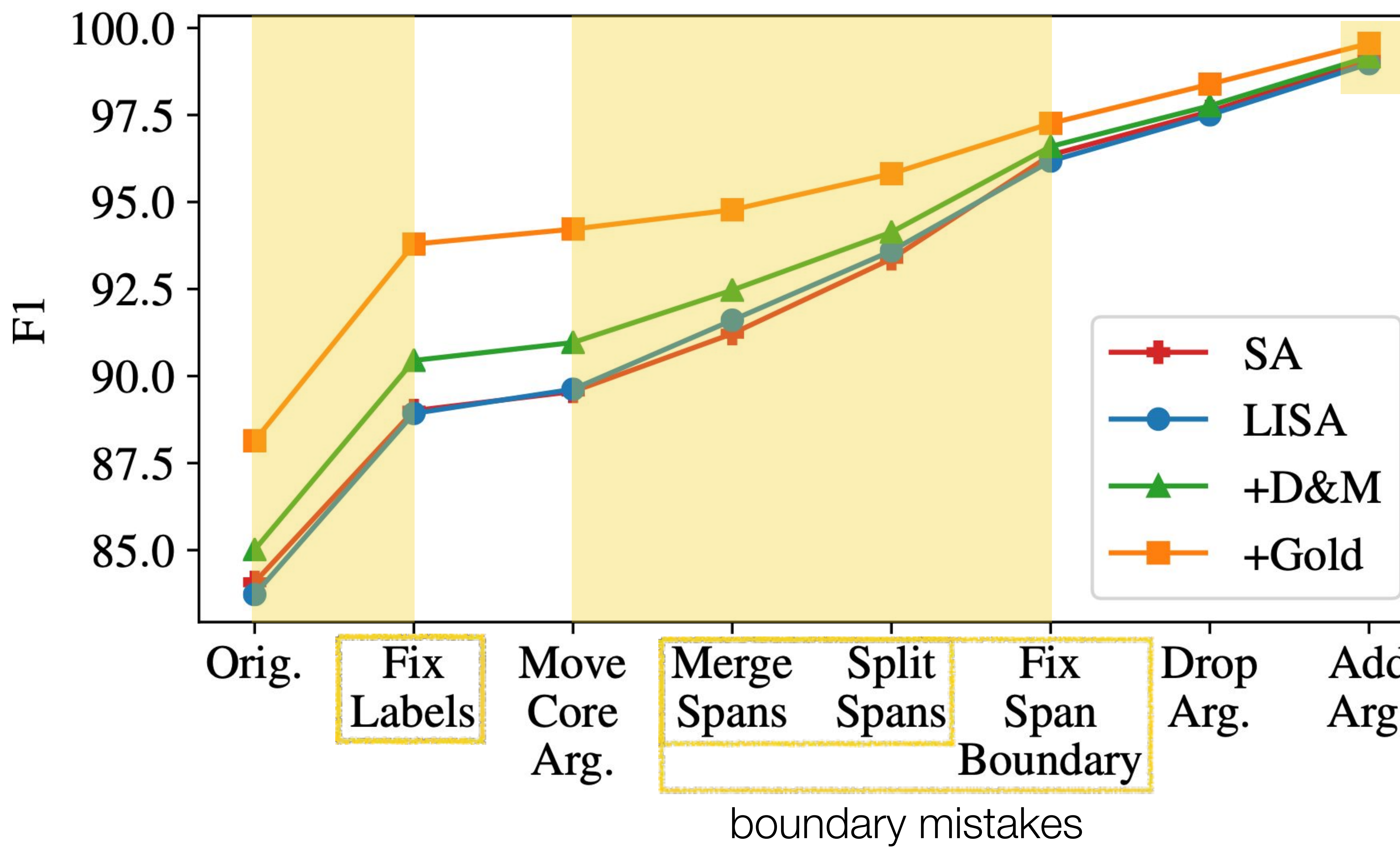
	GloVe		ELMo	
	CoNLL-05	**CoNLL-12**	**CoNLL-05**	**CoNLL-12**
+Gold ΔF1	*5.21*	*7.03*	*2.33*	*4.36*
+D&M ΔF1	1.98	2.65	-0.13	0.36

	GloVe		ELMo	
	CoNLL-05	**CoNLL-12**	**CoNLL-05**	**CoNLL-12**
LISA UAS	94.92	93.35	96.48	94.84
D&M UAS	---	---	96.28	94.99

Experimental results: Analysis



Experimental results: Analysis



Summary

Thank you!

- **LISA**: Multi-task learning + multi-head self attention trained to attend to syntactic parents
  - Achieves state-of-the-art F1 on PropBank SRL
  - Linguistic structure improves generalization
  - Fast: encodes sequence ***only once*** to predict predicates, parts-of-speech, labeled dependency parse, SRL

- Everyone wants to run NLP on the entire web:
  - **accuracy**, **robustness**, **computational efficiency**.

Models & Code: https://github.com/strubell/LISA

I am on the academic job market this spring!

Experimental results: CoNLL-2005

Gold predicates; GloVe embeddings 

WSJ Test (in-domain):

	Precision	**Recall**	**F1**
He et al. 2018	84.2	83.7	83.9
Tan et al. 2018	81.2	83.9	84.8
SA	84.7	84.24	84.47
LISA	84.72	84.57	84.64
+D&M	86.02	86.05	86.04

Brown Test (out-of-domain):

	Precision	**Recall**	**F1**
He et al. 2018	74.2	73.1	73.7
Tan et al. 2018	73.5	74.6	74.1
SA	73.89	72.39	73.13
LISA	74.77	74.32	74.55
+D&M	76.65	76.44	76.54

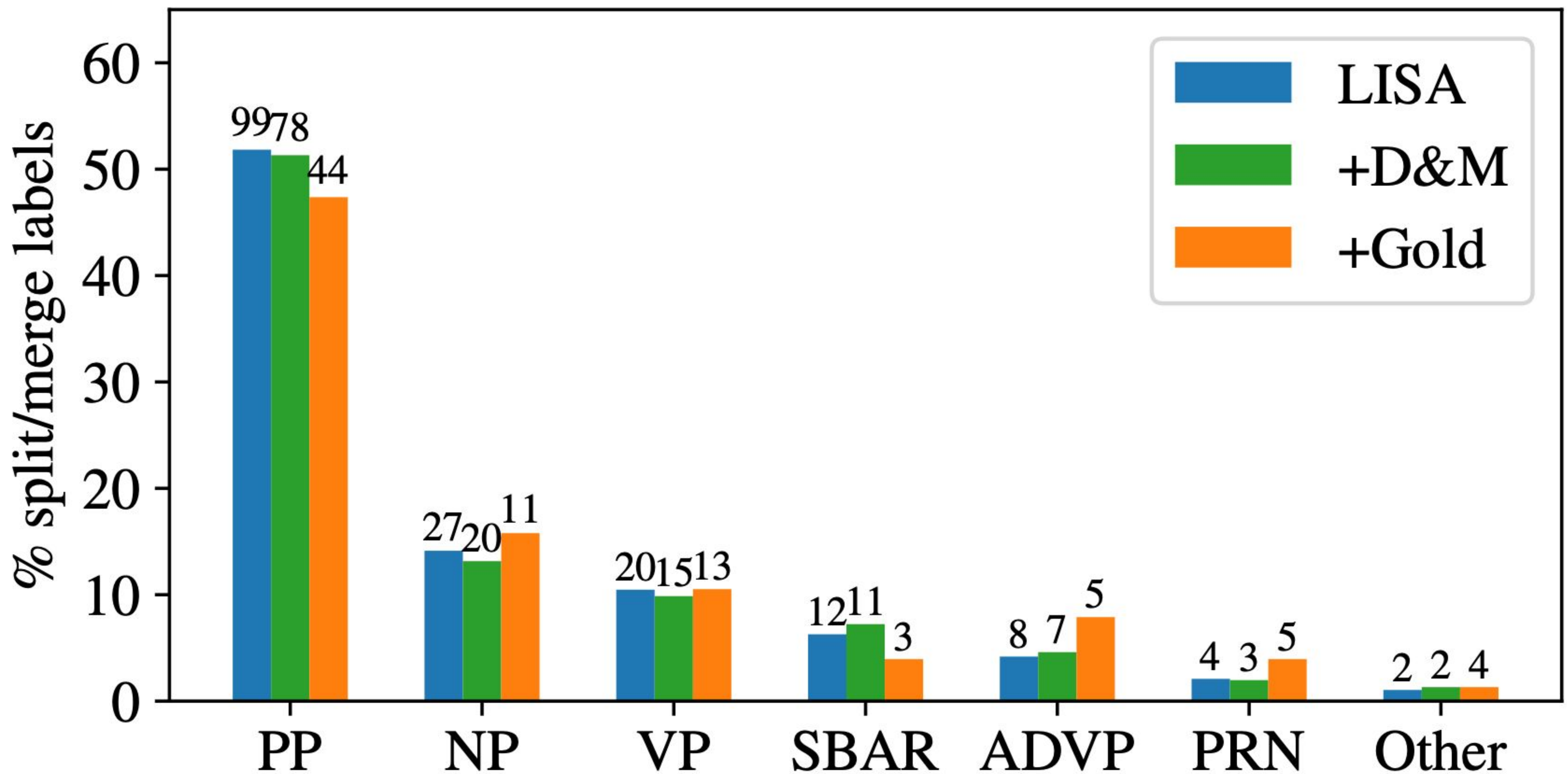
Experimental results: CoNLL-2012

Predicted predicates

	Precision	**Recall**	**F1**
He et al. 2018	79.4	80.1	79.8
SA	82.55	80.02	81.26
LISA	81.86	79.56	80.70
+D&M	83.3	81.38	82.33

	Precision	**Recall**	**F1**
He et al. 2018	81.9	84.0	82.9
SA	84.39	82.21	83.28
LISA	83.97	82.29	83.12
+D&M	84.14	82.64	83.38

Experimental results: Analysis



Experimental results: Analysis

	L+/D+	**L-/D+**	**L+/D-**	**L-/D-**
Proportion	26%	12%	4%	56%
SA	79.29	75.14	75.97	75.08
LISA	79.51	74.33	79.69	75.00
+D&M	79.03	76.96	77.73	76.52
+Gold	*79.61*	*78.38*	*81.41*	*80.47*

Experimental results: Analysis