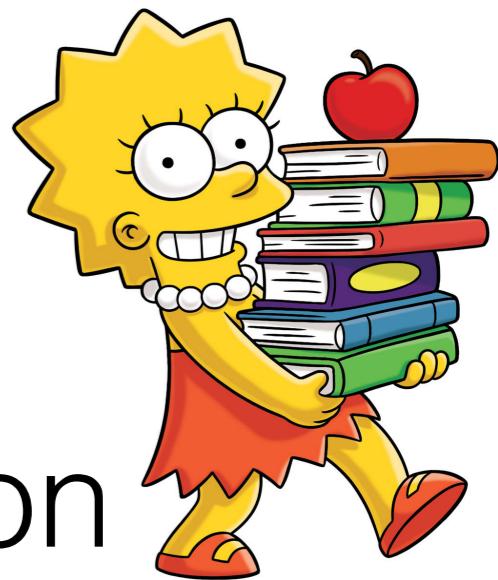


LISA

Linguistically-Informed Self-Attention for Semantic Role Labeling



Emma
Strubell¹



Patrick
Verga¹



Daniel
Andor²



David
Weiss²



Andrew
McCallum¹

¹ UMassAmherst



College of Information
and Computer Sciences

²



Google AI

Want fast, accurate, robust NLP

Nobel laureate Donna Strickland: 'I see myself as a scientist, not a woman'

For Just the Third Time in 117 Years, a Woman Wins the Nobel Prize in Physics



The 2018 Nobel laureate in Physics, Donna Strickland, is shown smiling. She has long, light-colored hair and is wearing dark-rimmed glasses. The background is slightly blurred, showing what appears to be a laboratory or office environment.

Sections ▾

The Washington Post
Democracy Dies in Darkness

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.

Volume 56, number 3 OPTICS COMMUNICATIONS 1 December 1985

COMPRESSION OF AMPLIFIED CHIRPED OPTICAL PULSES *

Donna STRICKLAND and Gerard MOUROU
Laboratory for Laser Energetics, University of Rochester, 250 East River Road, Rochester, NY 14623-1299, USA

Received 5 July

We have demo 1.06 μm laser pu

The onset of self-limiting the amplification of a similar problem arises for short, yet energetic pulses capable of handling the resolution for radar transponders by passing it through a lens before amplifying an echo is compressed to a negatively dispersive pulse.

We wish to report the technique employed and that in principle a short ($\lesssim 1$ ps) pulse can be compressed to a short, low-energy pulse. The pulse is linearly chirped by combination of group velocity modulation [2]. The pulse is then compressed by amplifying the stretched pulse. Amplifying the compressed pulse allows the pulse to be focused before self-focusing does not appear to affect the pulse. Pulses can be fully compressed without the risk of amplifying a chocked medium if gain sweet spots are avoided.

* This is a corrected version. Comm. 55 (1985) 447 was printed as fig. 1.

0 030-4018/85/\$03.25 (North-Holland Physics Publishing)

STEM Gems
@STEMGemsBook

Follow

#STEMGems 💎 "Physicist
#DonnaStrickland, a self-described 'laser
jock' who prefers to keep a low profile, won
the @NobelPrize in Physics in 2018.
3rd woman ever to win the Nobel Prize in Physics.
She calls 'surviving' her work.
#GirlsInSTEM

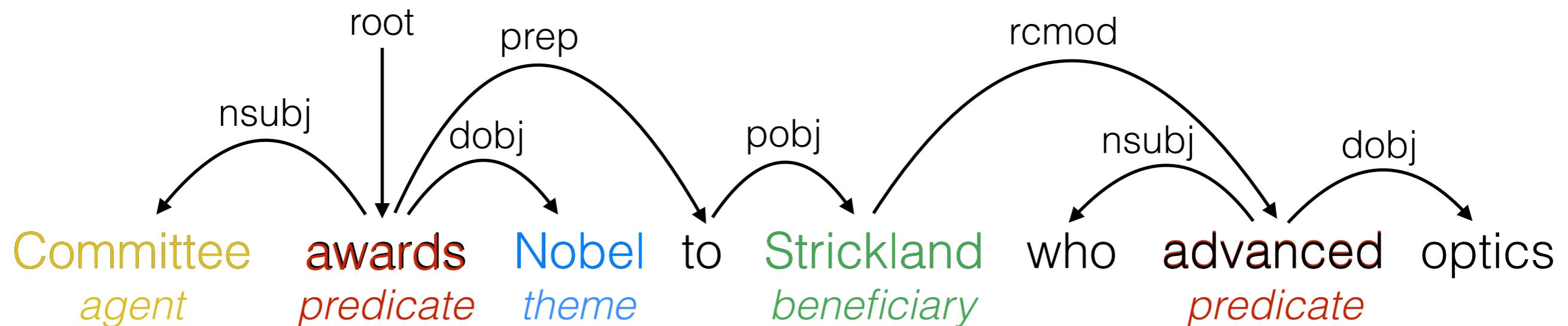
SECOND EDITION

NOBEL PRIZE WOMEN IN SCIENCE

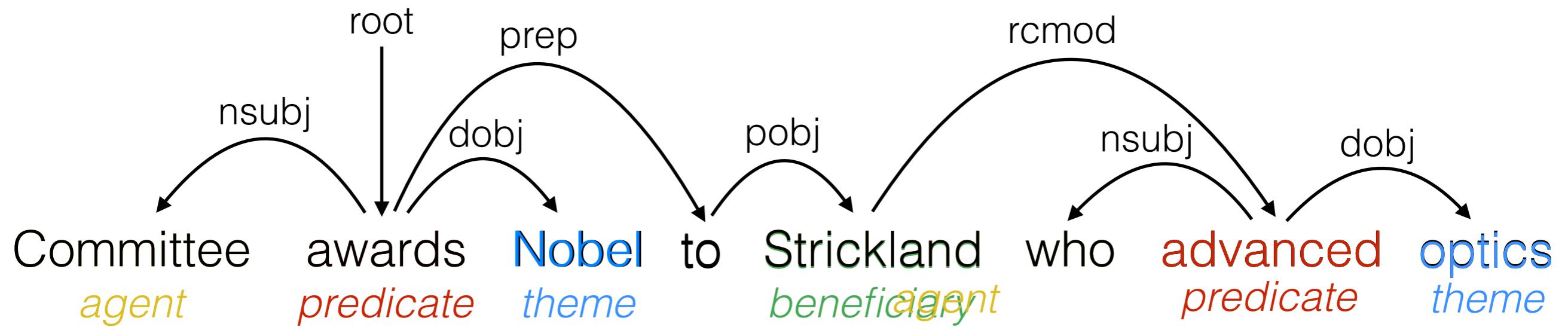
Their Lives, Struggles, and Momentous Discoveries

SHARON BERTSCH McGRAWNE

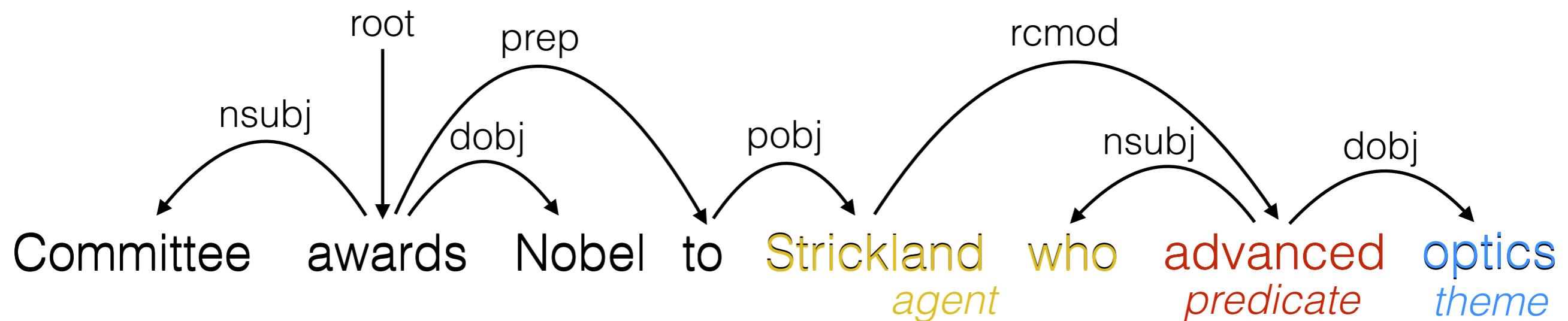
SRL: Who did what to whom?



SRL: Who did what to whom?

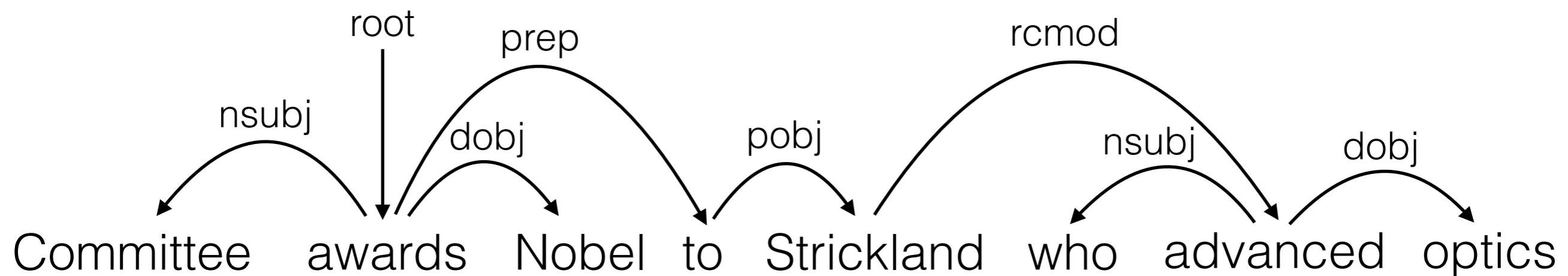


SRL: Who did what to whom?



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

PropBank SRL: Who did what to whom?



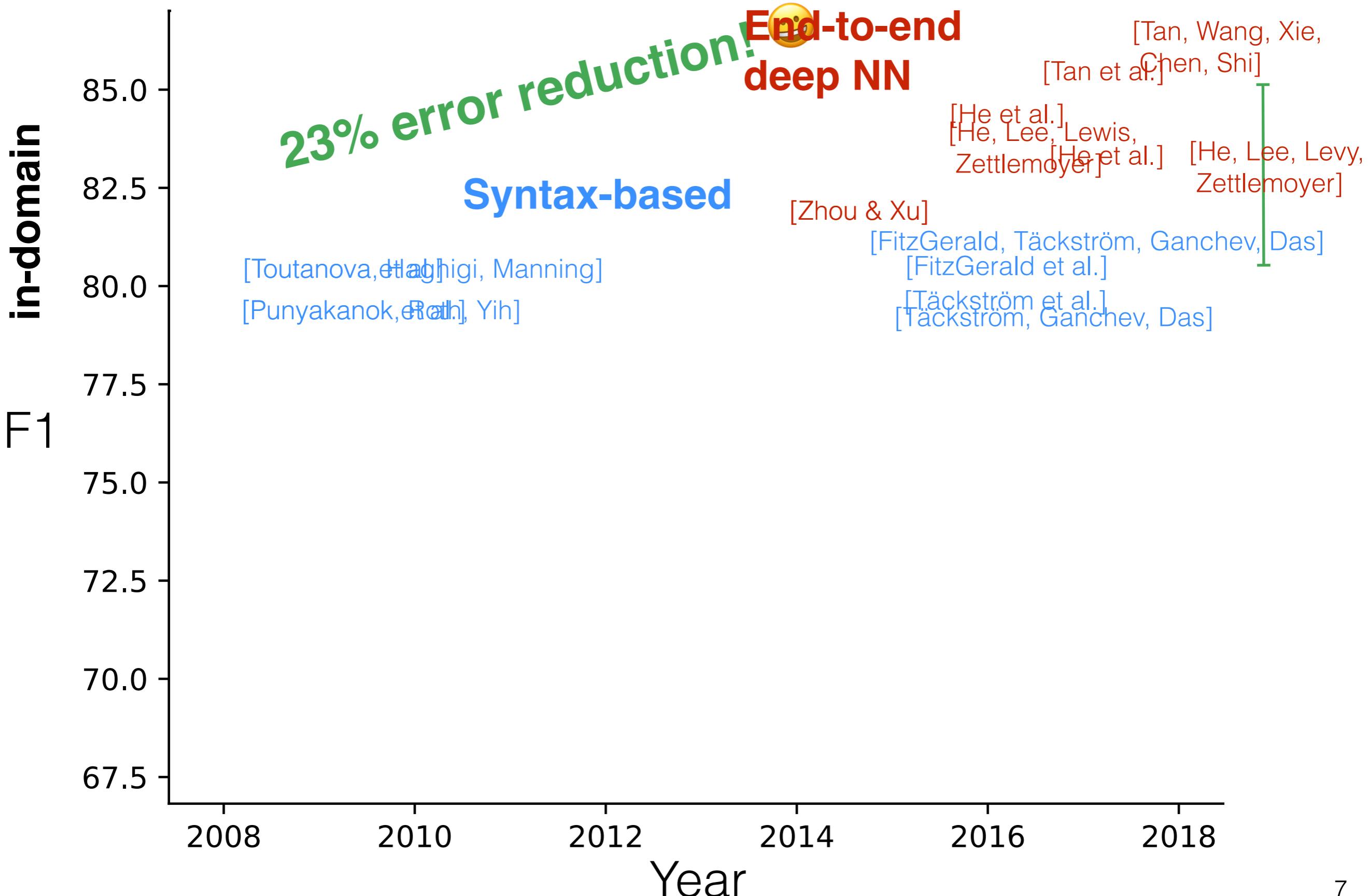
Committee awards Nobel to Strickland who advanced optics

ARG_{agent} R-ARG₀ predicate tARG₁

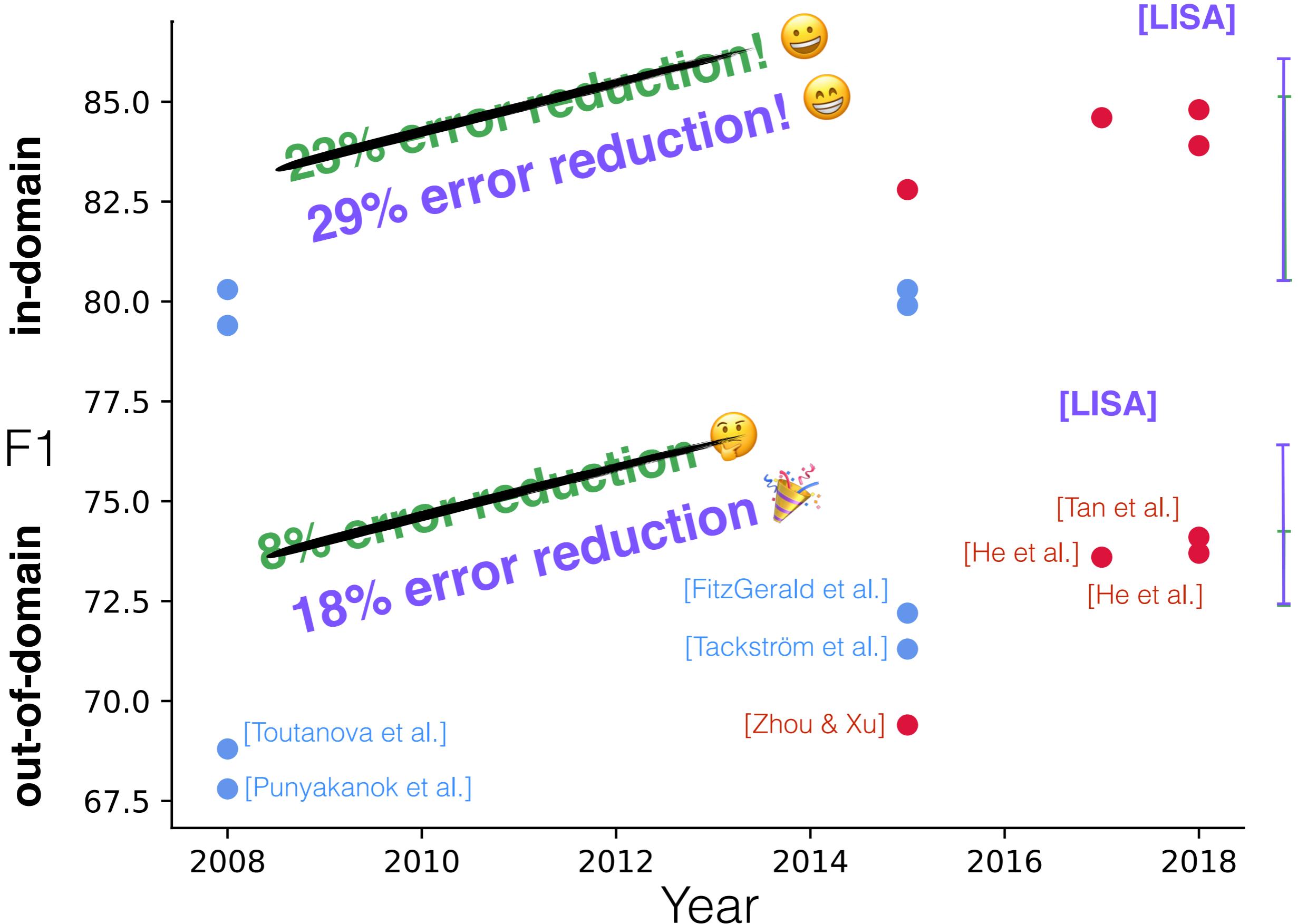
Committee awards Nobel to Strickland who advanced optics

Agent predicate Arg₀ Arg₁

10 years of PropBank SRL

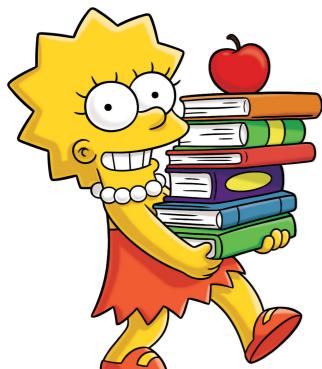


10 years of PropBank SRL



Linguistically-Informed Self-Attention

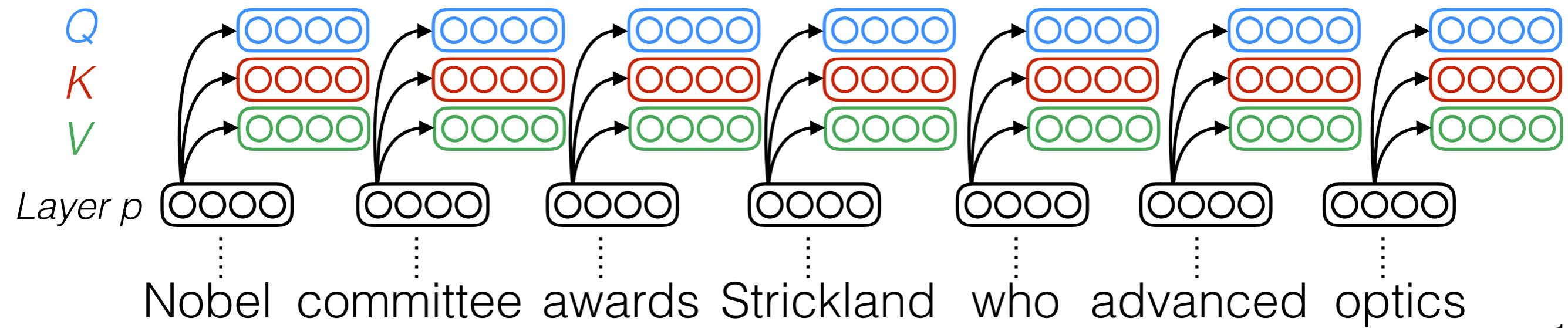
- **Multi-task learning**
 - Part-of-speech tagging
 - Labeled dependency parsing
 - Predicate detection
 - Semantic role spans & labeling
 - **Syntactically-informed self-attention**
 - Multi-head self-attention supervised by **syntax**
- 



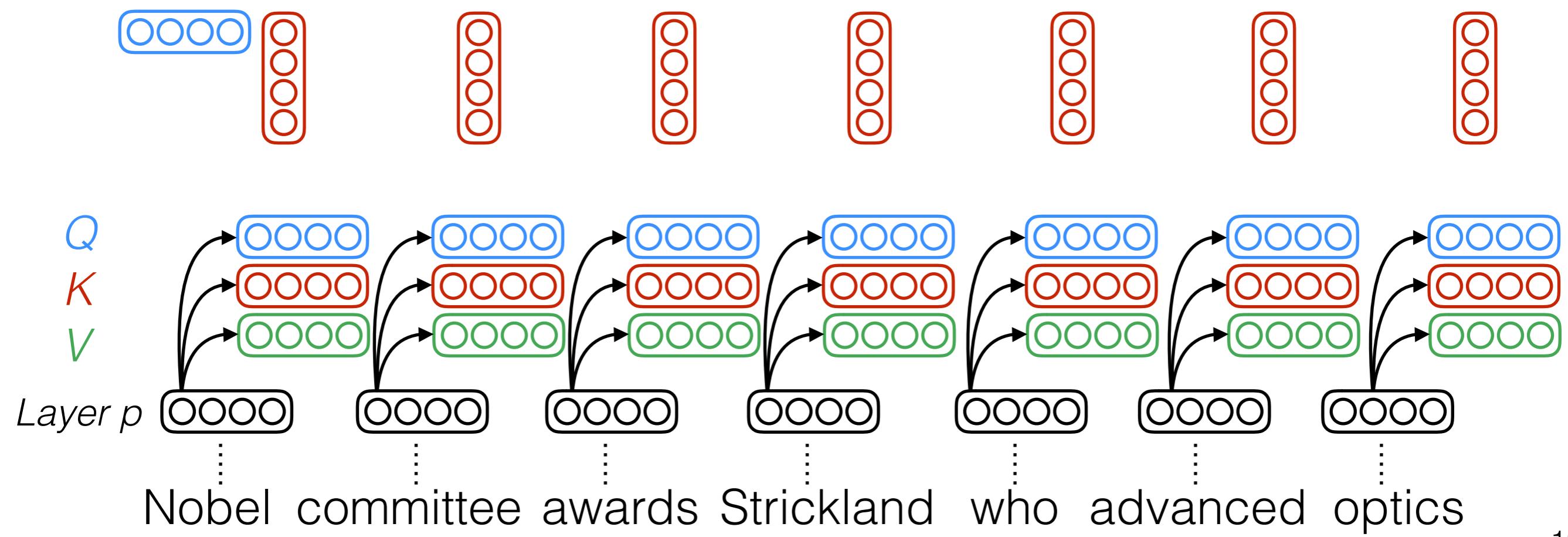
Outline

- 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
 - Training
 - Experimental results & error analysis

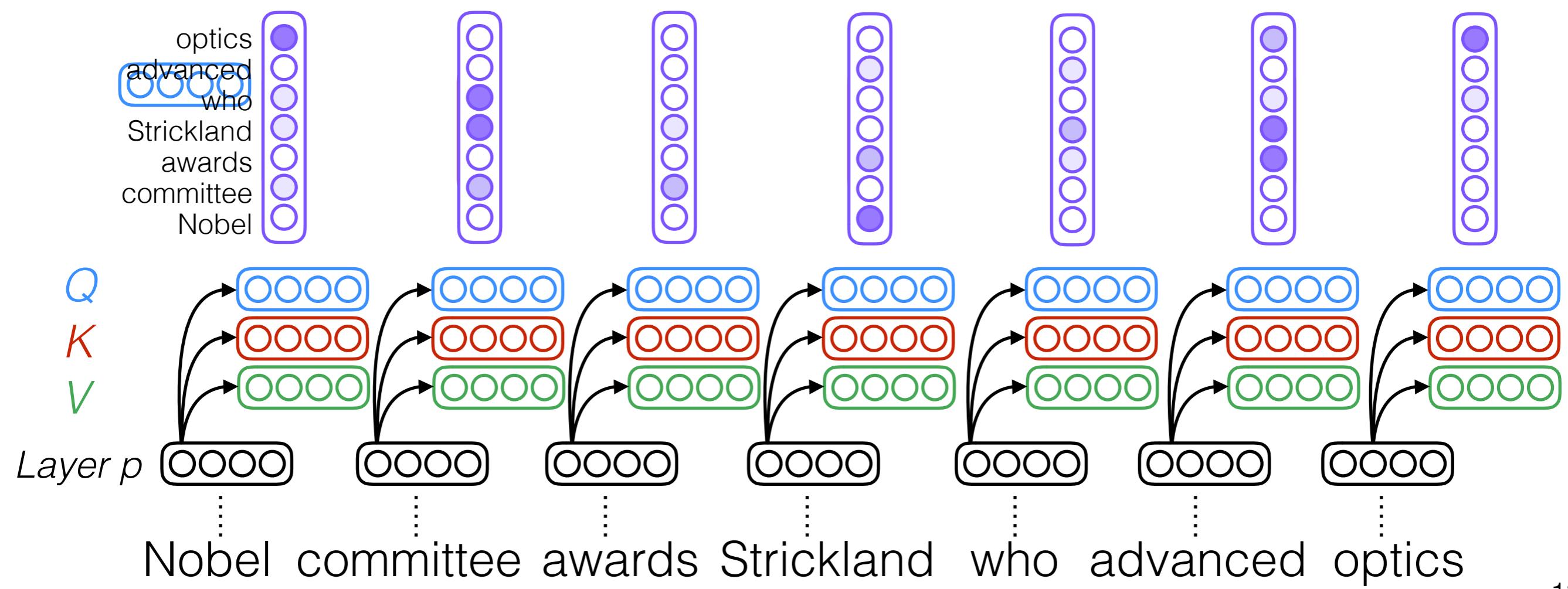
Self-attention



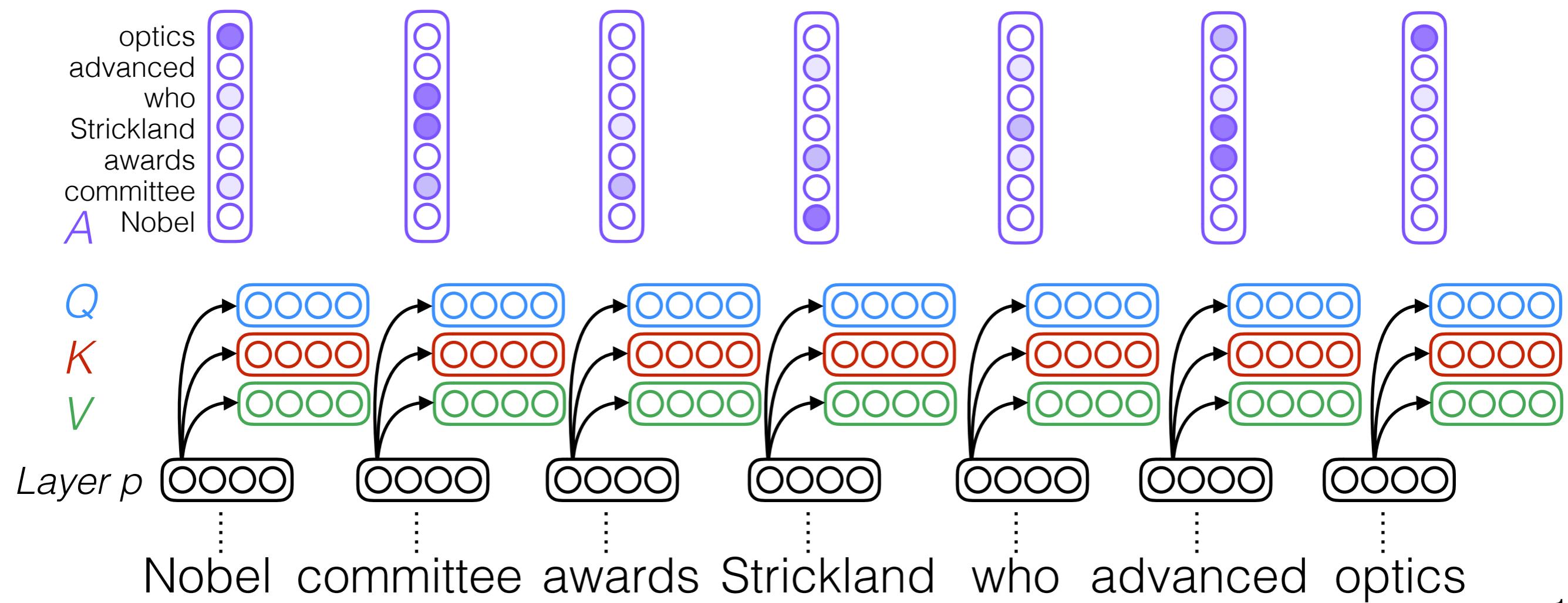
Self-attention



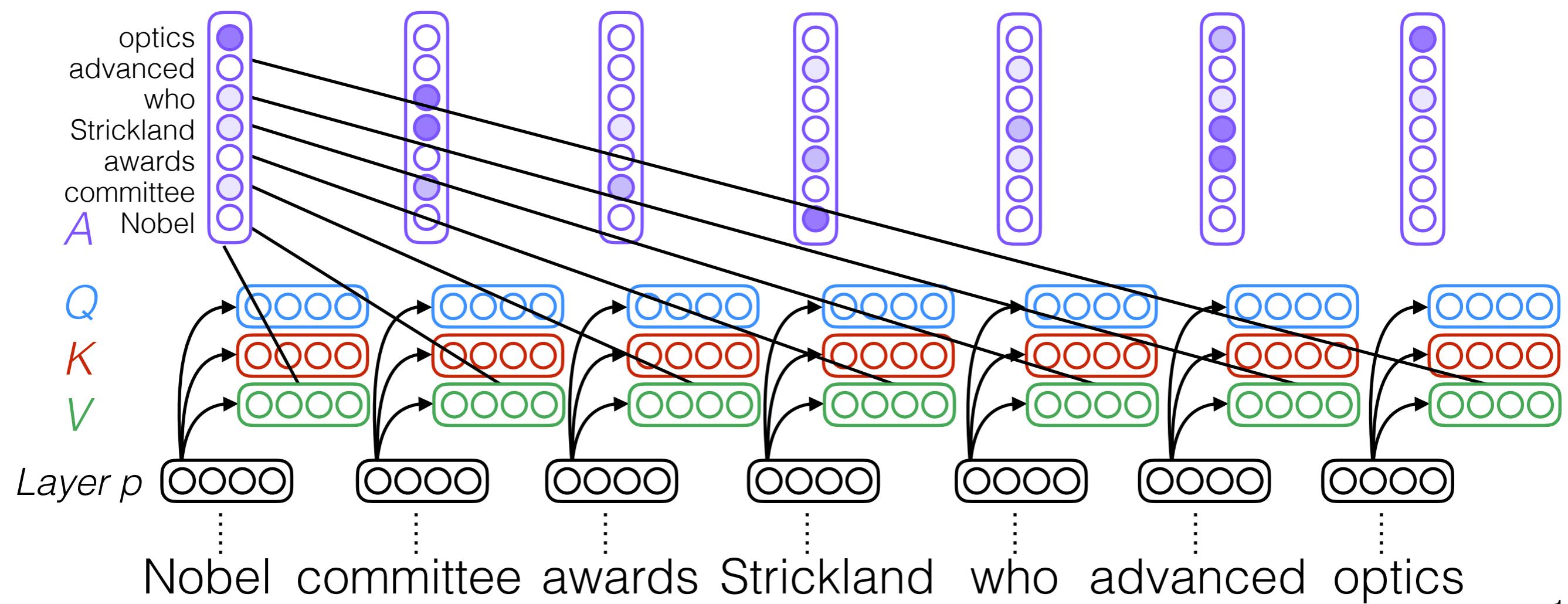
Self-attention



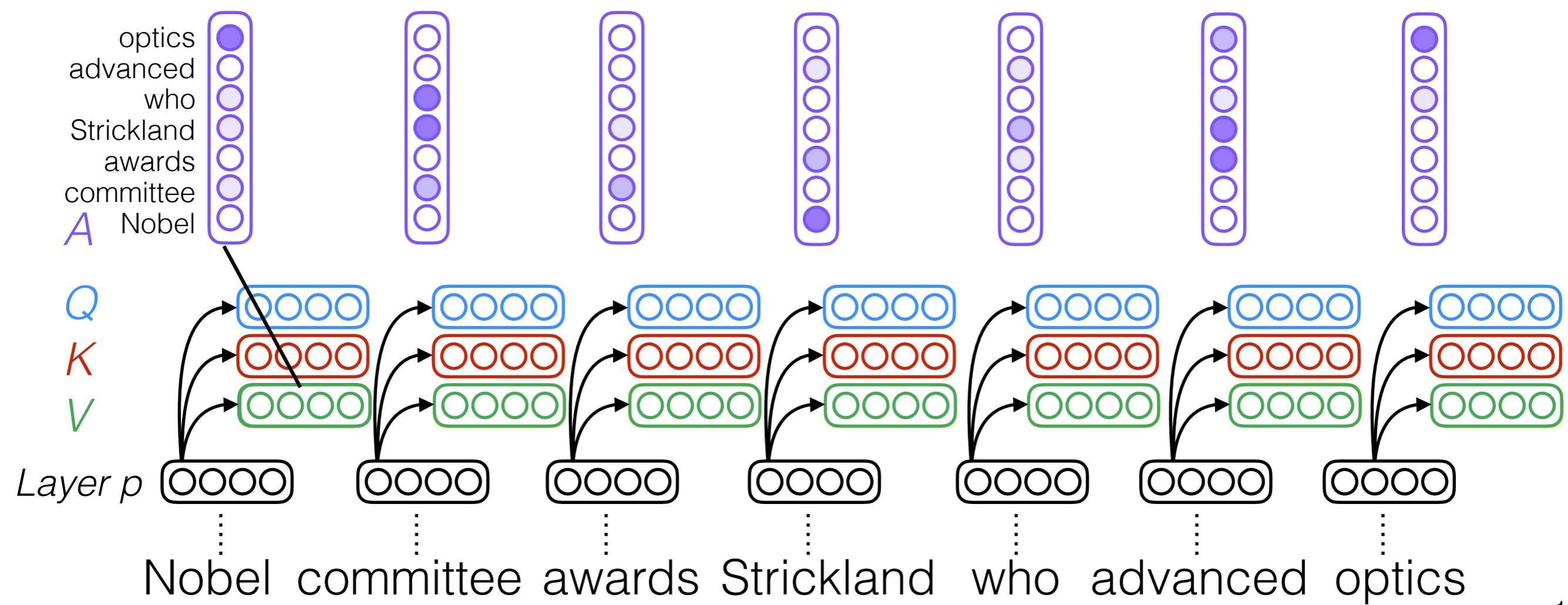
Self-attention



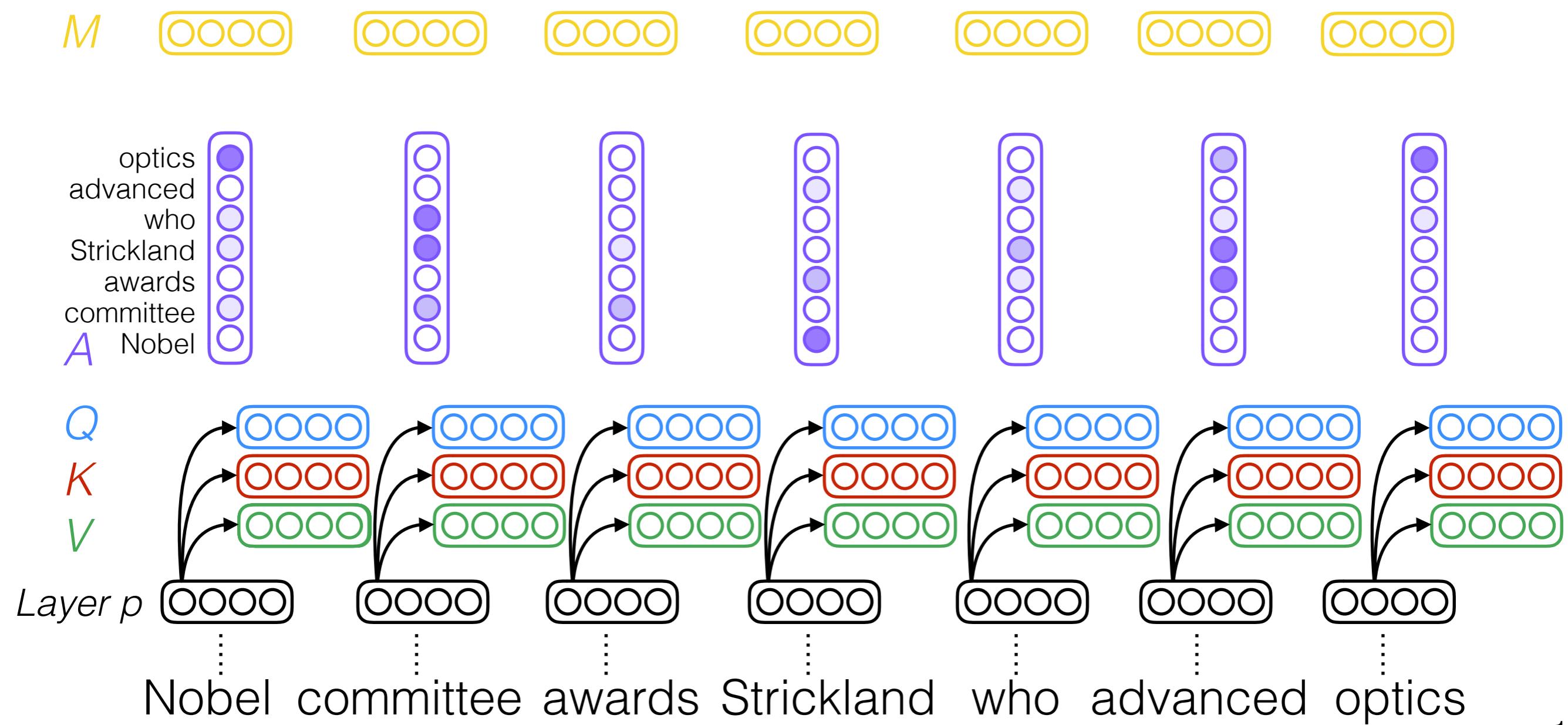
Self-attention



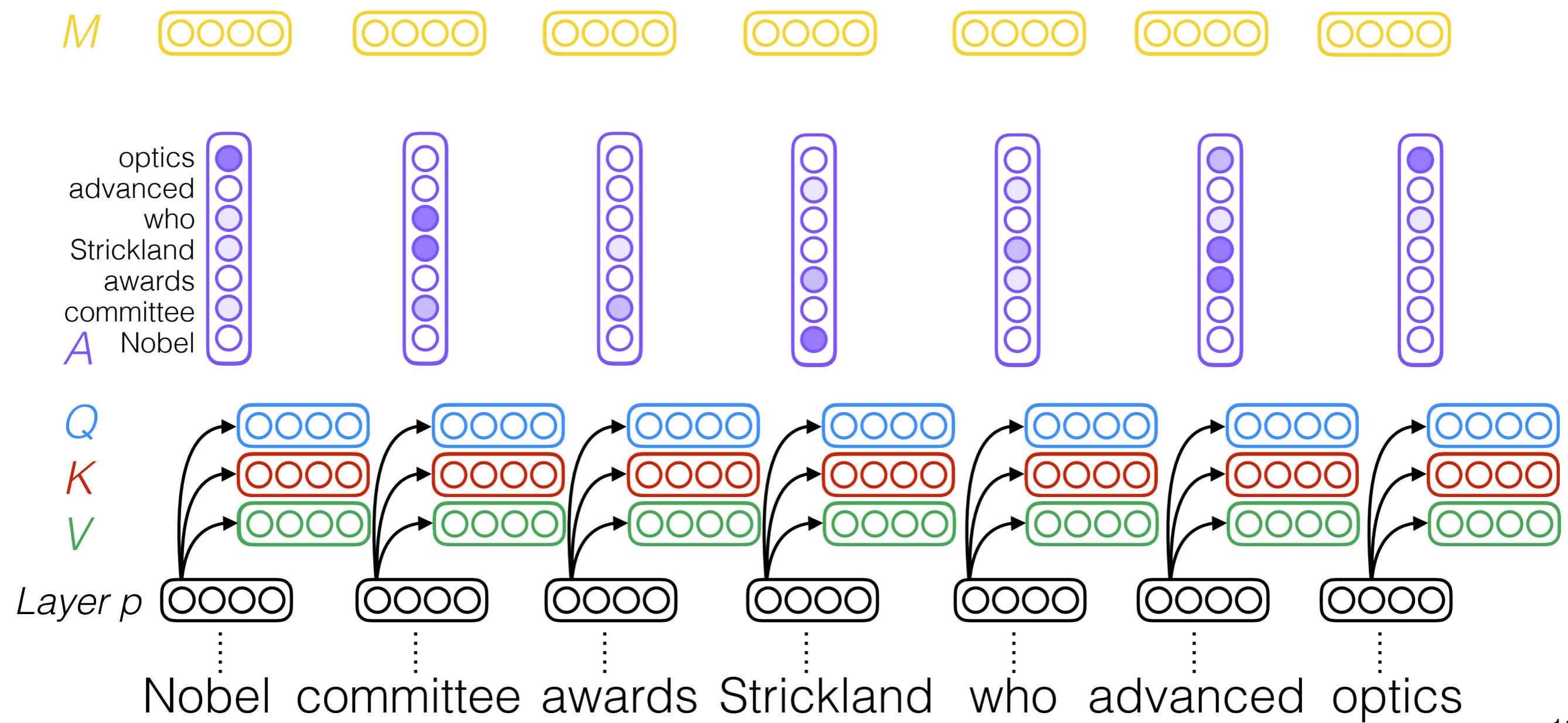
Self-attention



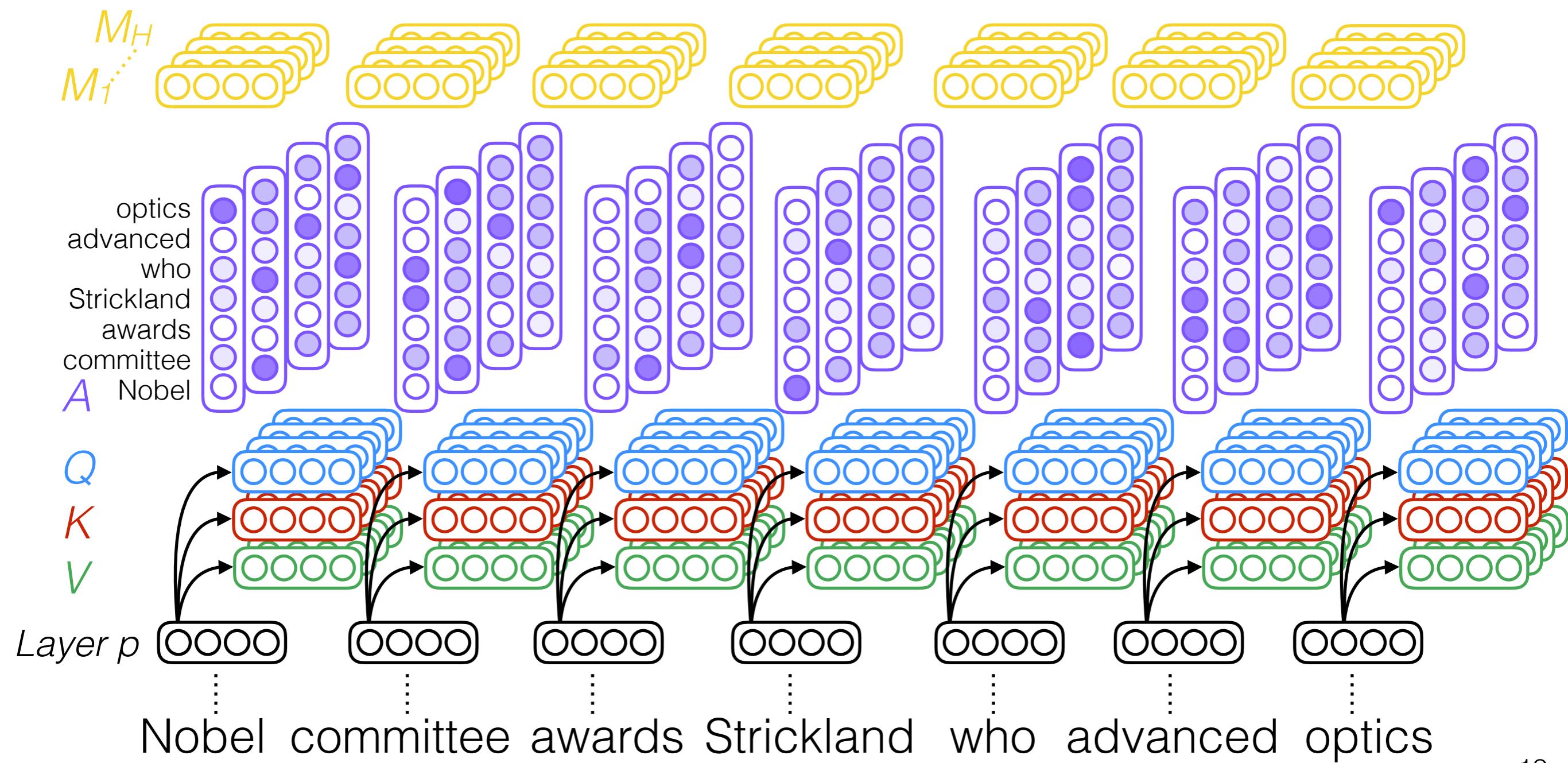
Self-attention



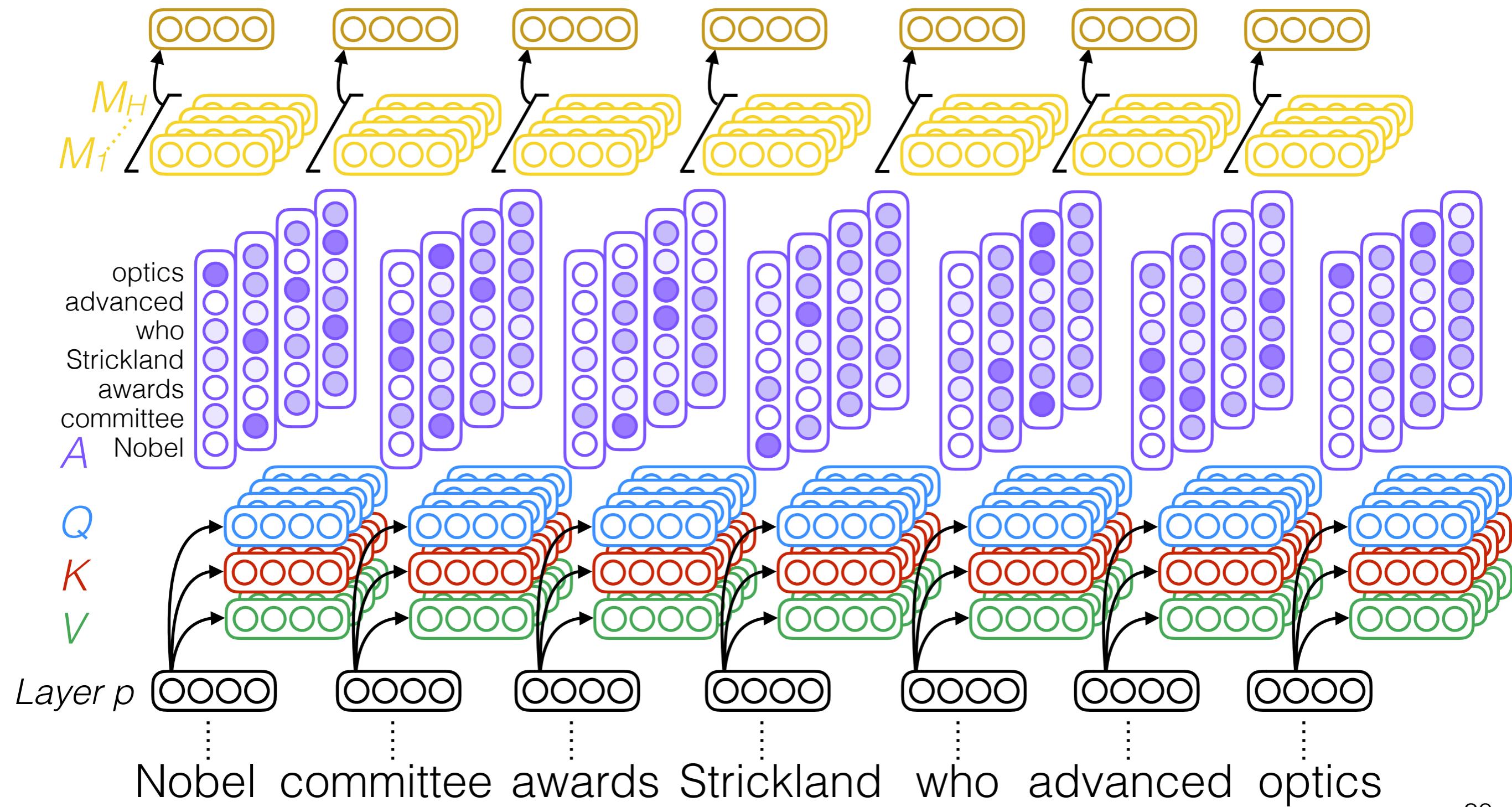
Self-attention



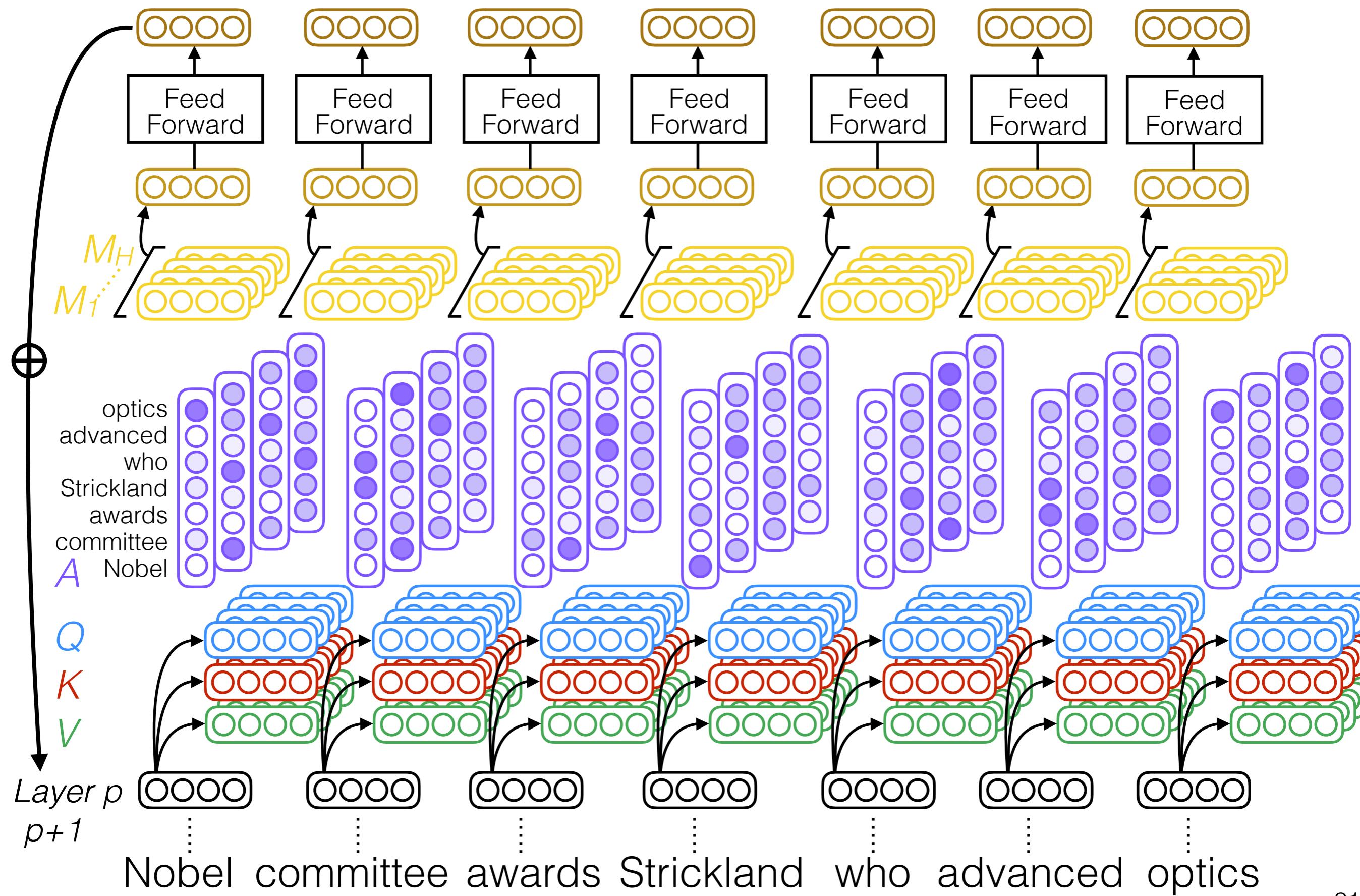
Multi-head self-attention



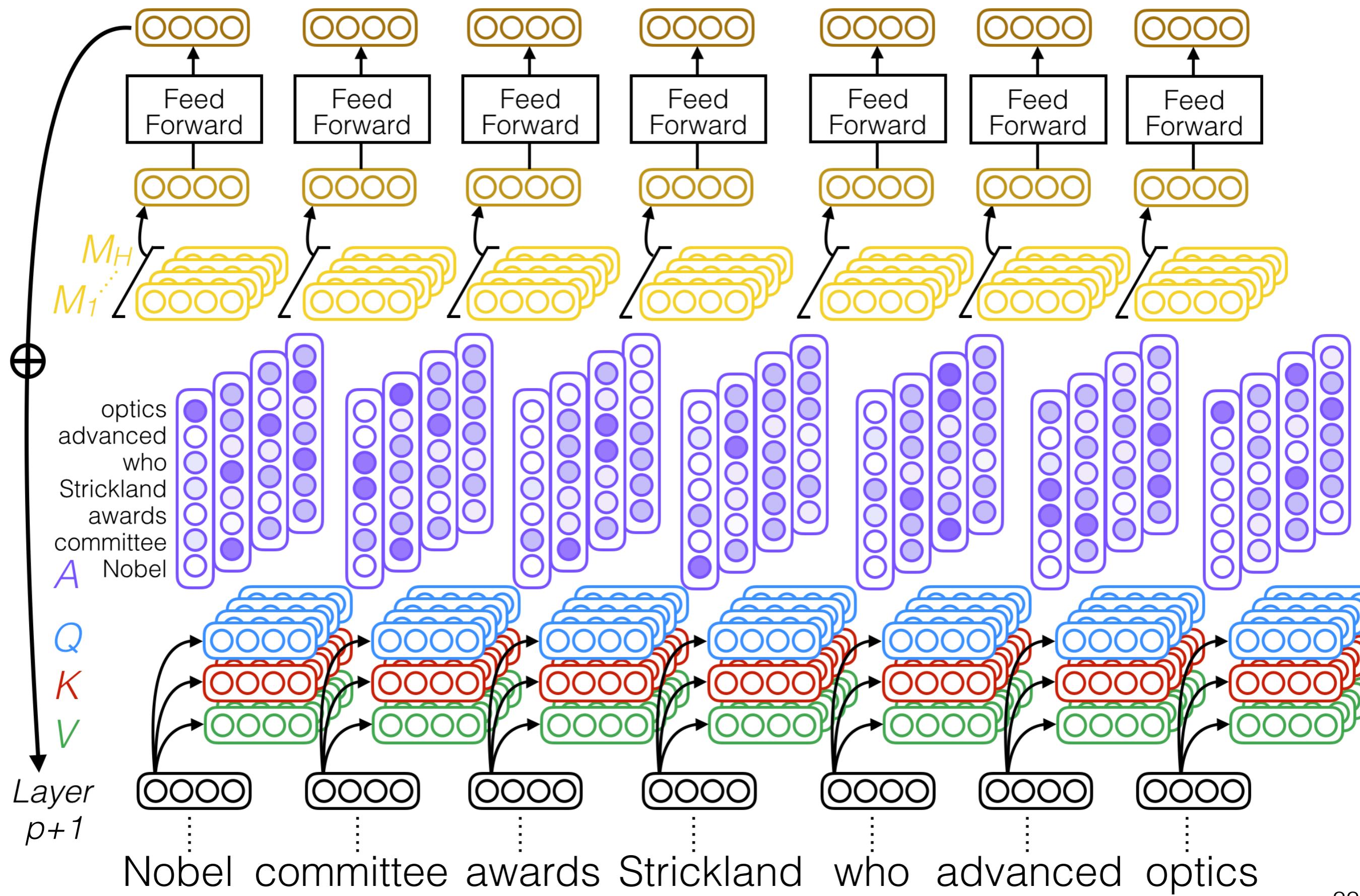
Multi-head self-attention



Multi-head self-attention

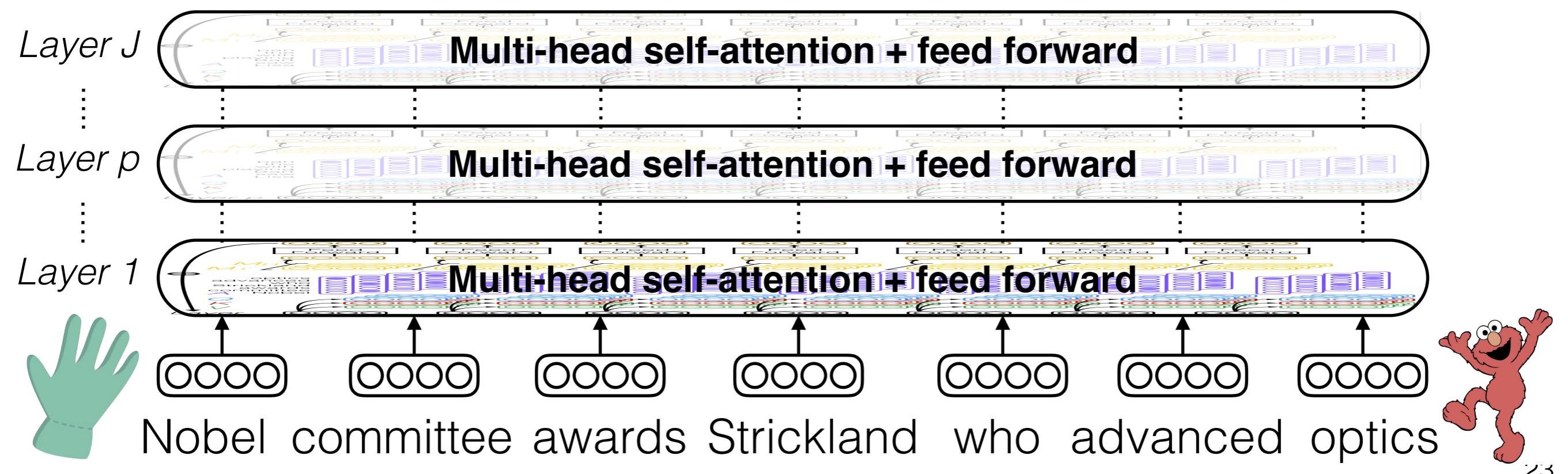


Multi-head self-attention



Multi-head self-attention

$$s_t^{(j)} = LN(s_t^{(j-1)} + T^{(j)}(s_t^{(j-1)}))$$



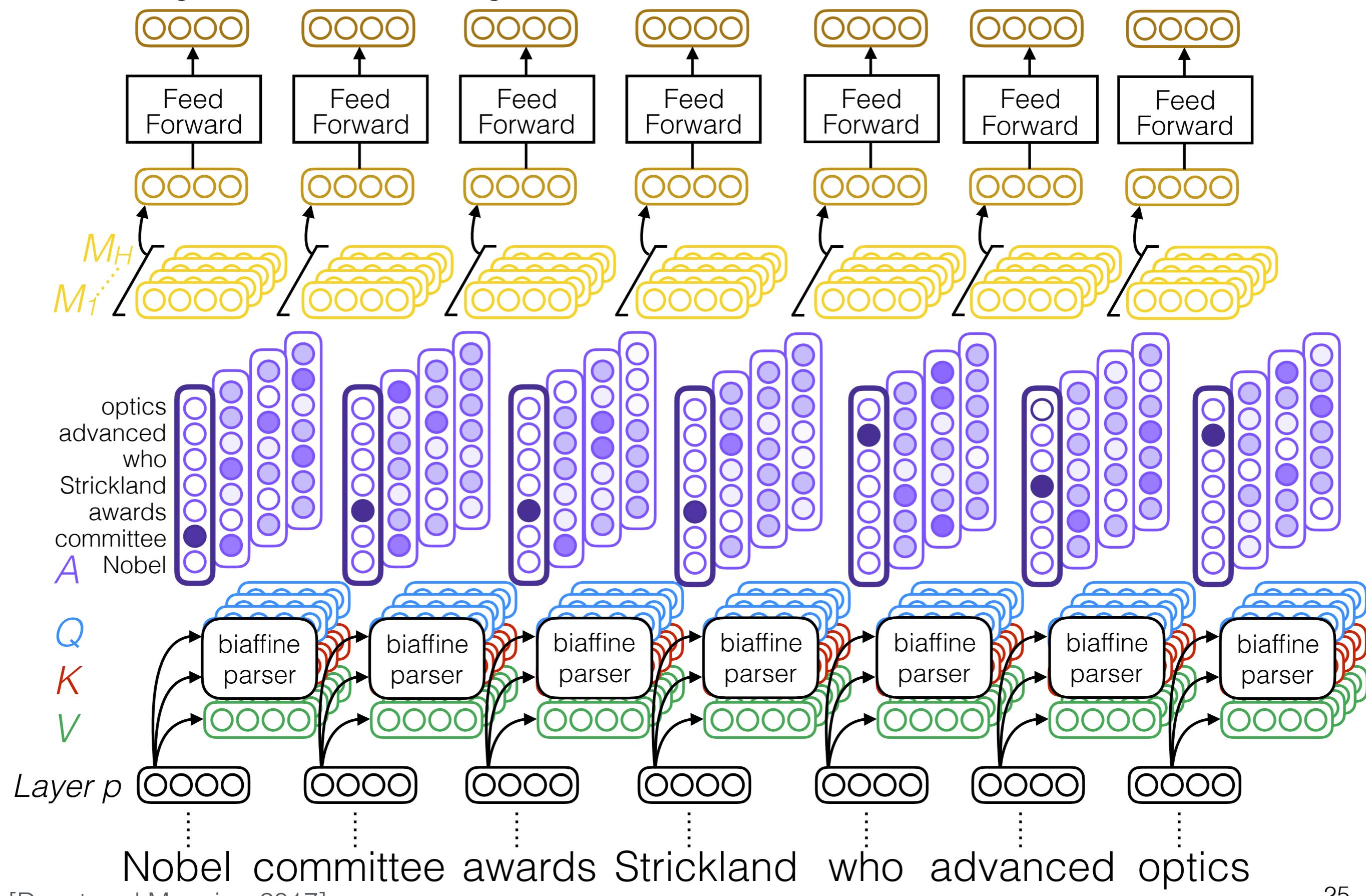
Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention

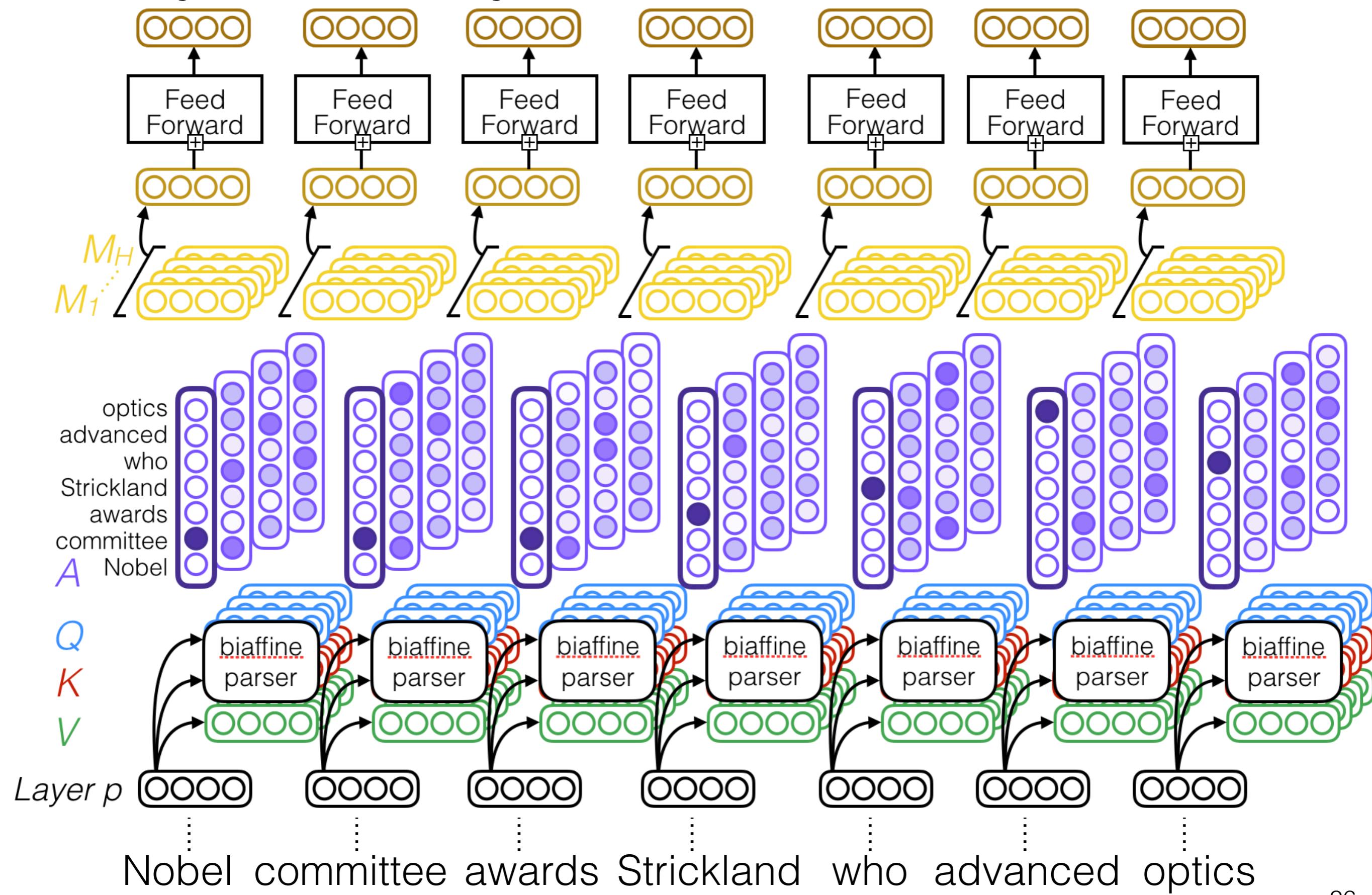
- Multi-head self-attention [Vaswani et al. 2017]

- Syntactically-informed self-attention
- Multi-task learning, single-pass inference
- Training
- Experimental results & error analysis

Syntactically-informed self-attention



Syntactically-informed self-attention



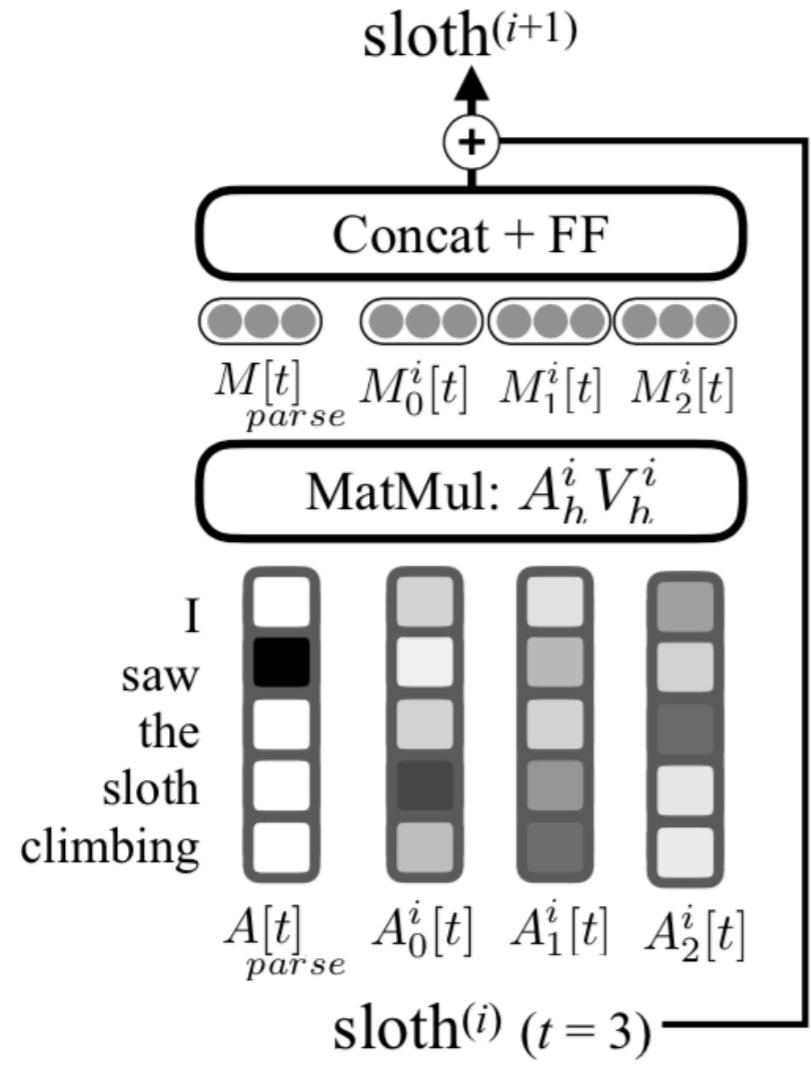


Figure 2: Syntactically-informed self-attention for the query word *sloth*. Attention weights A_{parse} heavily weight the token’s syntactic governor, *saw*, in a weighted average over the token values V_{parse} . The other attention heads act as usual, and the attended representations from all heads are concatenated and projected through a feed-forward layer to produce the syntactically-informed representation for *sloth*.

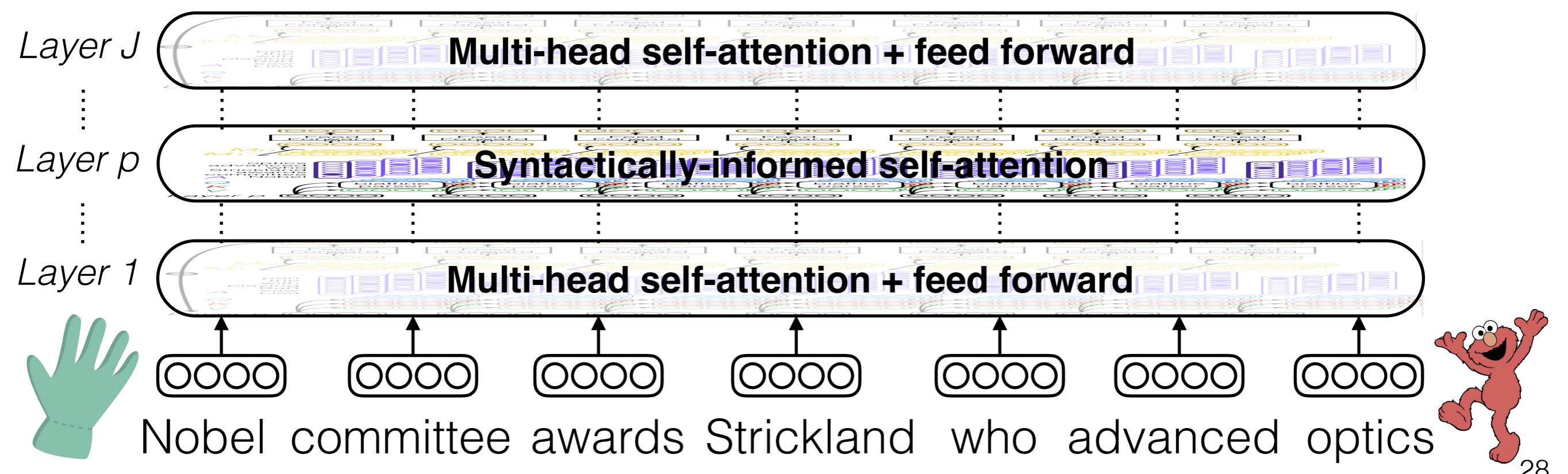
$$A_h^{(j)} = \text{softmax}(d_k^{-0.5} Q_h^{(j)} K_h^{(j)T})$$

$$M_h^{(j)} = A_h^{(j)} V_h^{(j)}$$

$$A_{parse} = \text{softmax}(Q_{parse} U_{heads} K_{parse}^T)$$

$$P(q = \text{head}(t) \mid \mathcal{X}) = A_{parse}[t, q]$$

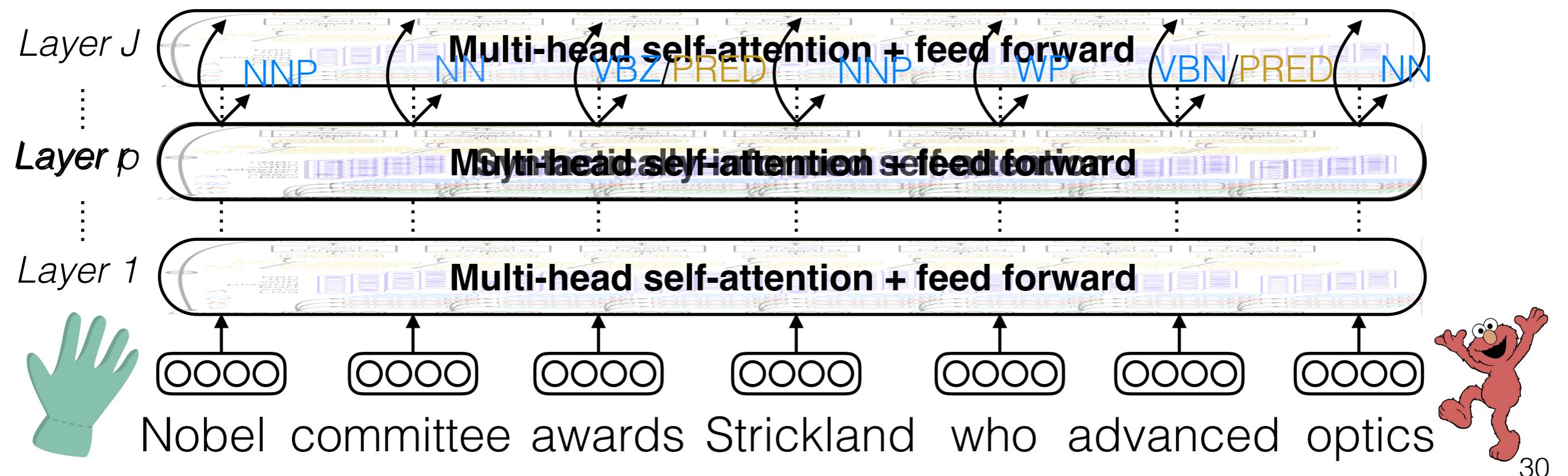
Syntactically-informed self-attention



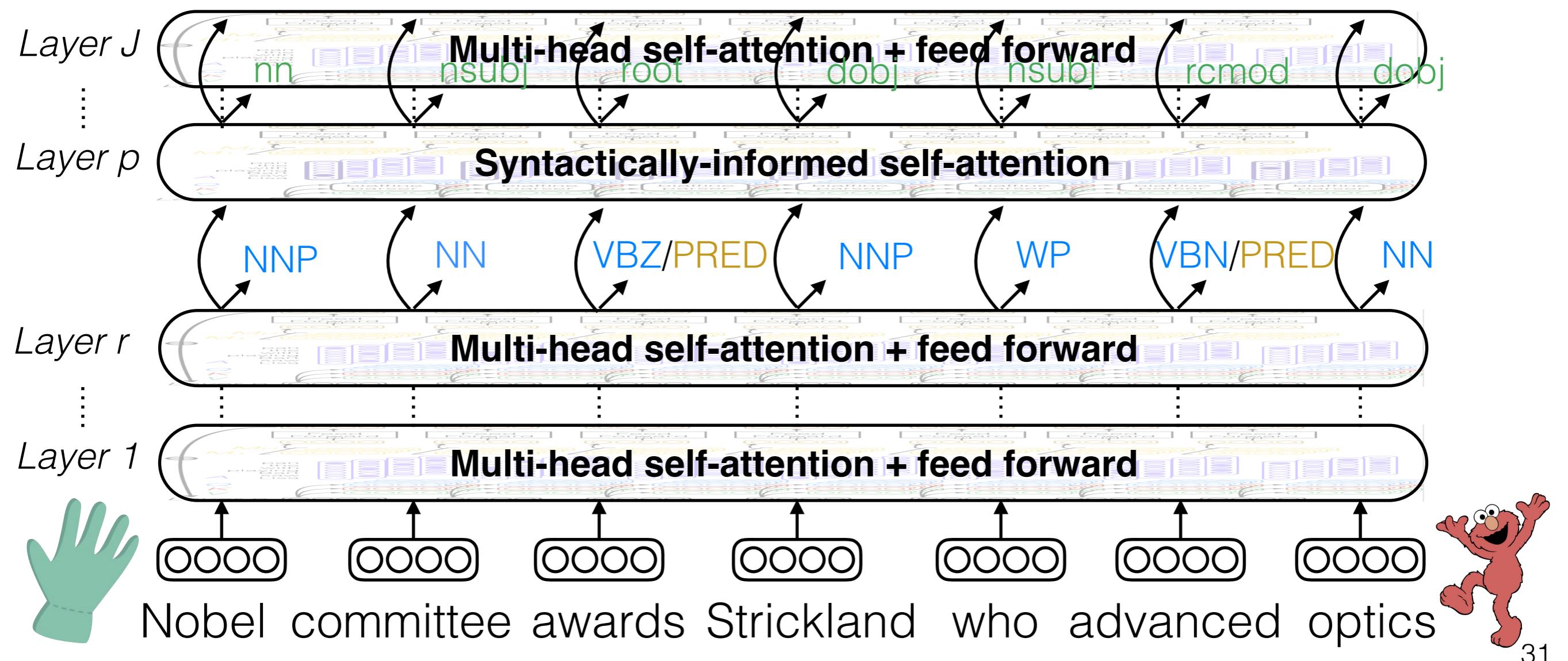
Outline

- 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
 - Training
 - Experimental results & error analysis

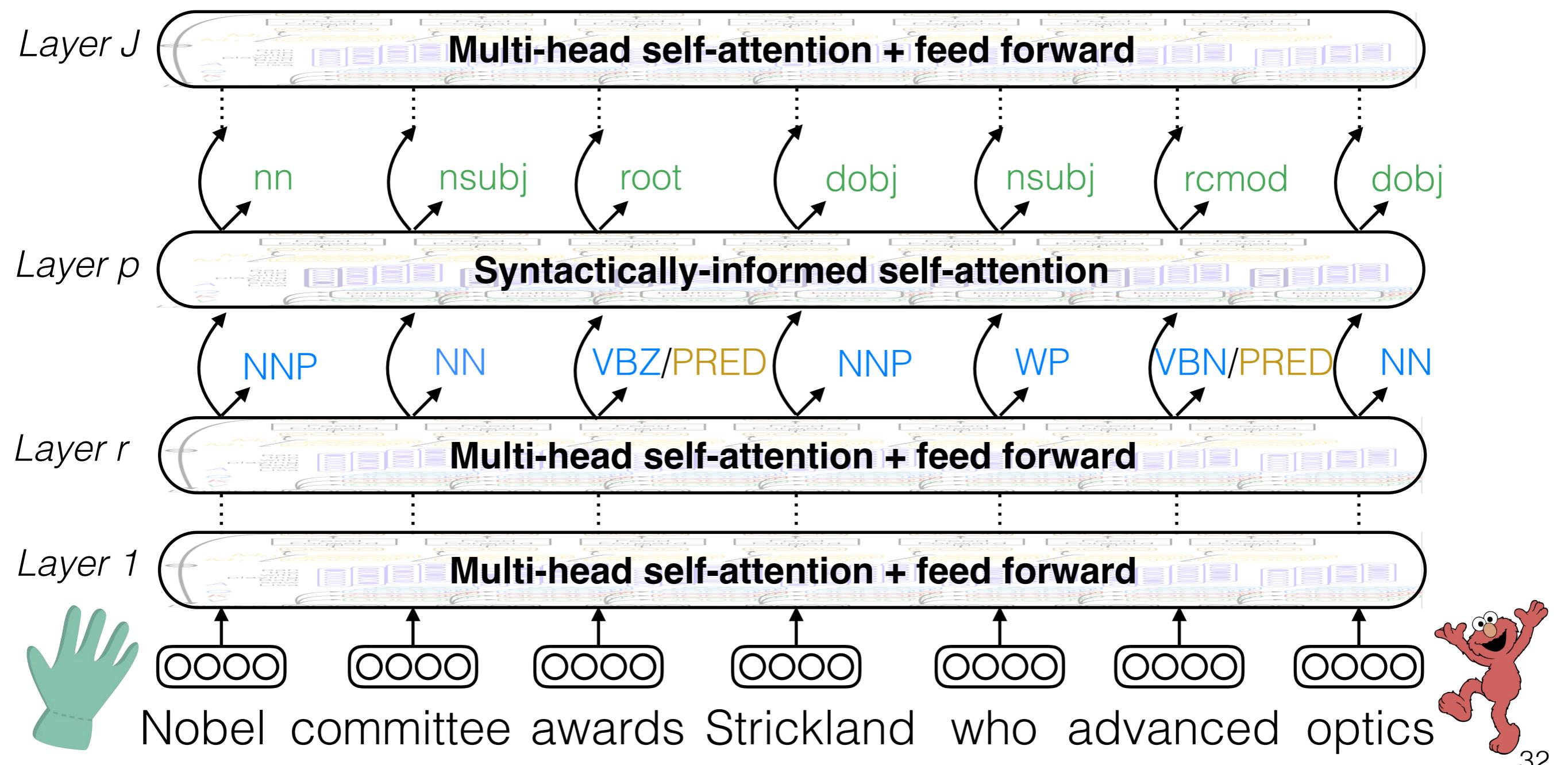
LISA: Linguistically-Informed Self-Attention



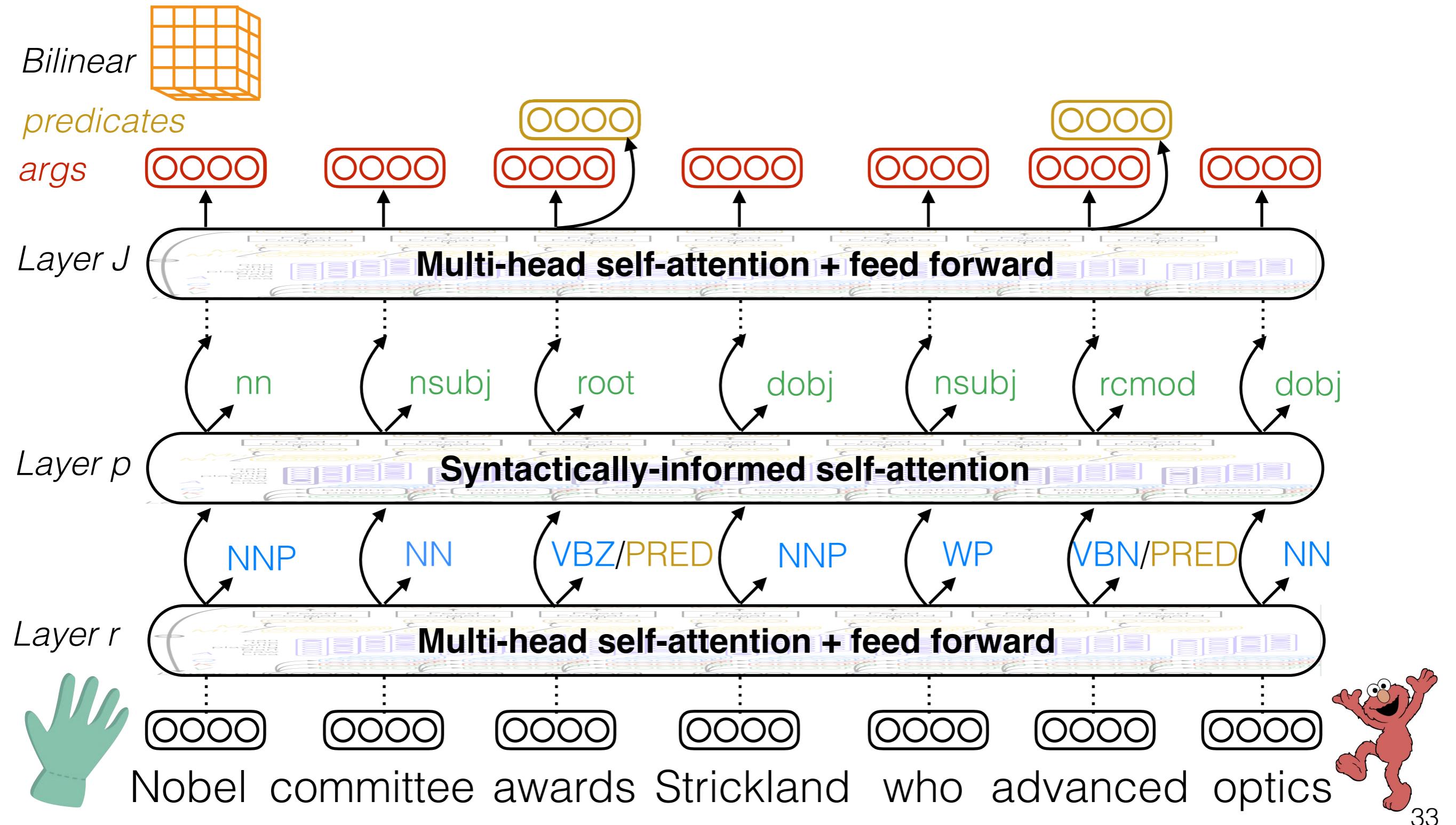
LISA: Linguistically-Informed Self-Attention



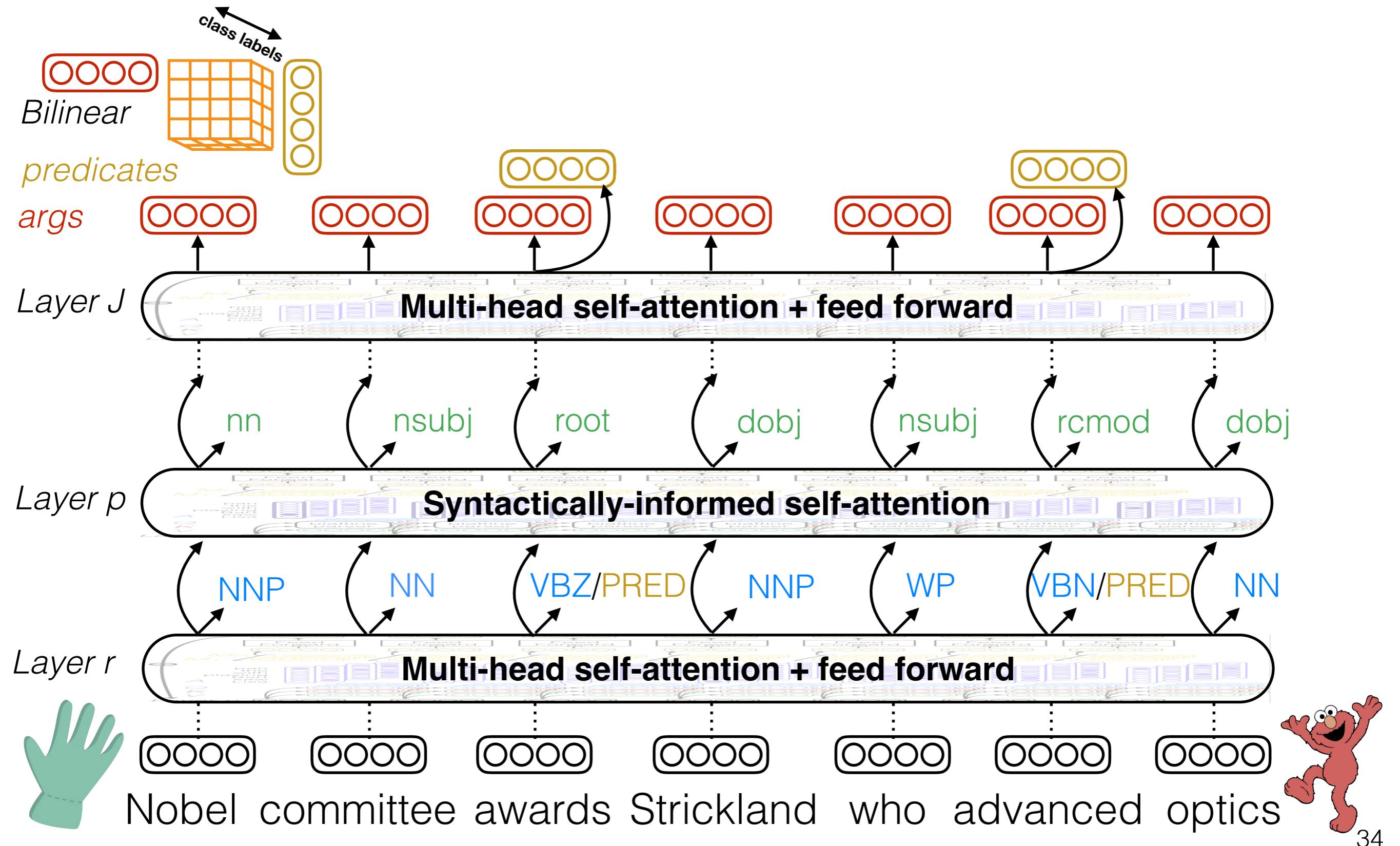
LISA: Linguistically-Informed Self-Attention



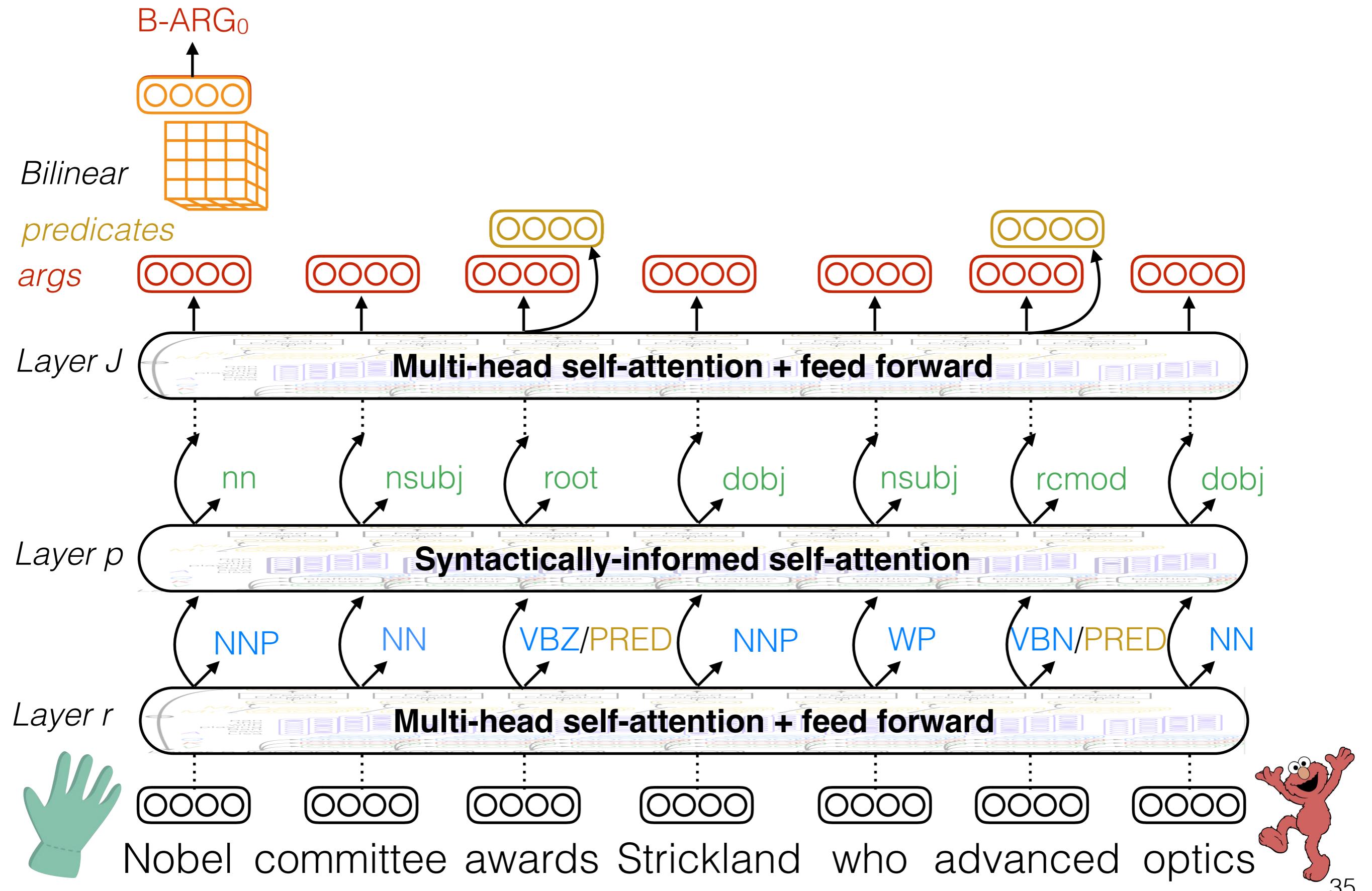
LISA: Linguistically-Informed Self-Attention



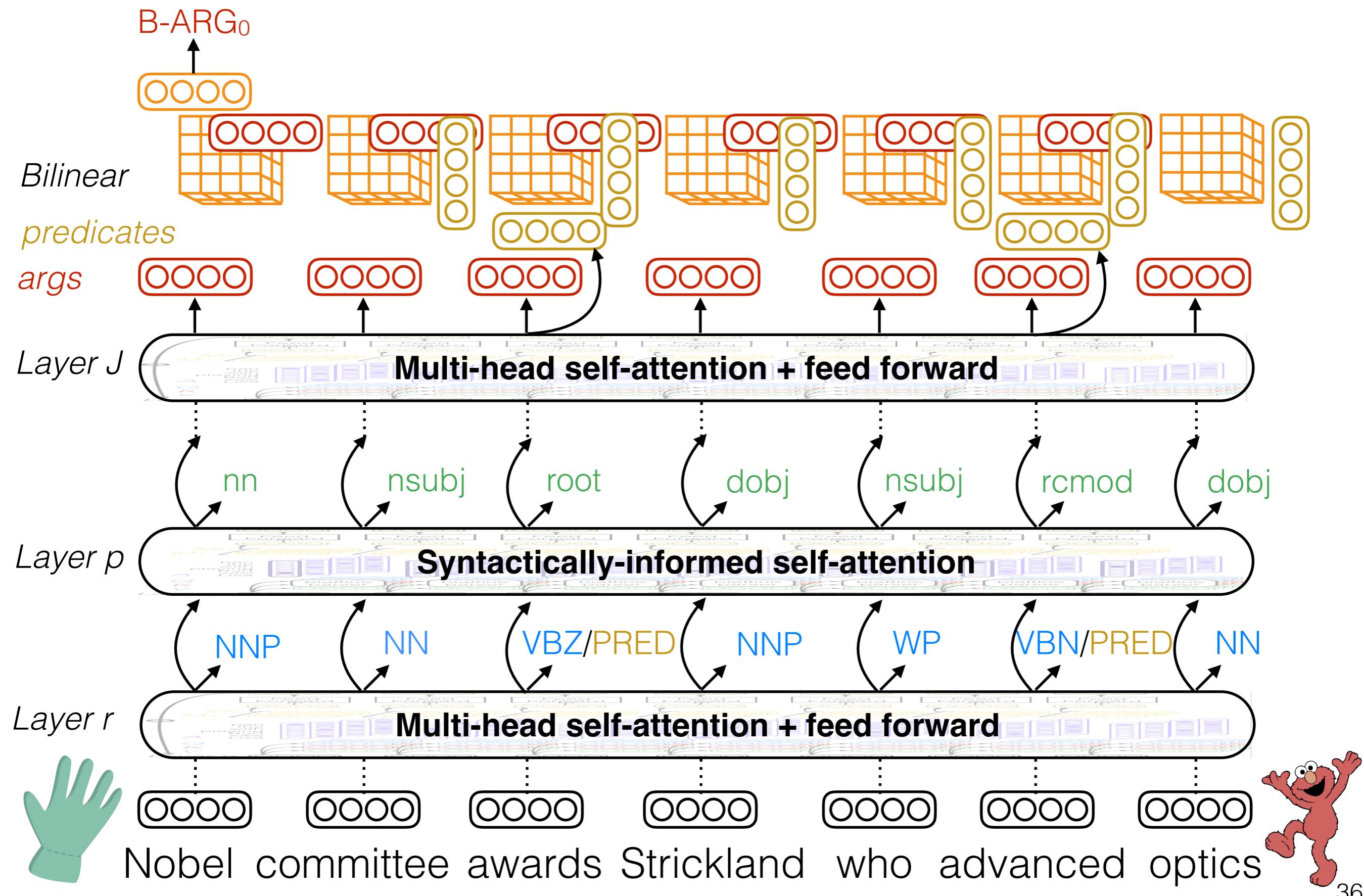
LISA: Linguistically-Informed Self-Attention



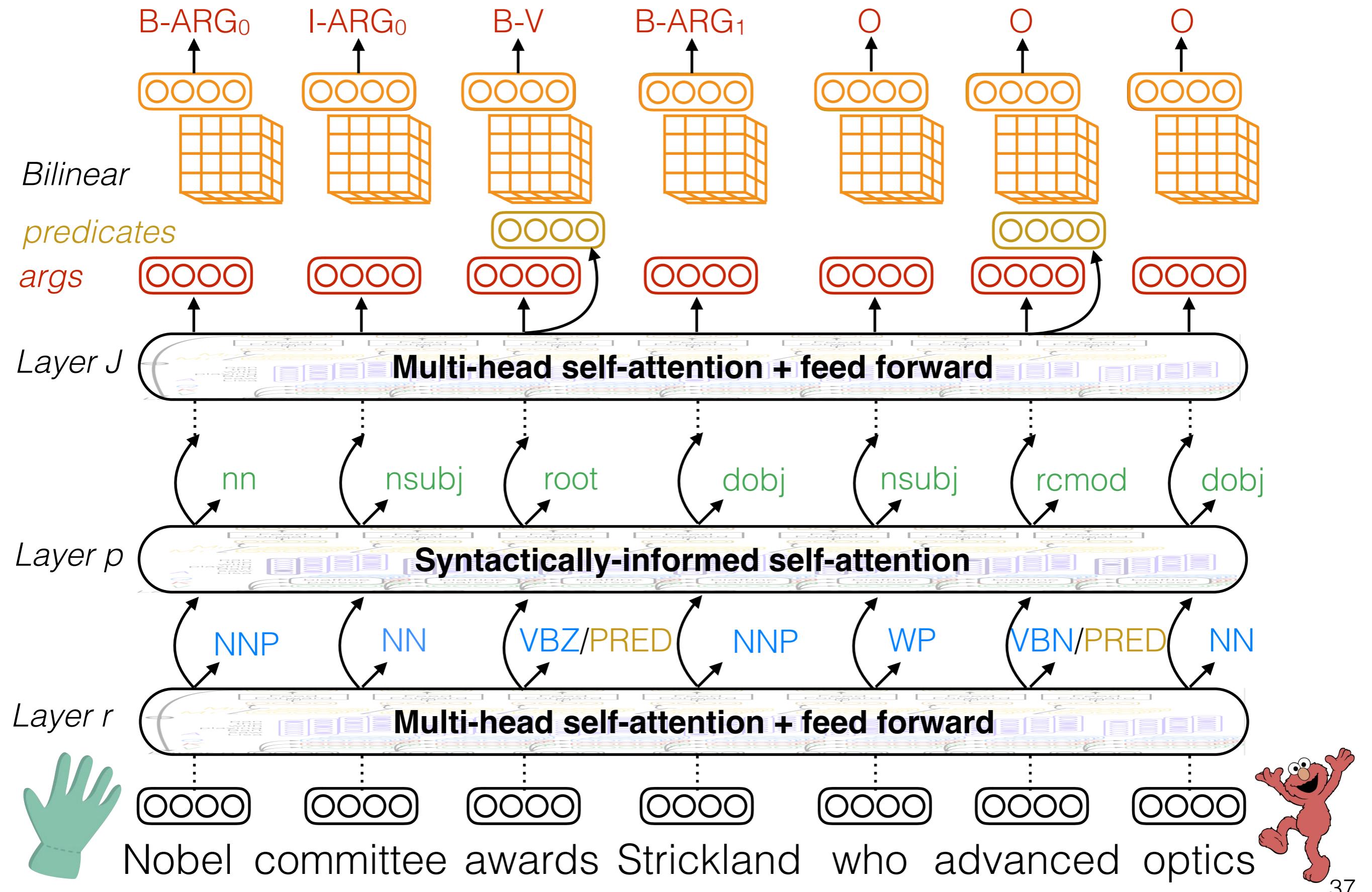
LISA: Linguistically-Informed Self-Attention



LISA: Linguistically-Informed Self-Attention

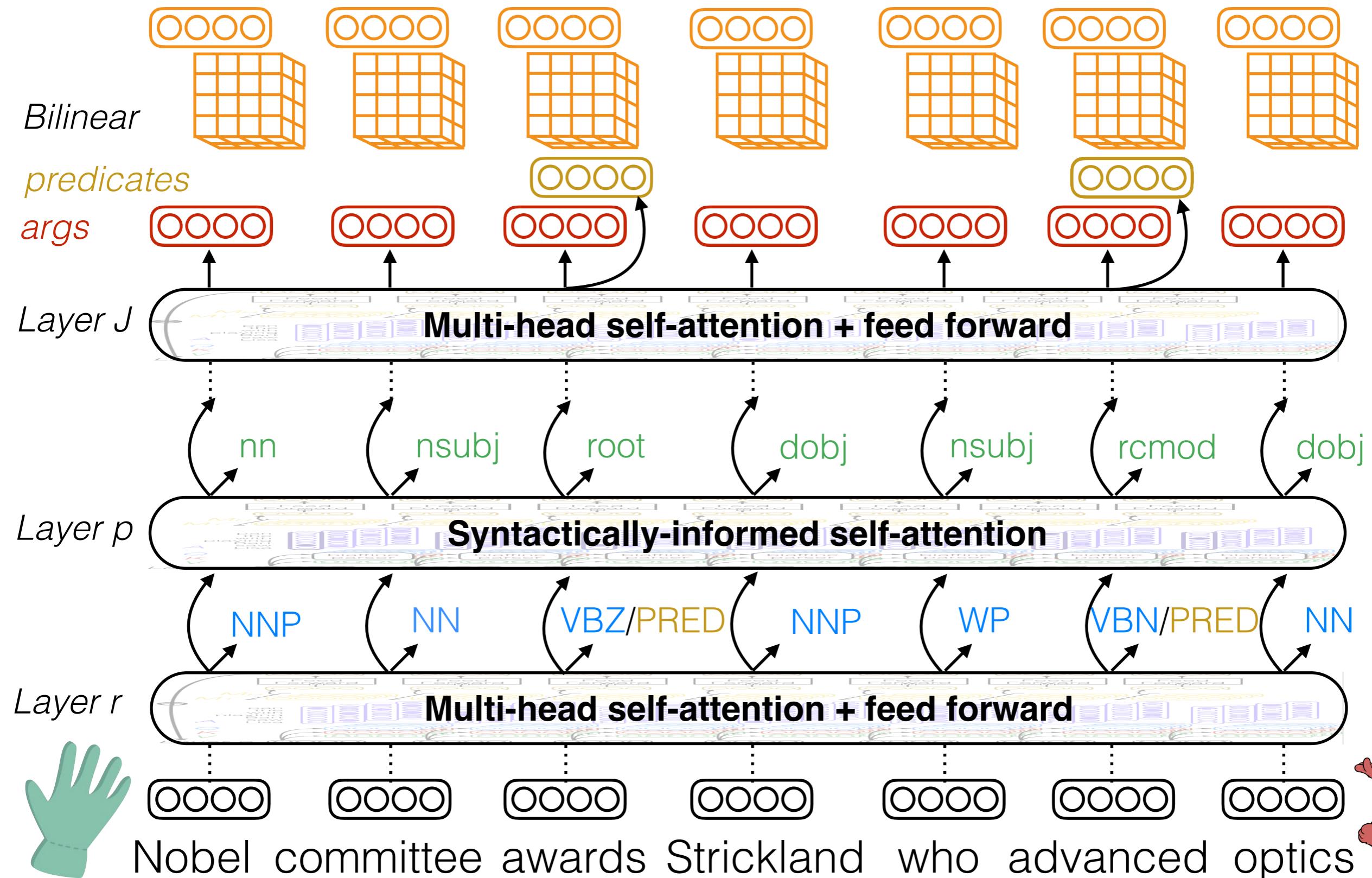


LISA: Linguistically-Informed Self-Attention



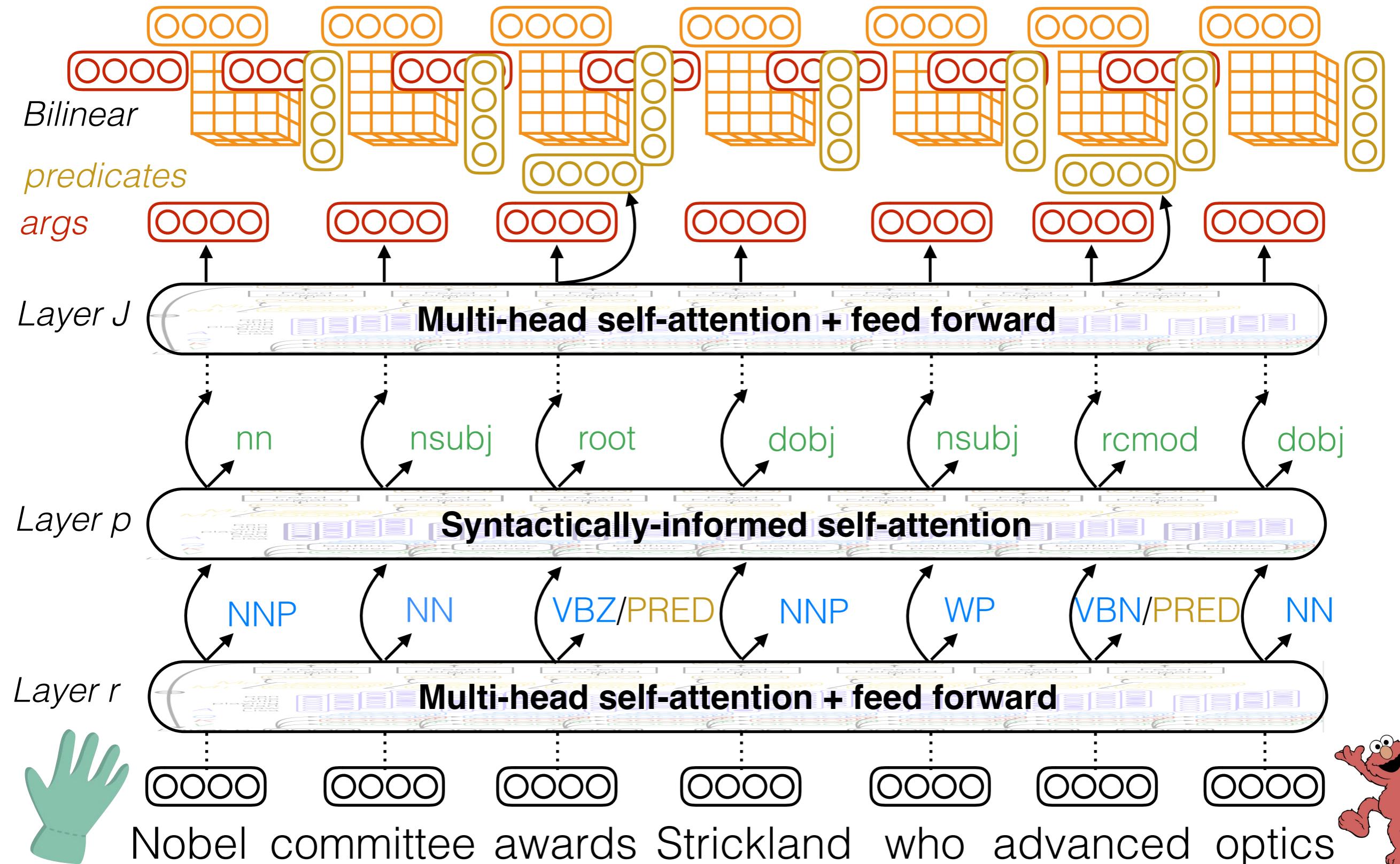
LISA: Linguistically-Informed Self-Attention

B-ARG₀ I-ARG₀ B-V B-ARG₁ O O O



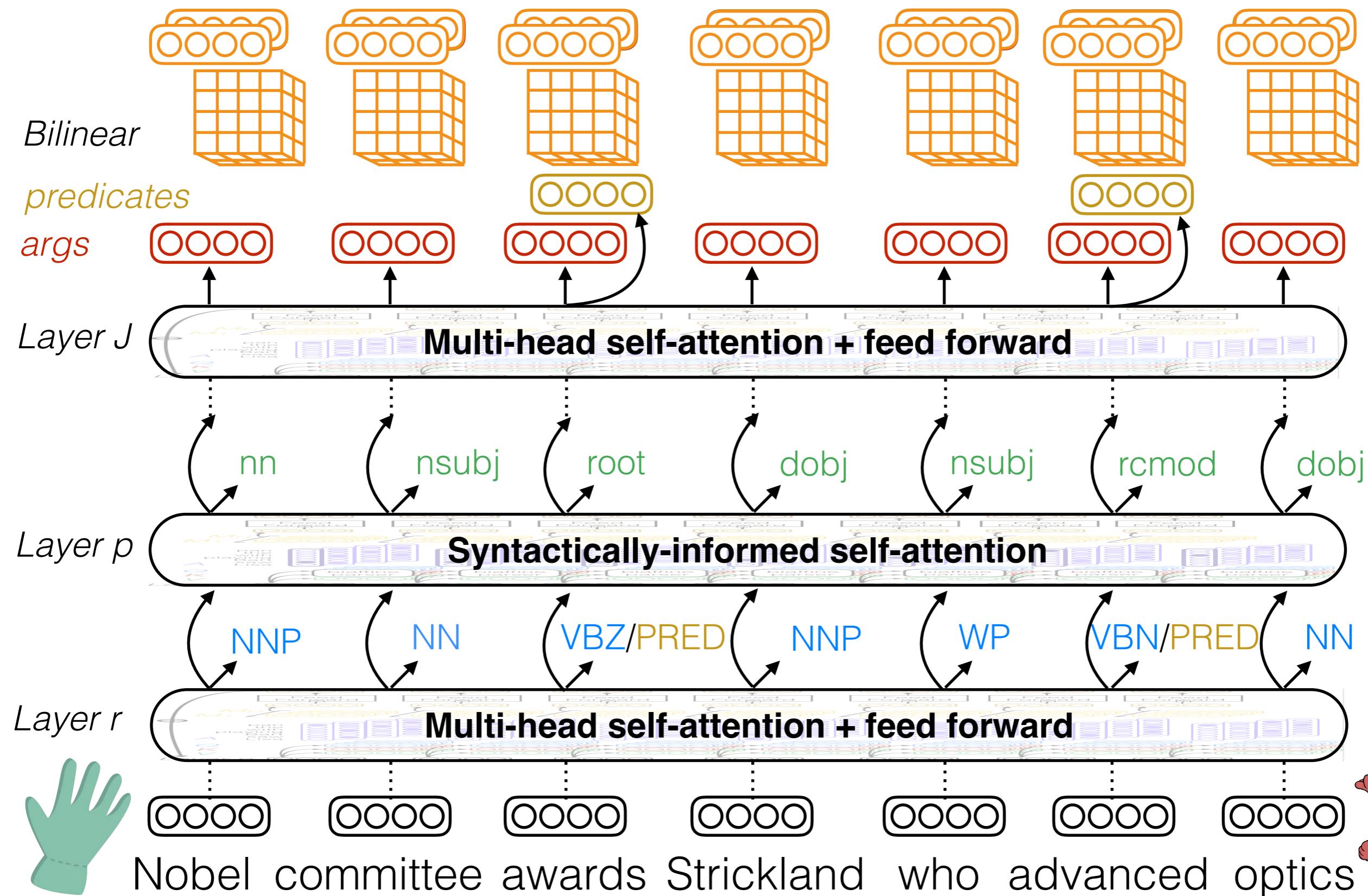
LISA: Linguistically-Informed Self-Attention

B-ARG₀ I-ARG₀ B-V B-ARG₁ O O O

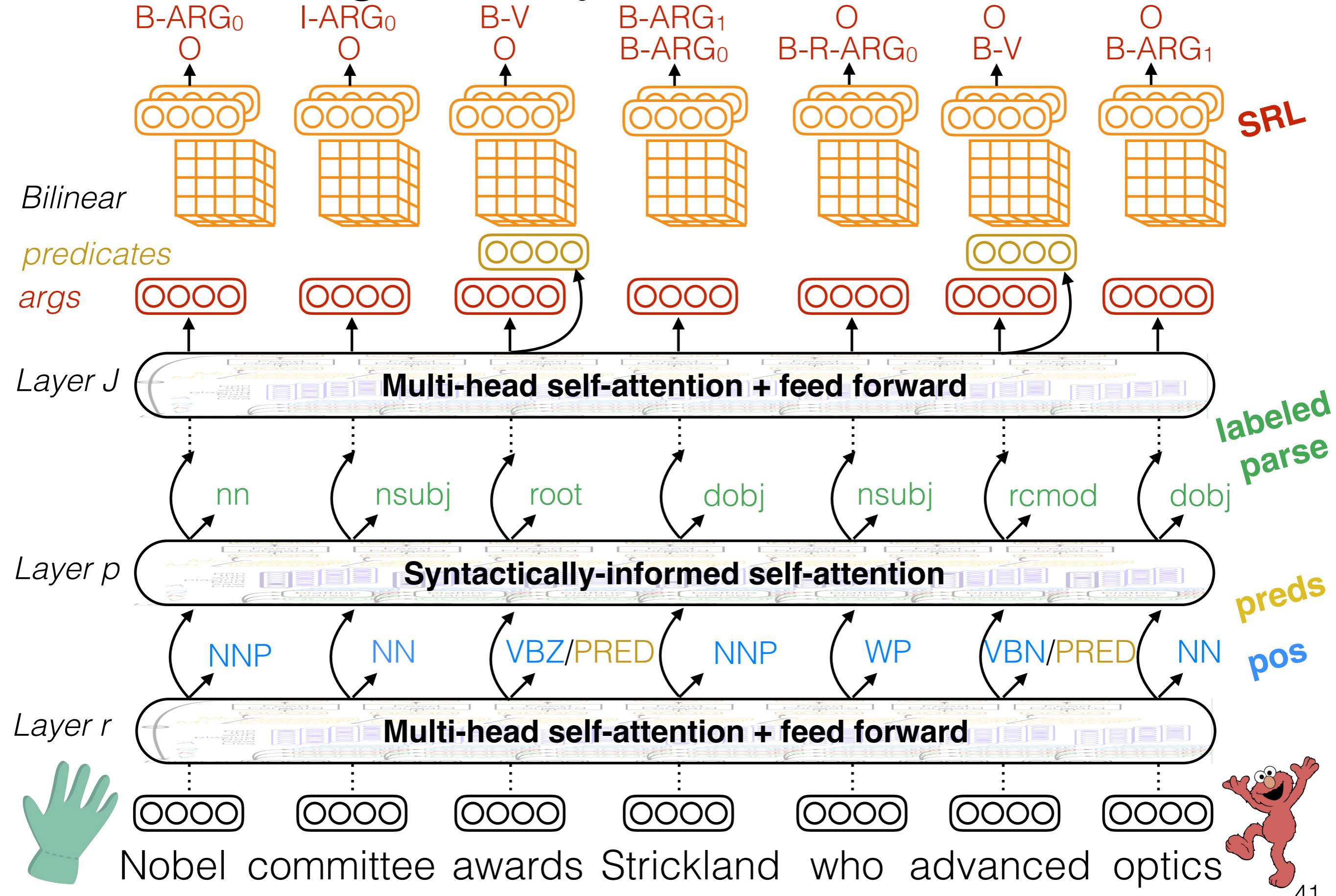


LISA: Linguistically-Informed Self-Attention

B-ARG₀ I-ARG₀ B-V B-ARG₁ O O O

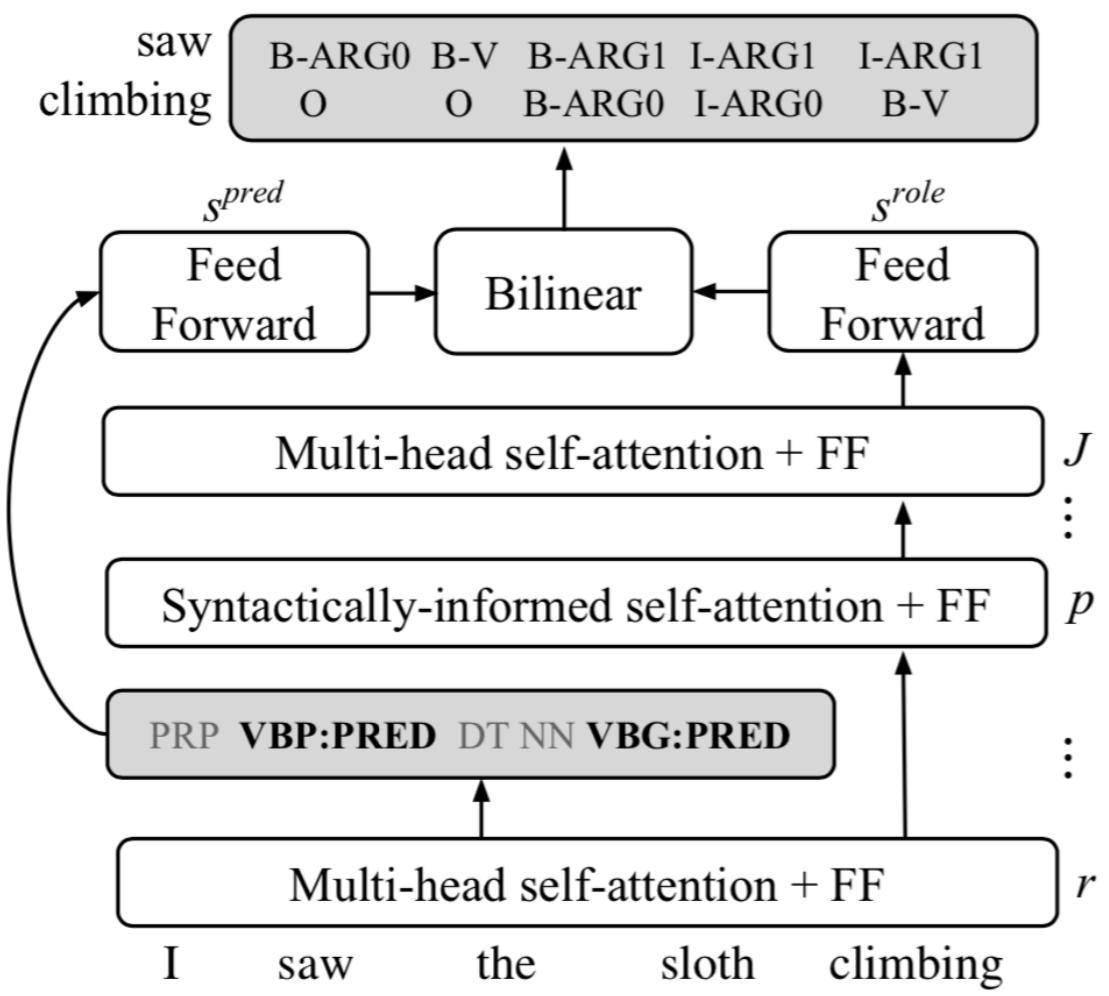


LISA: Linguistically-Informed Self-Attention



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
- Training
- Experimental results & error analysis



$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \left[\sum_{f=1}^F \log P(y_{ft}^{role} | \mathcal{P}_G, \mathcal{V}_G, \mathcal{X}) \right. \\ & \quad + \log P(y_t^{prp} | \mathcal{X}) \\ & \quad + \lambda_1 \log P(\text{head}(t) | \mathcal{X}) \\ & \quad \left. + \lambda_2 \log P(y_t^{dep} | \mathcal{P}_G, \mathcal{X}) \right] \end{aligned}$$

y_{ft}^{role} : SRL Label

\mathcal{P}_G : Gold Parse

$\text{head}(t)$, y_t^{dep} : Dependency Parsing Annotation
(Edge Existence, Arc-label)

\mathcal{V}_G : Gold Predicates

X : Input Sequence

y_t^{prp} : POS+Predicate Joint Label

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
- Training

- 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

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	94.9 JAS 84.99	90.3 JAS 74.66	96.6 JAS 86.90	93.4 JAS 88.25

+2.49 F1 +3.86 F1 +0.9 F1 ?

Experimental results: CoNLL-2005



96.5 UAS!

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: POS&Parsing

Data	Model	POS	UAS	LAS
WSJ	D&M _E	—	96.48	94.40
	LISA _G	96.92	94.92	91.87
	LISA _E	97.80	96.28	93.65
Brown	D&M _E	—	92.56	88.52
	LISA _G	94.26	90.31	85.82
	LISA _E	95.77	93.36	88.75
CoNLL-12	D&M _E	—	94.99	92.59
	LISA _G	96.81	93.35	90.42
	LISA _E	98.11	94.84	92.23

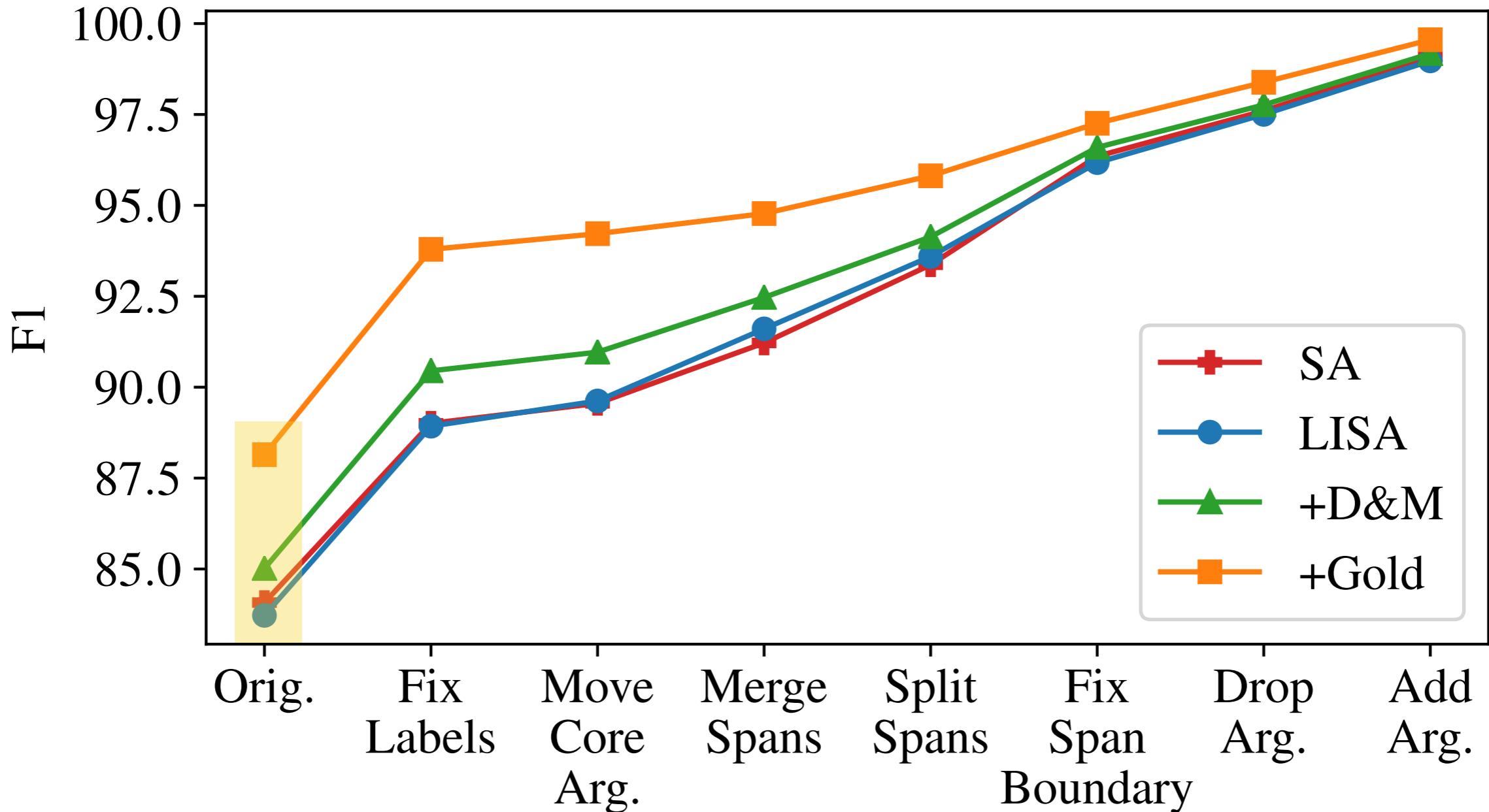
Table 4: Parsing (labeled and unlabeled attachment) and POS accuracies attained by the models used in SRL experiments on test datasets. Subscript *G* denotes GloVe and *E* ELMo embeddings.

Experimental results: Predicate Prediction

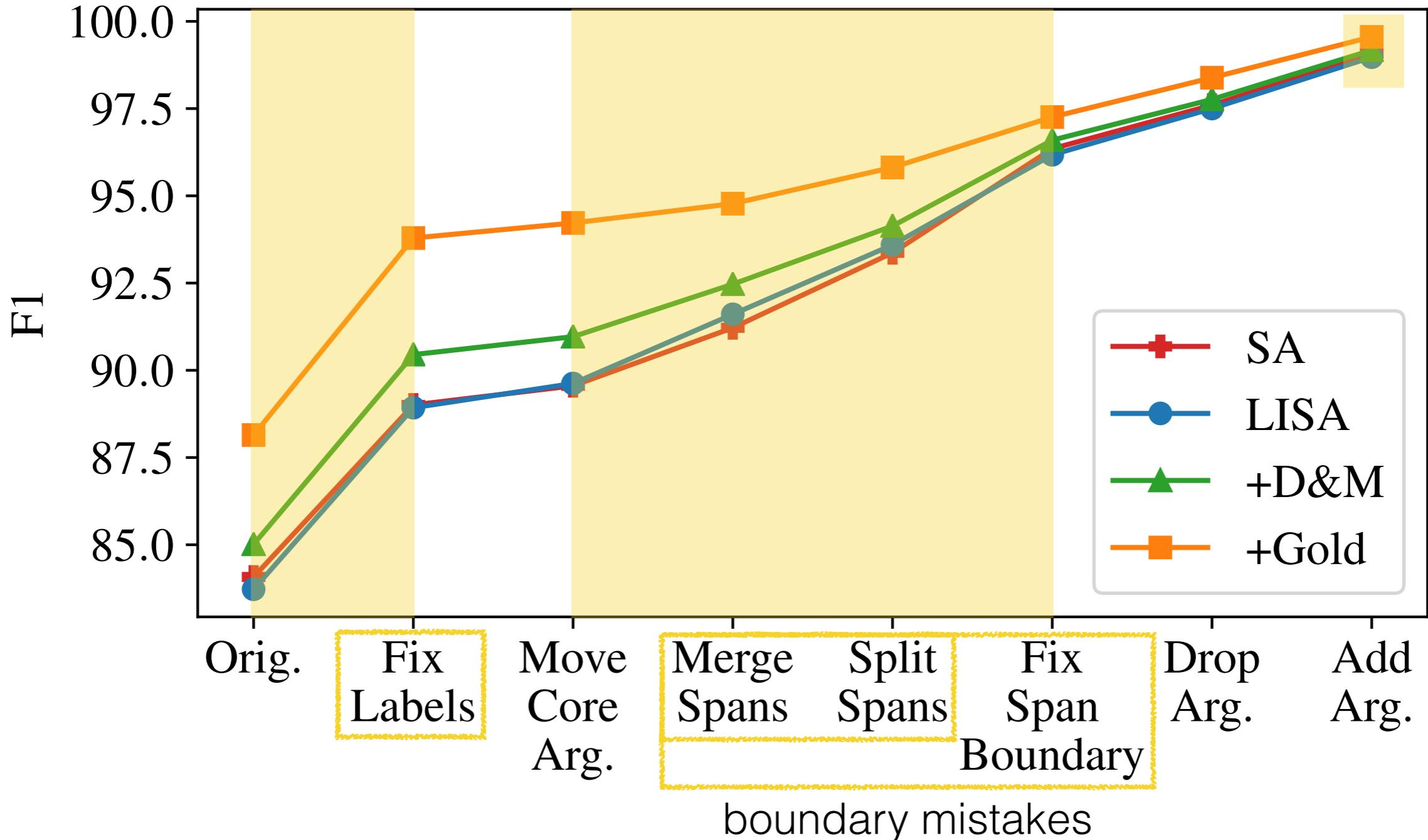
	Model	P	R	F1
WSJ	He et al. (2017)	94.5	98.5	96.4
	LISA	98.9	97.9	98.4
Brown	He et al. (2017)	89.3	95.7	92.4
	LISA	95.5	91.9	93.7
CoNLL-12	LISA	99.8	94.7	97.2

Table 5: Predicate detection precision, recall and F1 on CoNLL-2005 and CoNLL-2012 test sets.

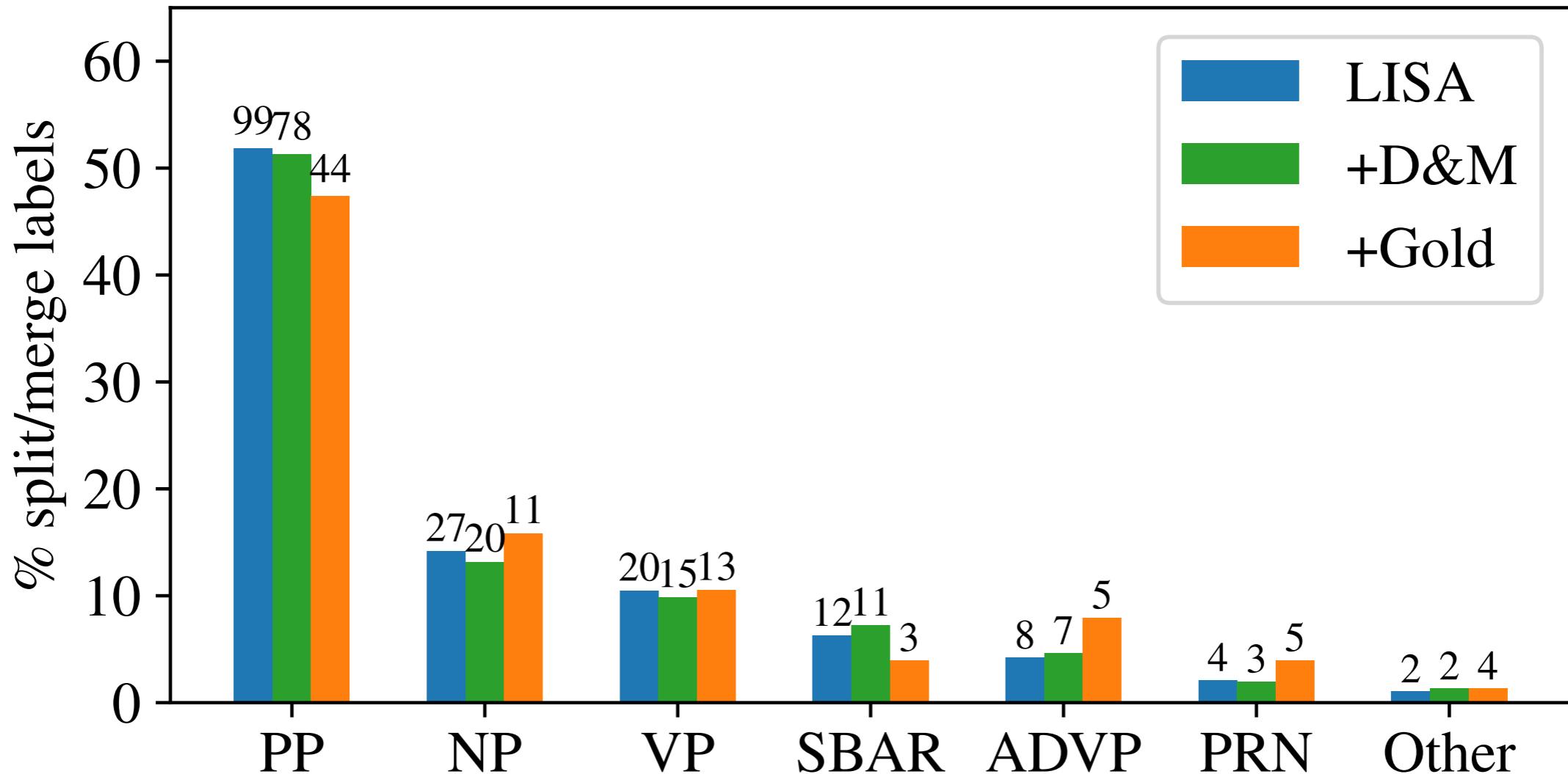
Experimental results: Analysis



Experimental results: Analysis



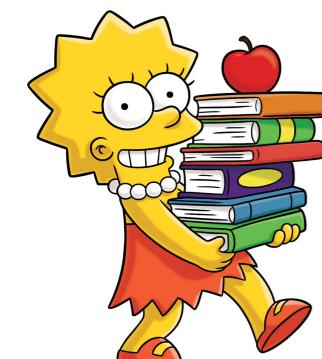
Experimental results: Analysis



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>