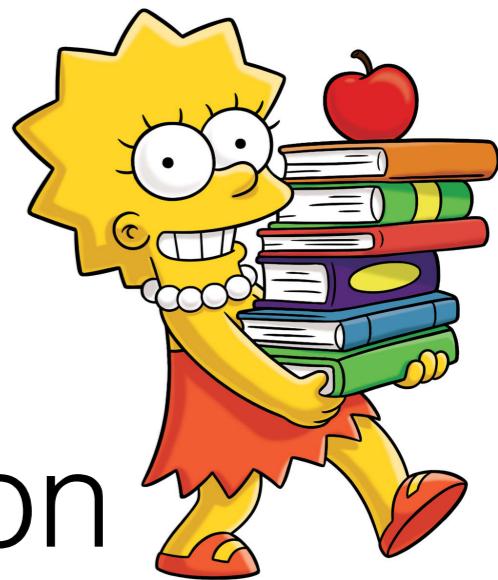


# LISA

## Linguistically-Informed Self-Attention for Semantic Role Labeling



Emma  
Strubell<sup>1</sup>



Patrick  
Verga<sup>1</sup>



Daniel  
Andor<sup>2</sup>



David  
Weiss<sup>2</sup>



Andrew  
McCallum<sup>1</sup>

<sup>1</sup> UMassAmherst



College of Information  
and Computer Sciences

<sup>2</sup>



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 first female recipient of a Nobel Prize in Physics was Marie Curie in 1903. Since then, only three other women have won the award: Venetia Burney in 1914, Maria Goeppert-Mayer in 1963, and Donna Strickland in 2018.

Strickland, 59, is the third woman to win the Nobel Prize in Physics since it was established in 1901. She is also the youngest person to receive the award in 2018.

**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 1985; accepted 10 September 1985

We have demonstrated the compression of chirped optical pulses with a peak power of 1.06 μm laser pulses.

The onset of self-focusing limits the amplification of short pulses. A similar problem arises for short, yet energetic pulses capable of handling the required compression. We have demonstrated the compression of chirped optical pulses before amplifying an echo is compressed to a negative dispersive pulse.

We wish to report the compression of chirped optical pulses and that in principle it is possible to compress short (<1 ps) pulses. A long pulse is deliberately stretched into a short, low-energy pulse. The pulse is linearly chirped by the combination of group velocity dispersion and frequency modulation [2]. The pulse is then compressed by self-focusing. Amplifying the stretched pulse allows the compressed pulse to be compressed before self-focusing does not appear to affect the compression. Pulses can be fully compressed without the need for amplifying a chirped pulse if the gain sweep rate is sufficiently slow.

\* This is a corrected version of an article published in Optics Comm., 55 (1985) 447. It was printed as fig. 1.

0 030-4018/85/\$03.00  
(North-Holland Physics)

**STEM Gems**  
@STEMGemsBook

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

**SECOND EDITION**

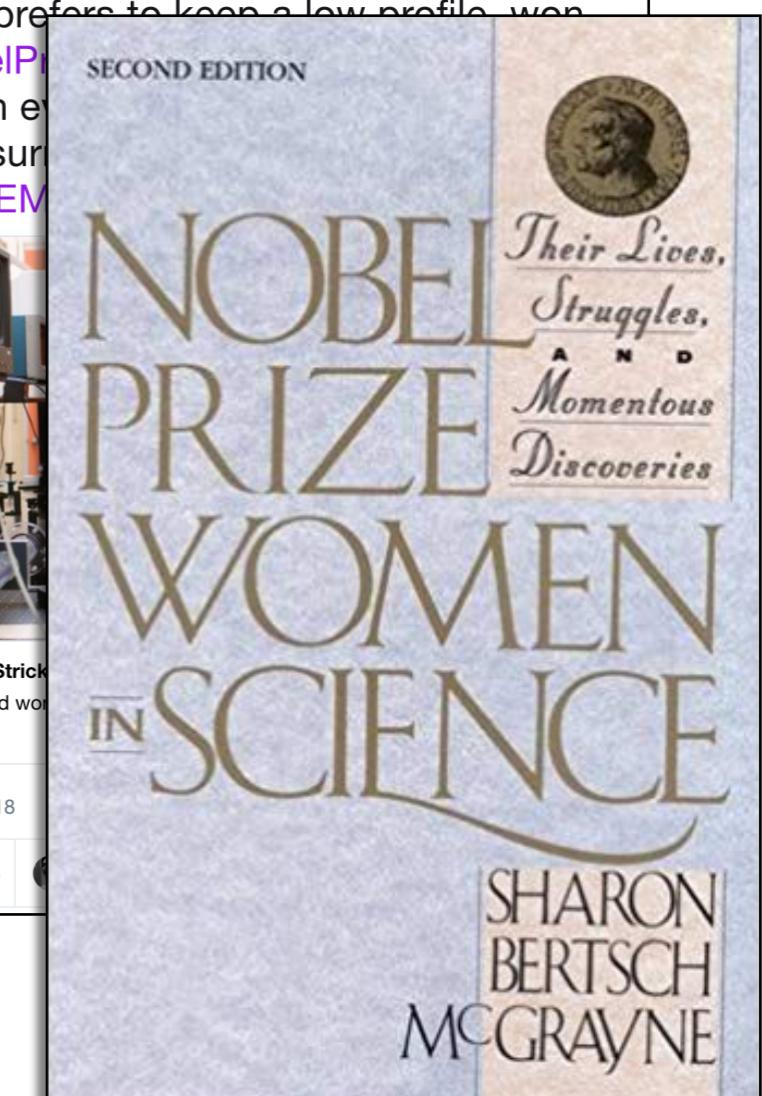
**NOBEL PRIZE WOMEN IN SCIENCE**

*Their Lives, Struggles, AND Momentous Discoveries*

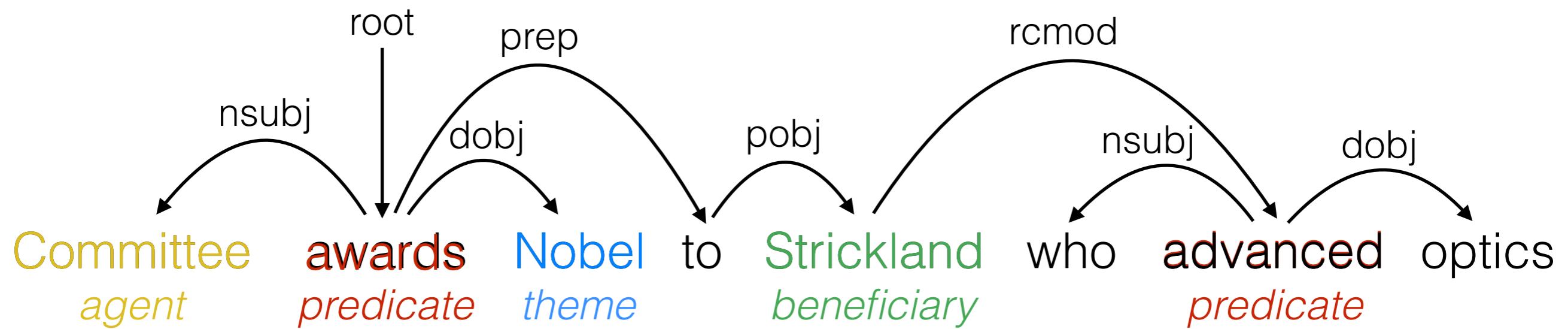
**Physicist Donna Strickland**  
She is only the third woman to win the Nobel Prize in Physics.  
[nobelprize.org](http://www.nobelprize.org)

6:02 AM - 23 Oct 2018

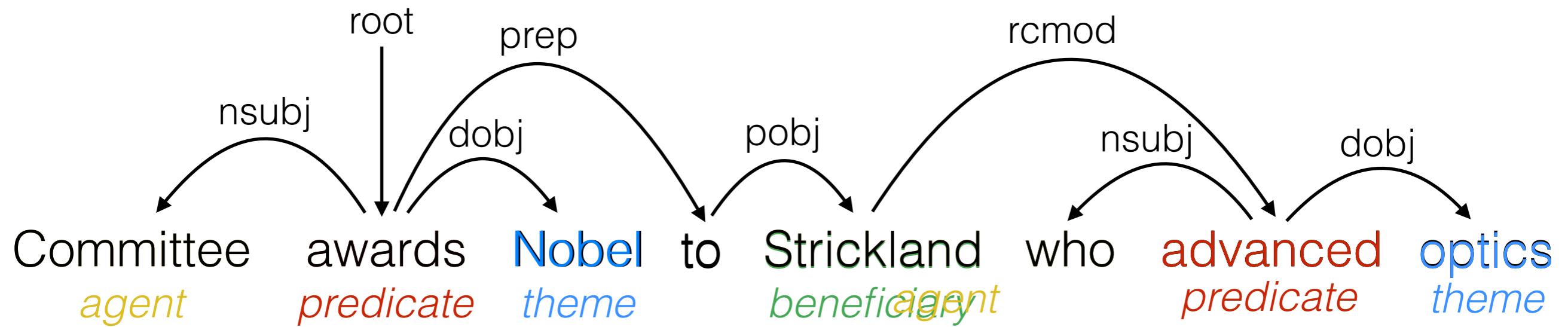
9 Retweets 30 Likes



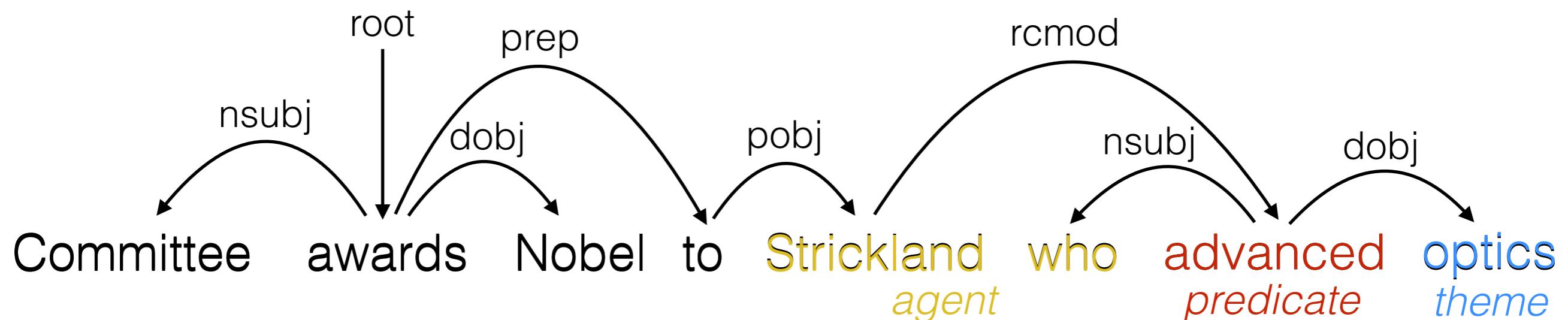
# SRL: Who did what to whom?



# SRL: Who did what to whom?

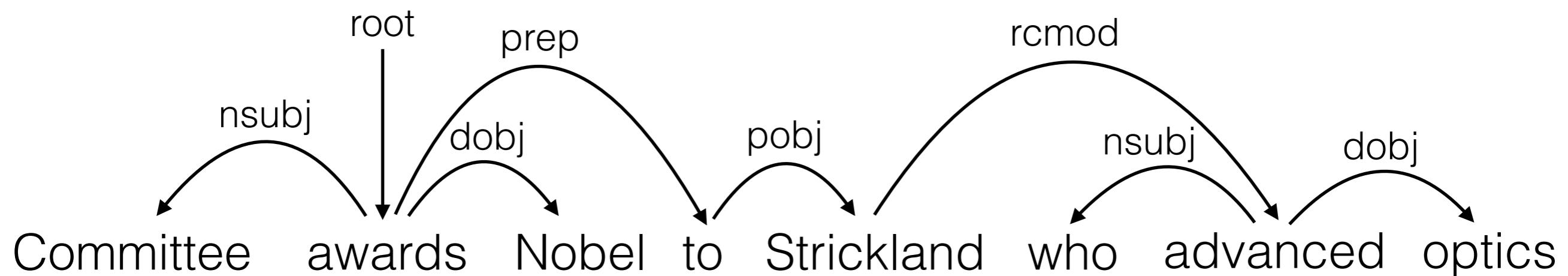


# SRL: Who did what to whom?



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

# PropBank SRL: Who did what to whom?



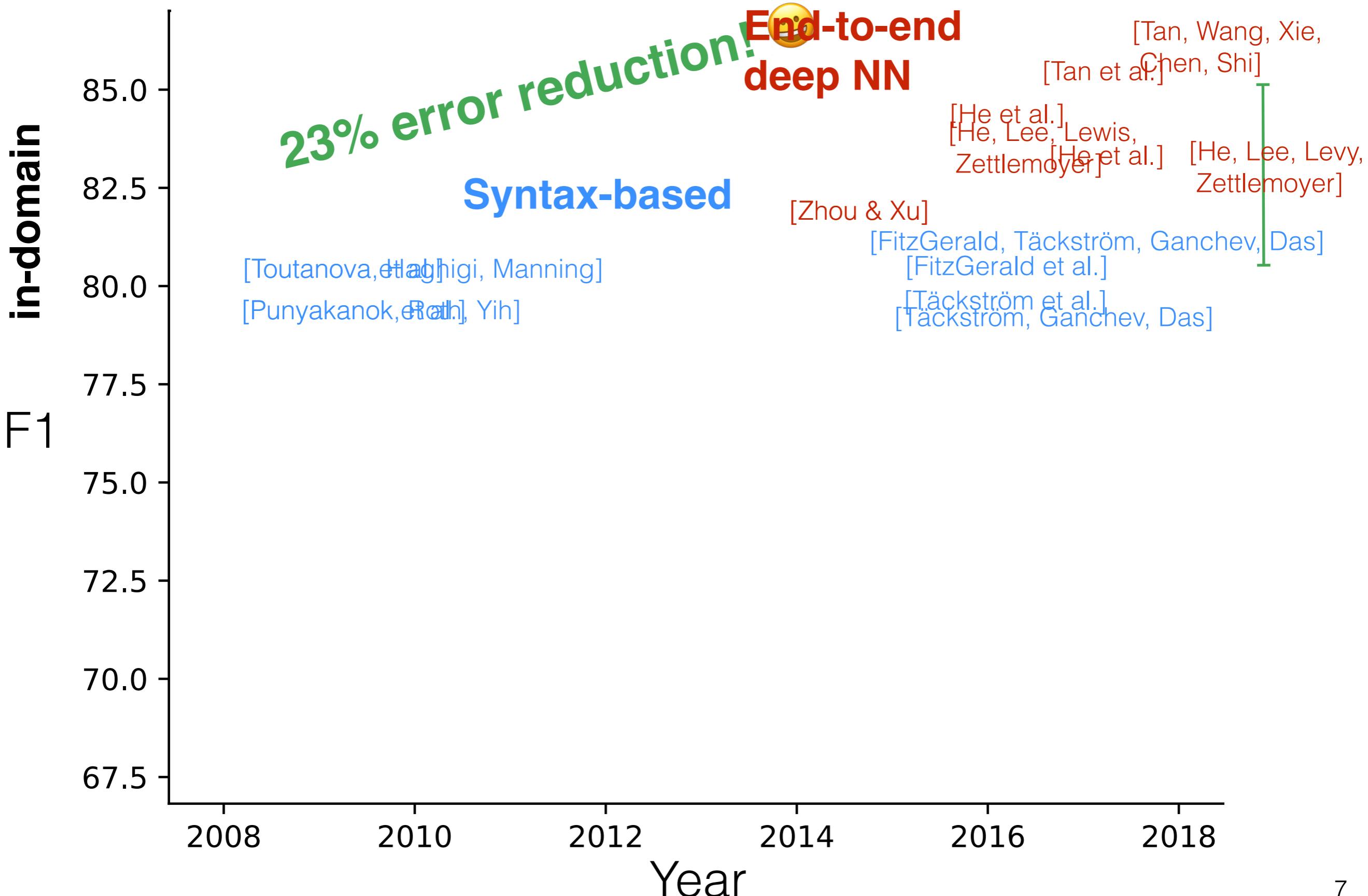
Committee awards Nobel to Strickland who advanced optics

ARG<sub>agent</sub> R-ARG<sub>0</sub> predicate

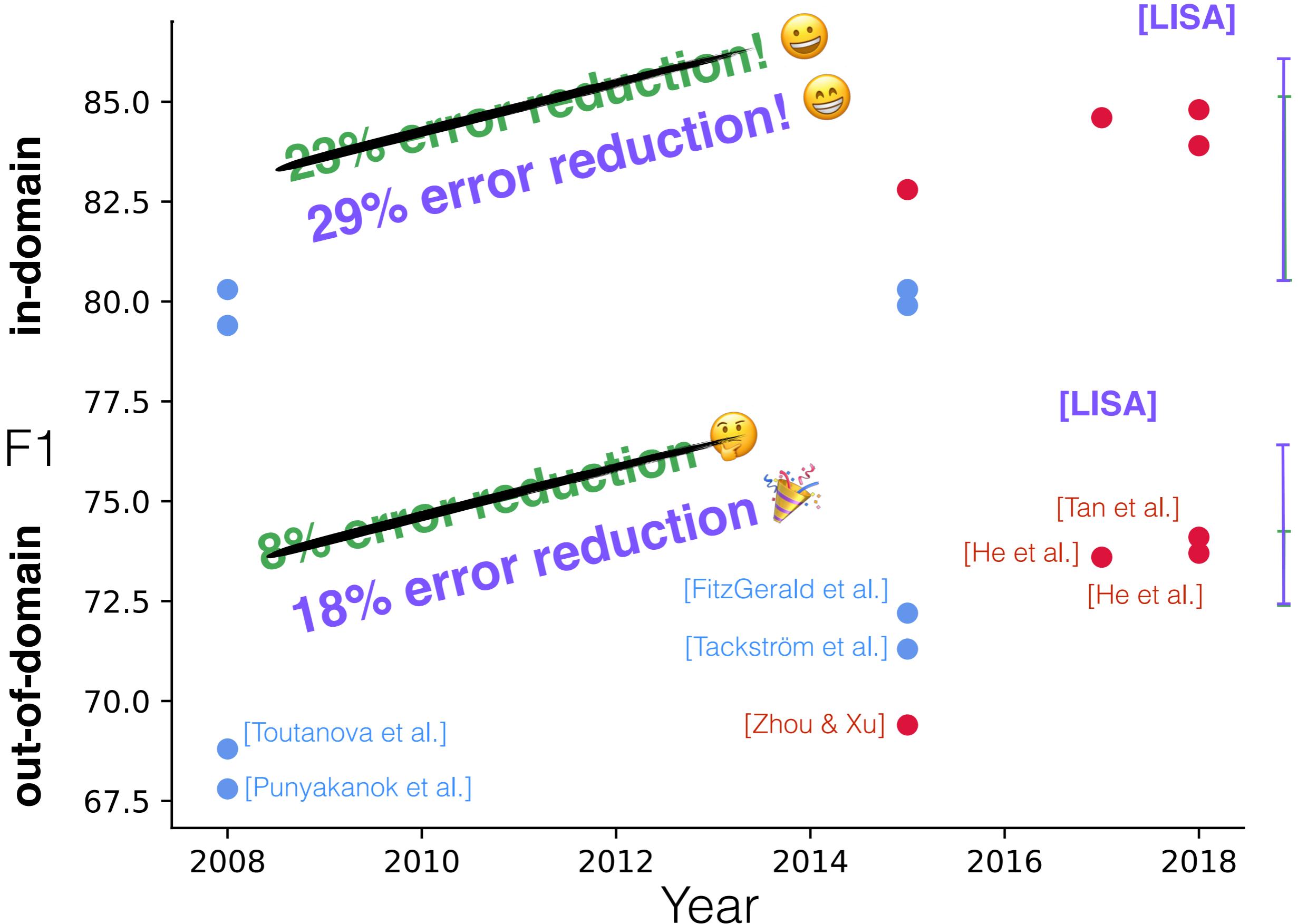
Committee awards Nobel to Strickland who advanced optics

Agent predicate Arg<sub>0</sub> Arg<sub>1</sub>

# 10 years of PropBank SRL

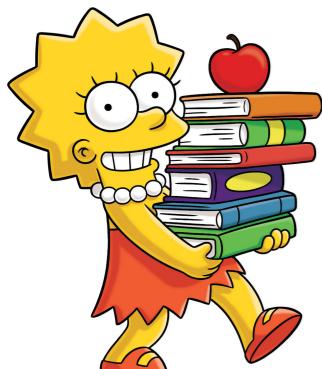


# 10 years of PropBank SRL



# Linguistically-Informed Self-Attention

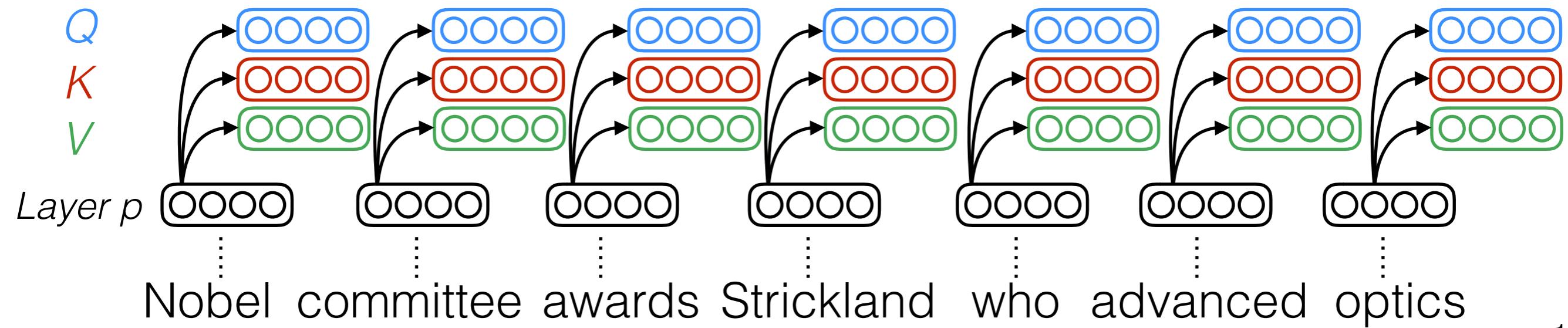
- **Multi-task learning**
    - Part-of-speech tagging
    - Labeled dependency parsing
    - Predicate detection
    - Semantic role spans & labeling
  - **Syntactically-informed self-attention**
    - Multi-head self-attention supervised by **syntax**
- 



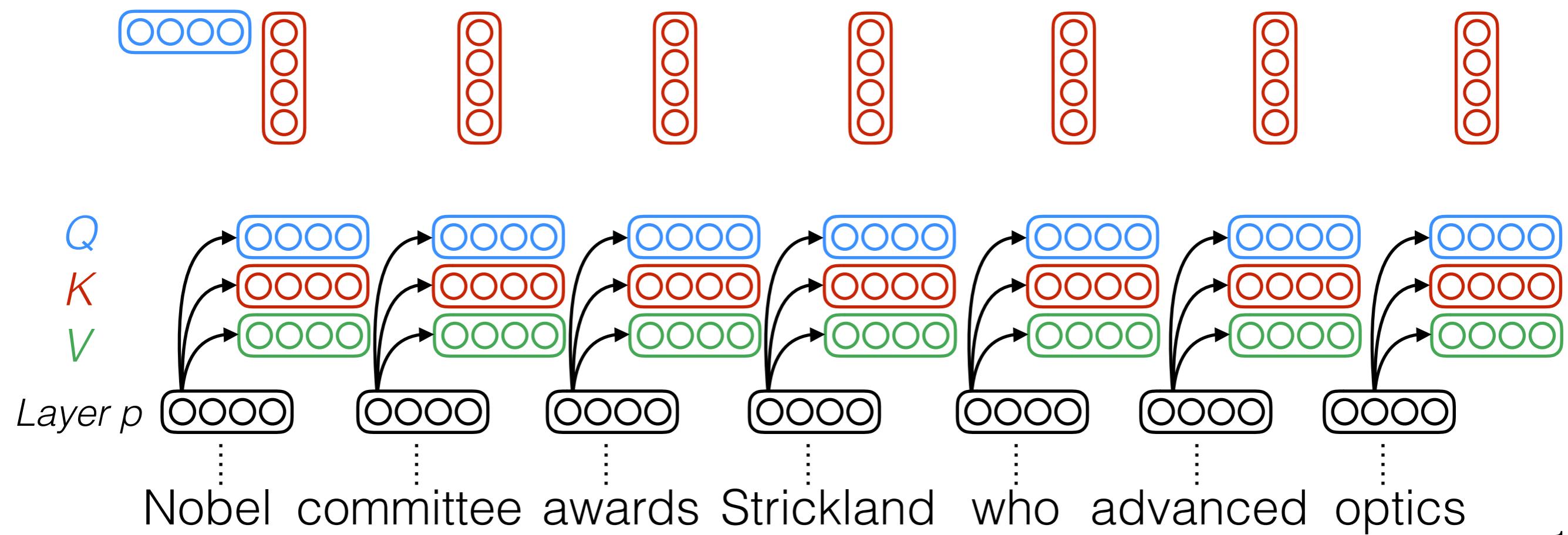
# Outline

- 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
  - Experimental results & error analysis

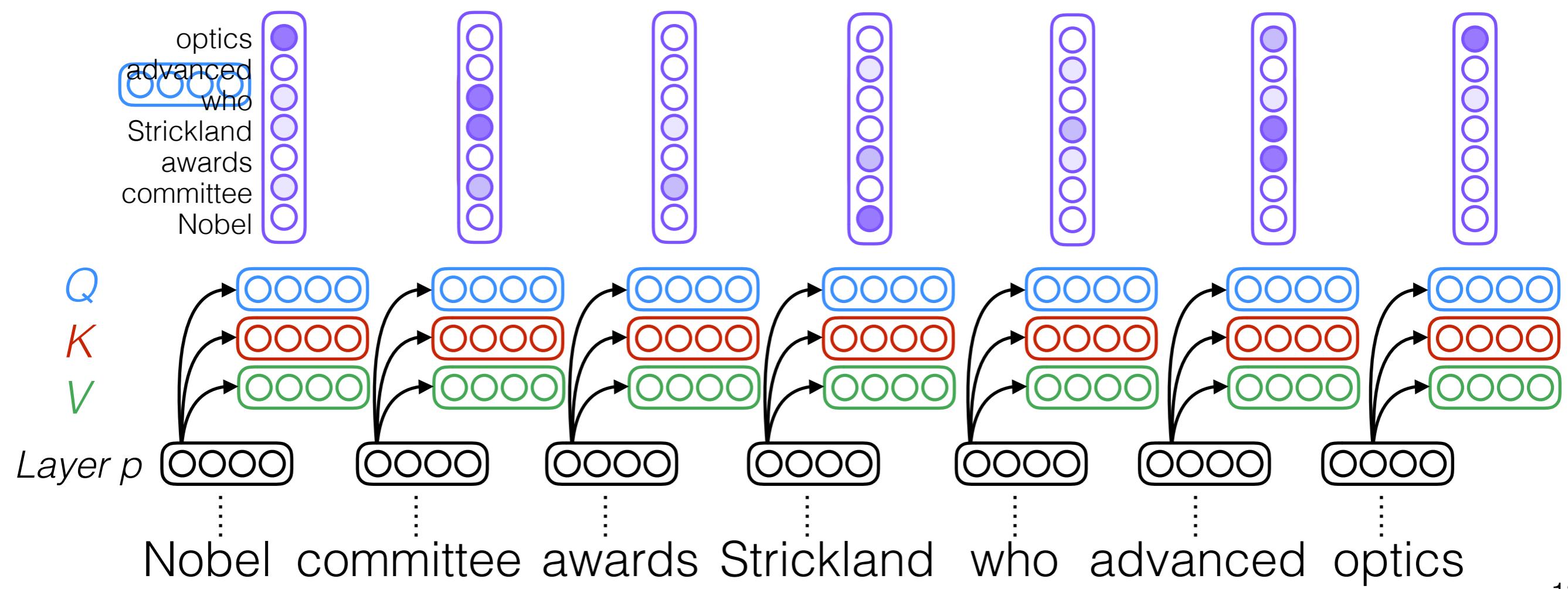
# Self-attention



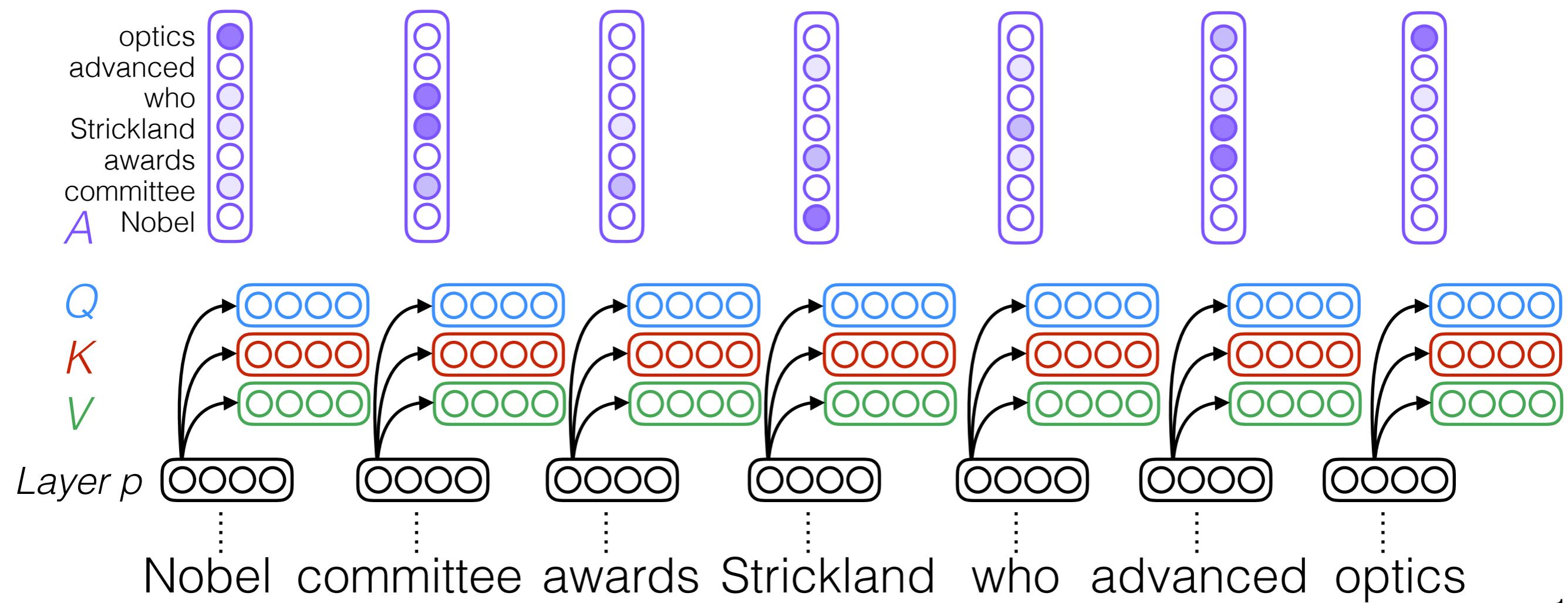
# Self-attention



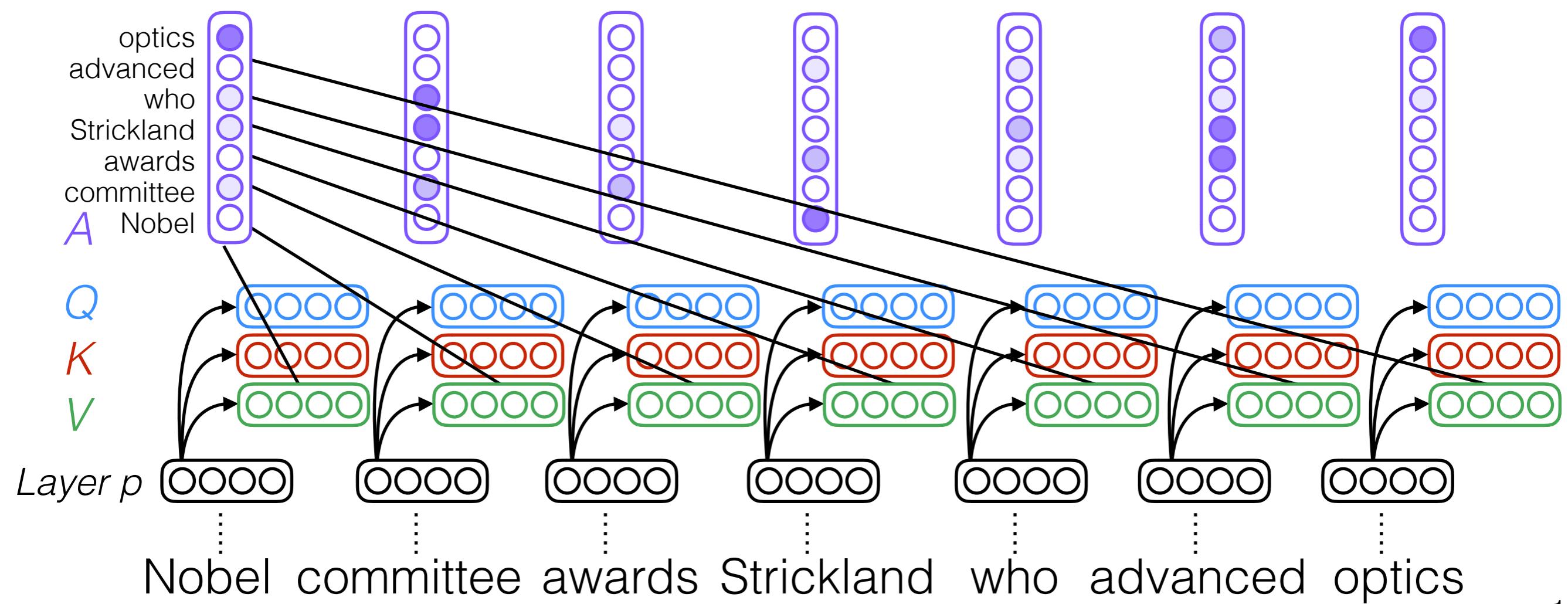
# Self-attention



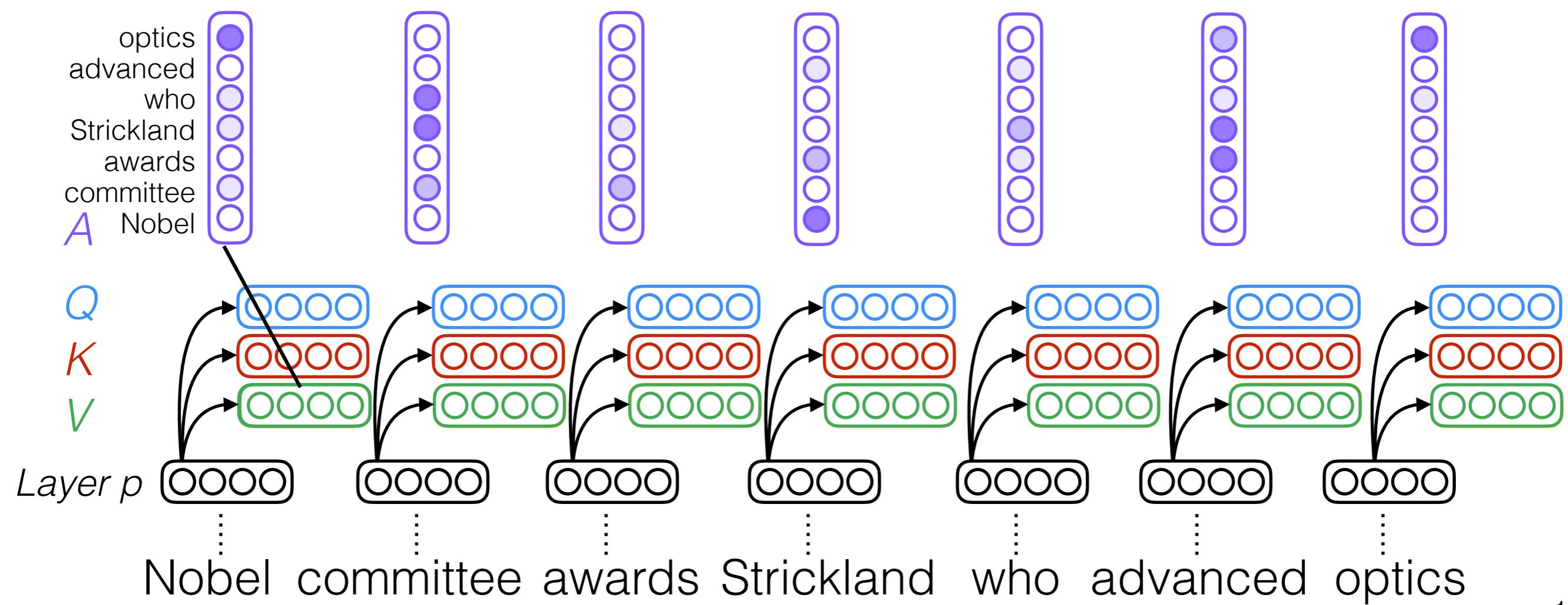
# Self-attention



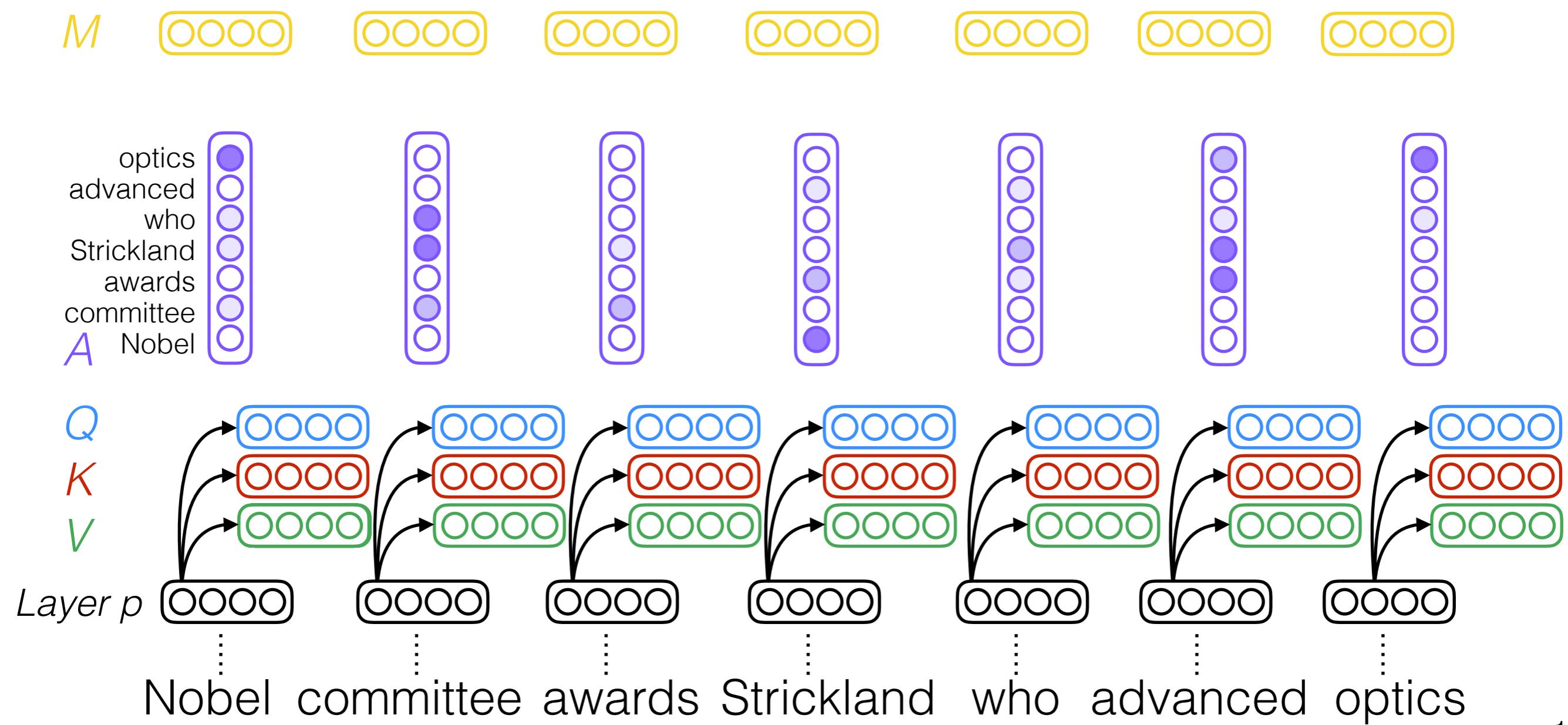
# Self-attention



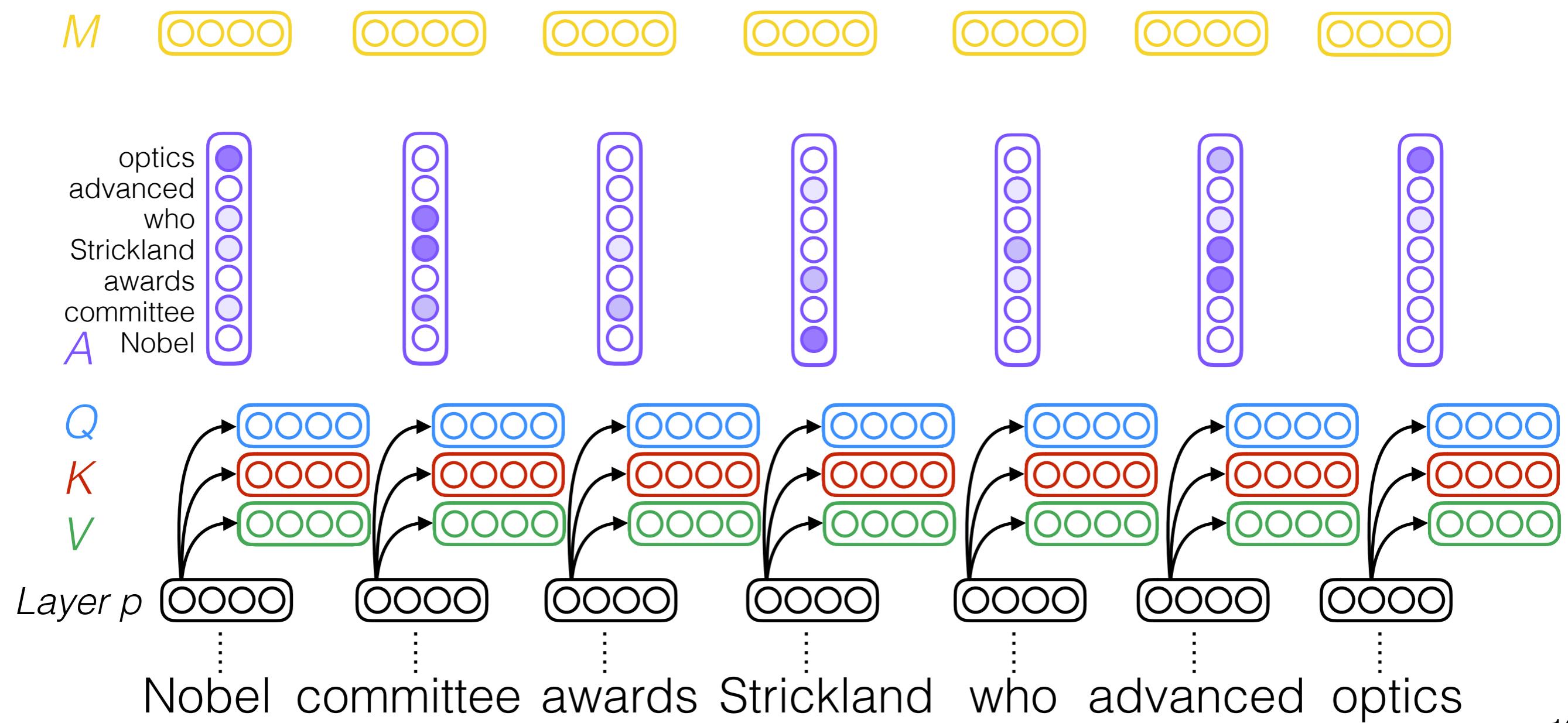
# Self-attention



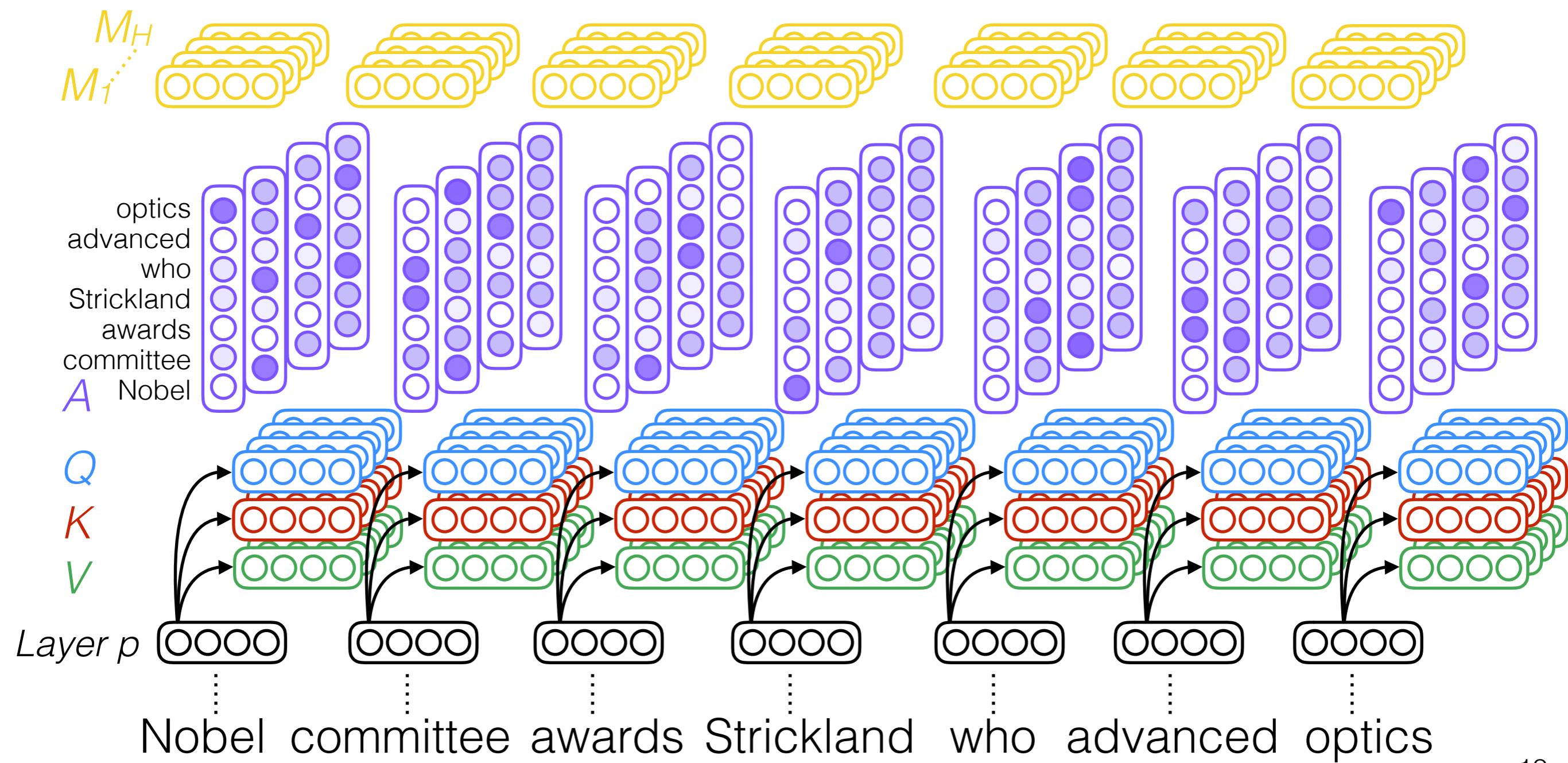
# Self-attention



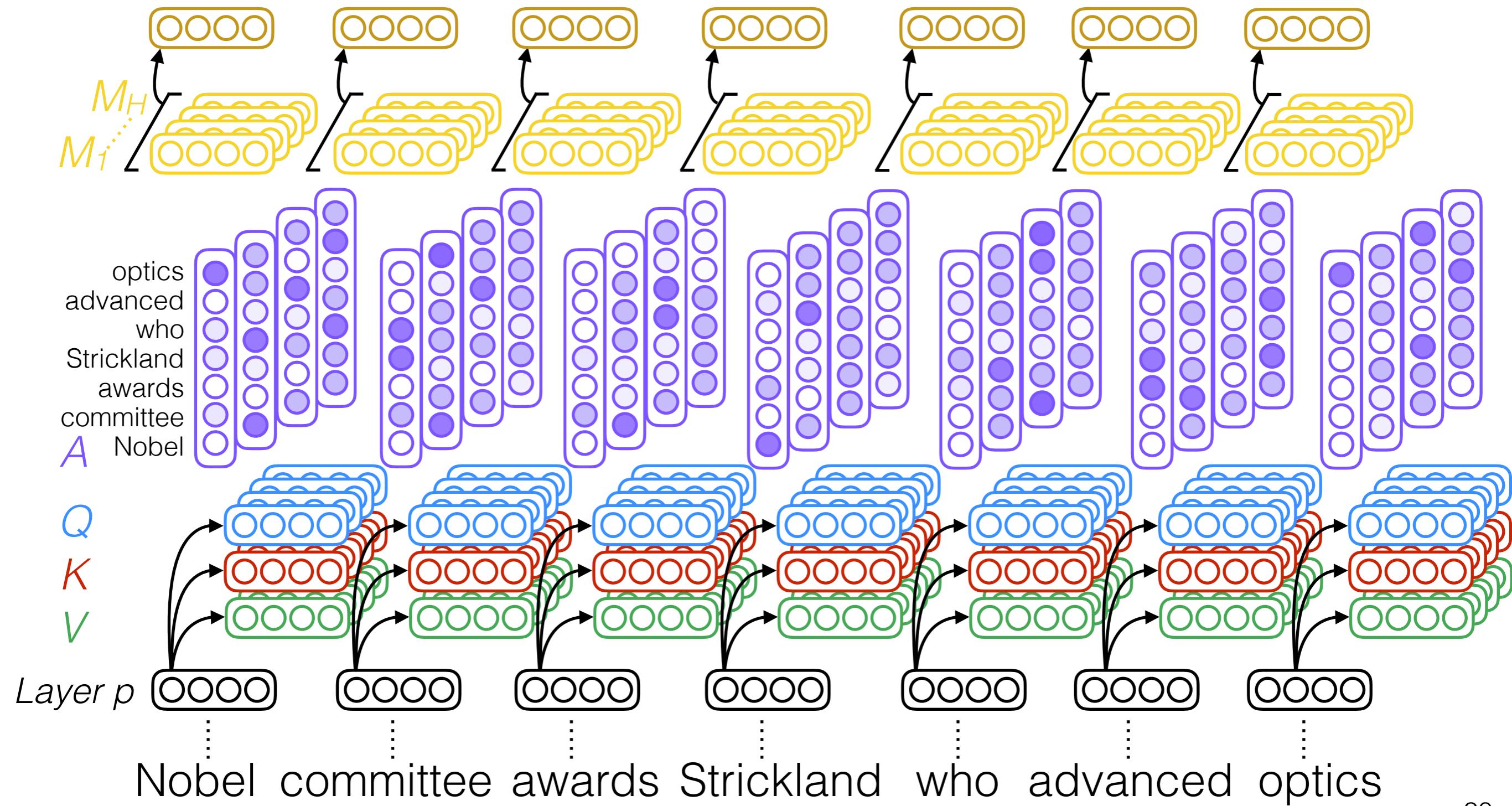
# Self-attention



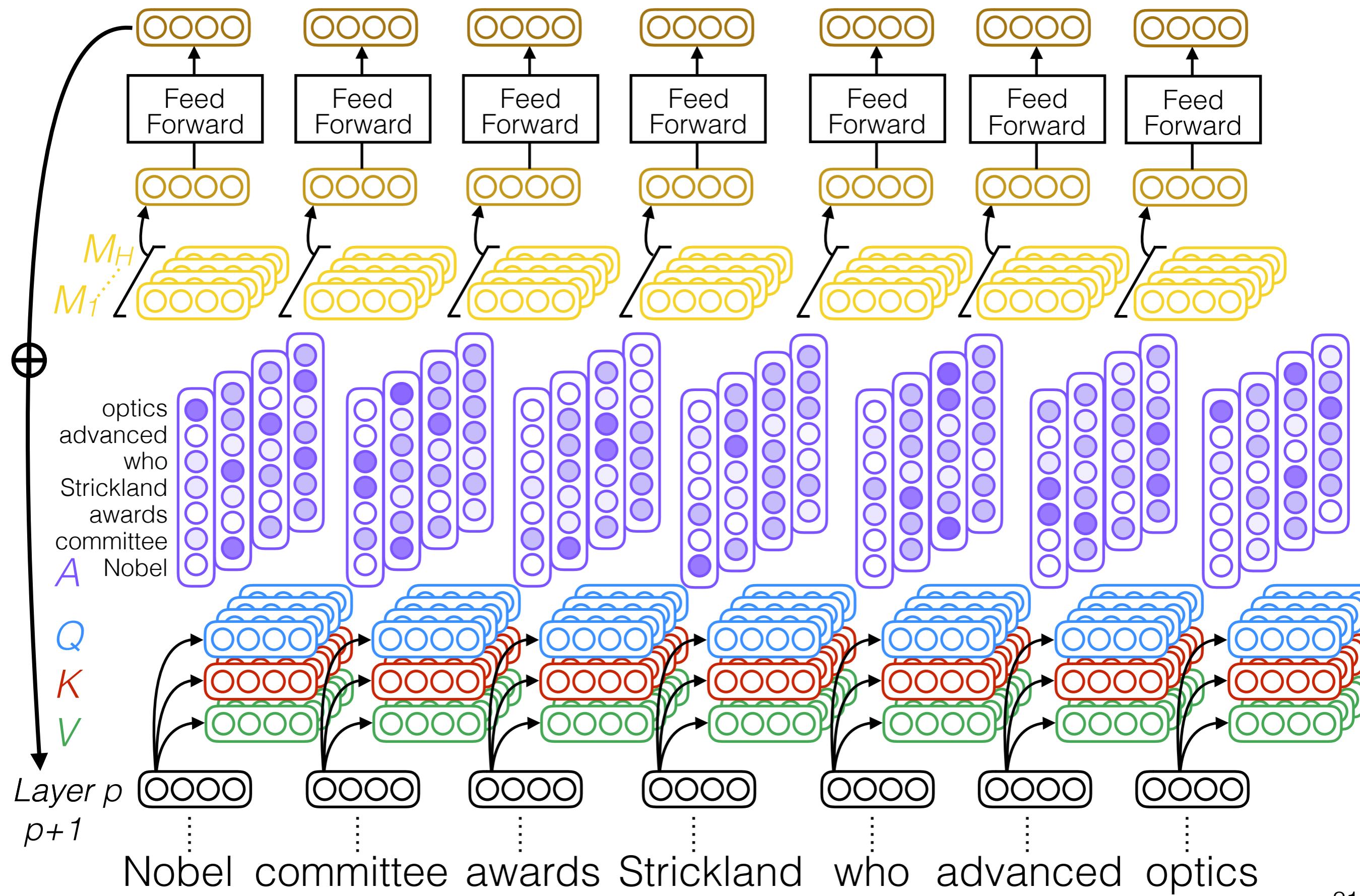
# Multi-head self-attention



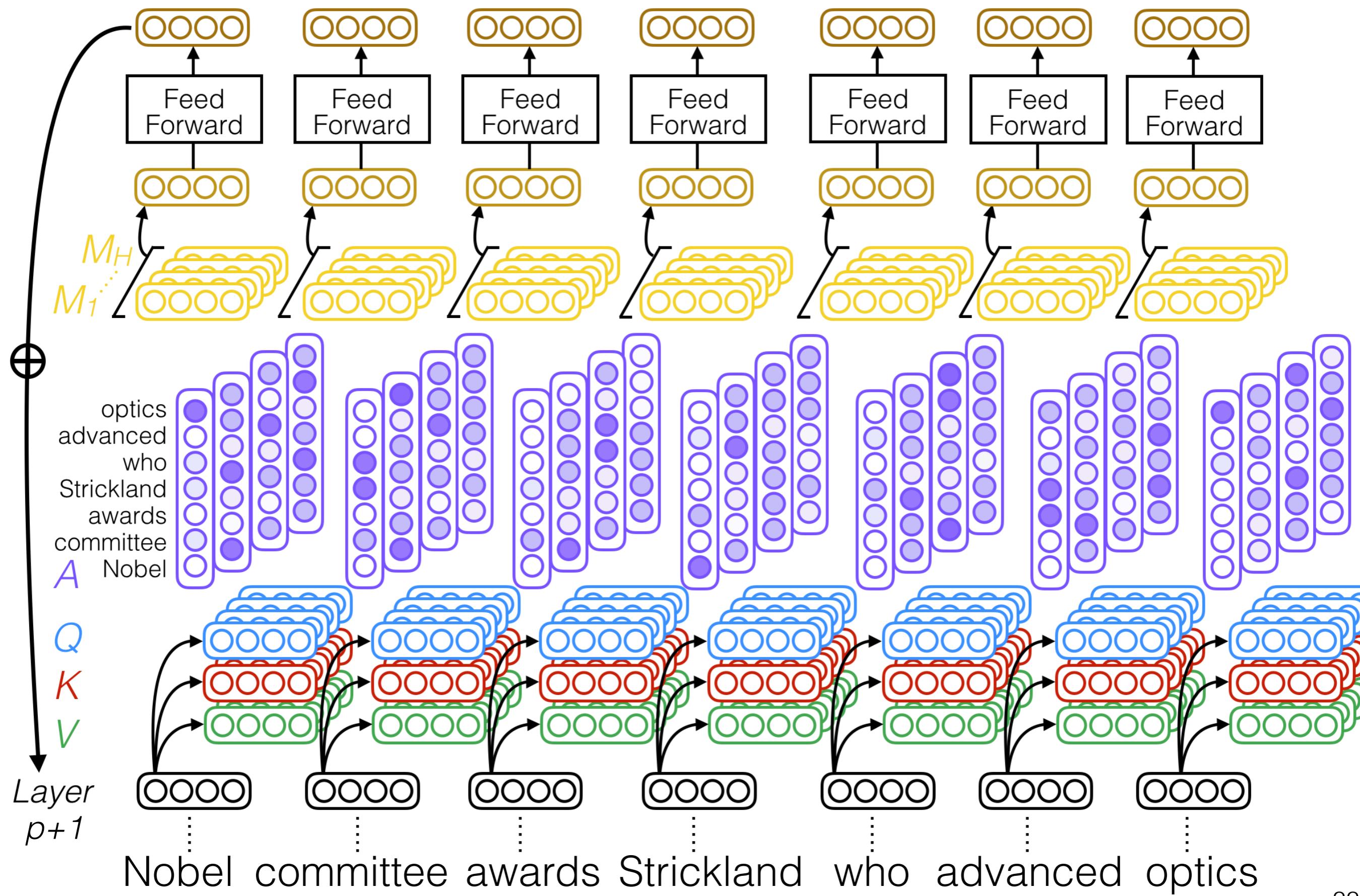
# Multi-head self-attention



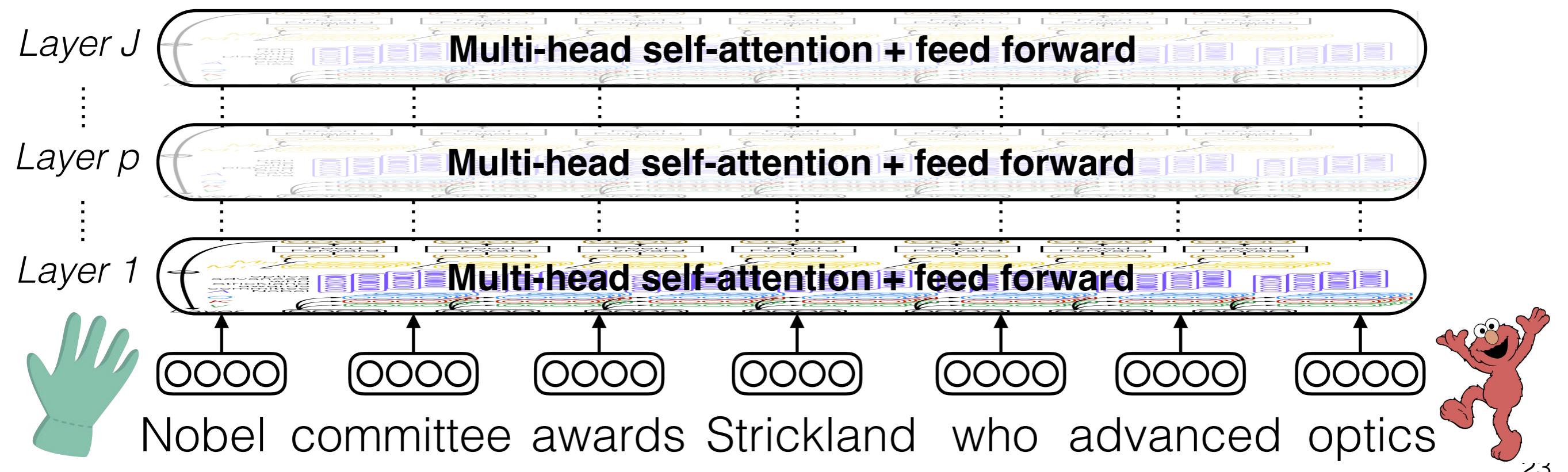
# Multi-head self-attention



# Multi-head self-attention



# Multi-head self-attention



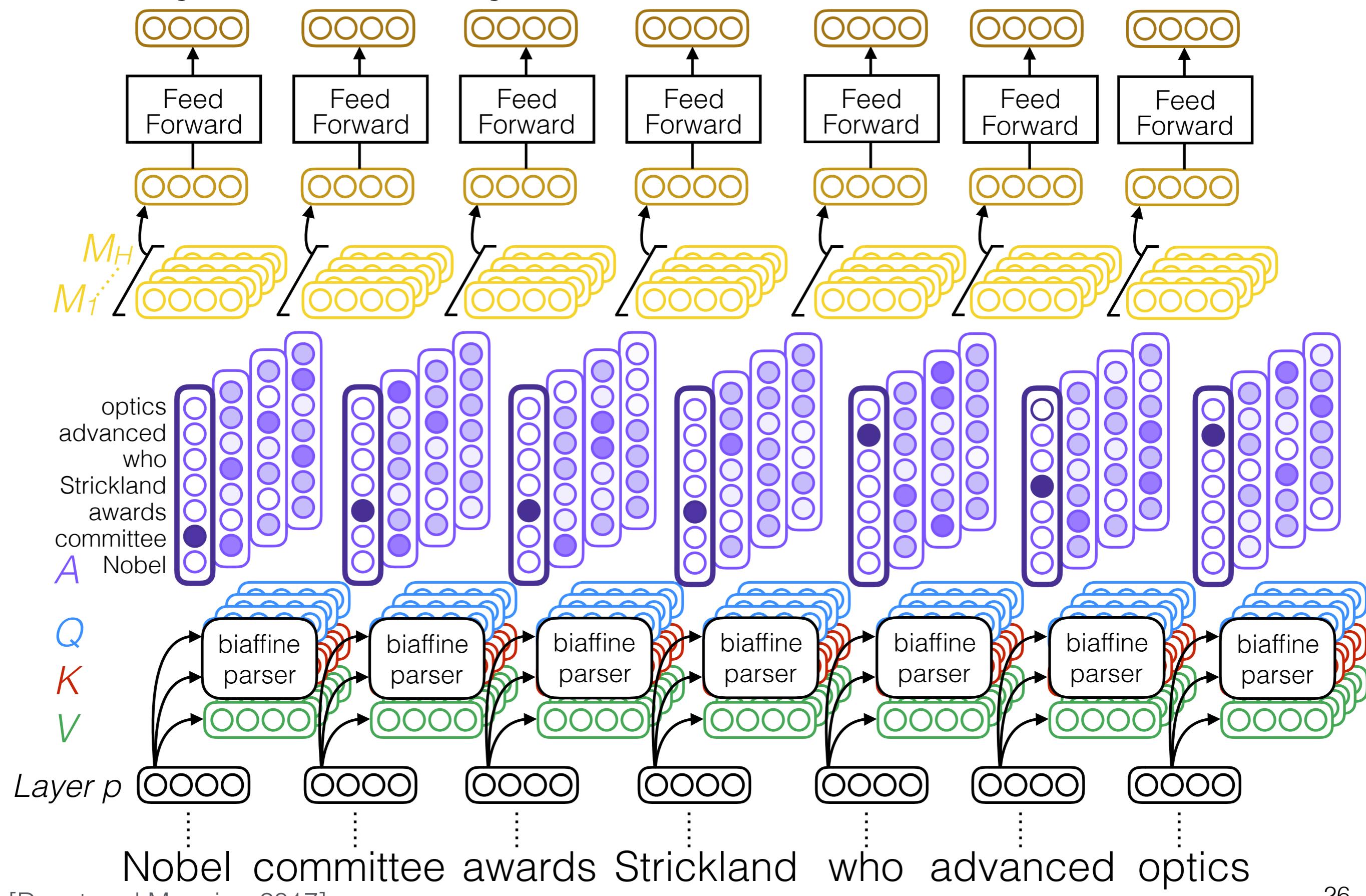
# Outline

- 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

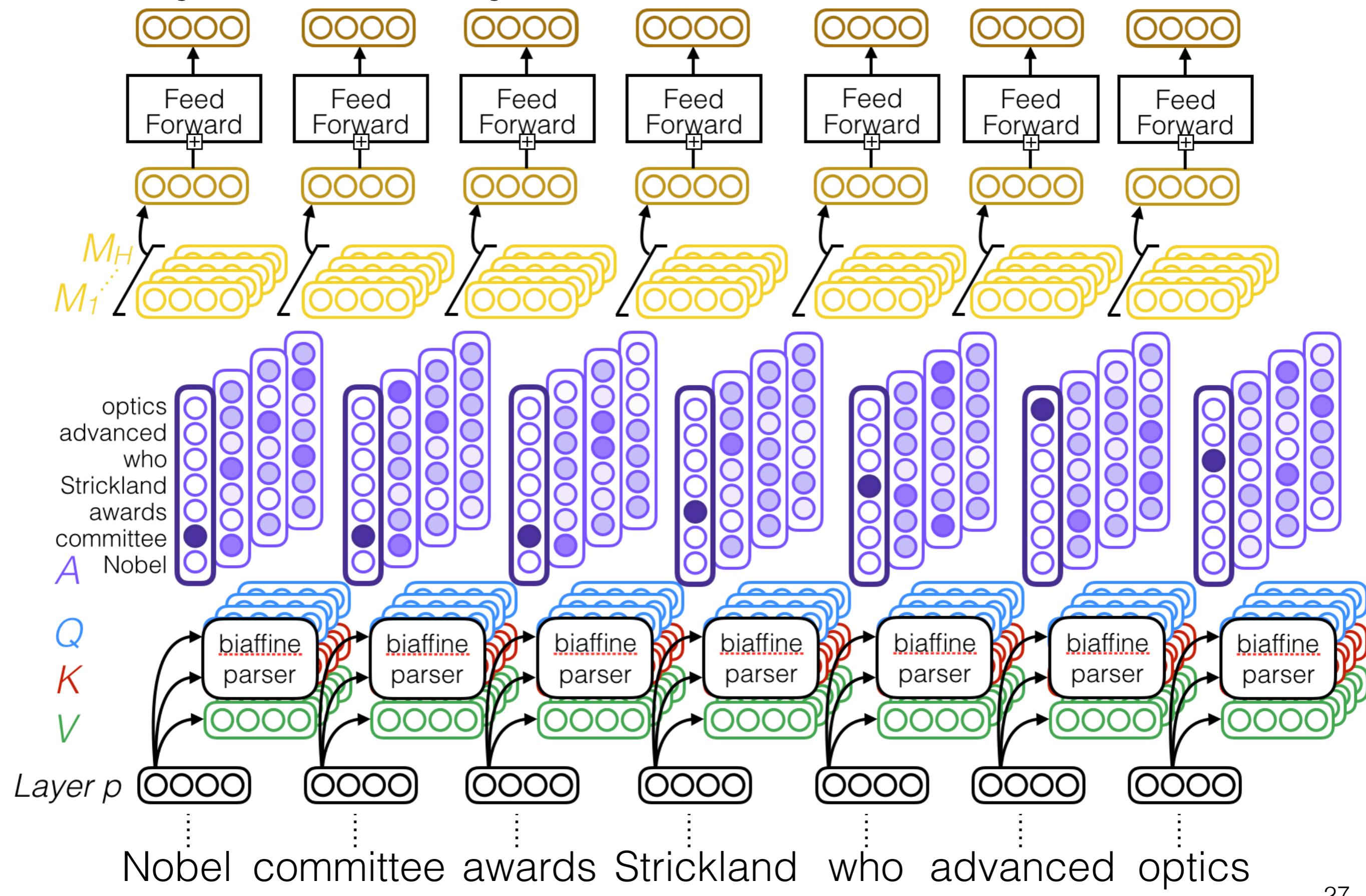
# How to incorporate syntax?

- 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.

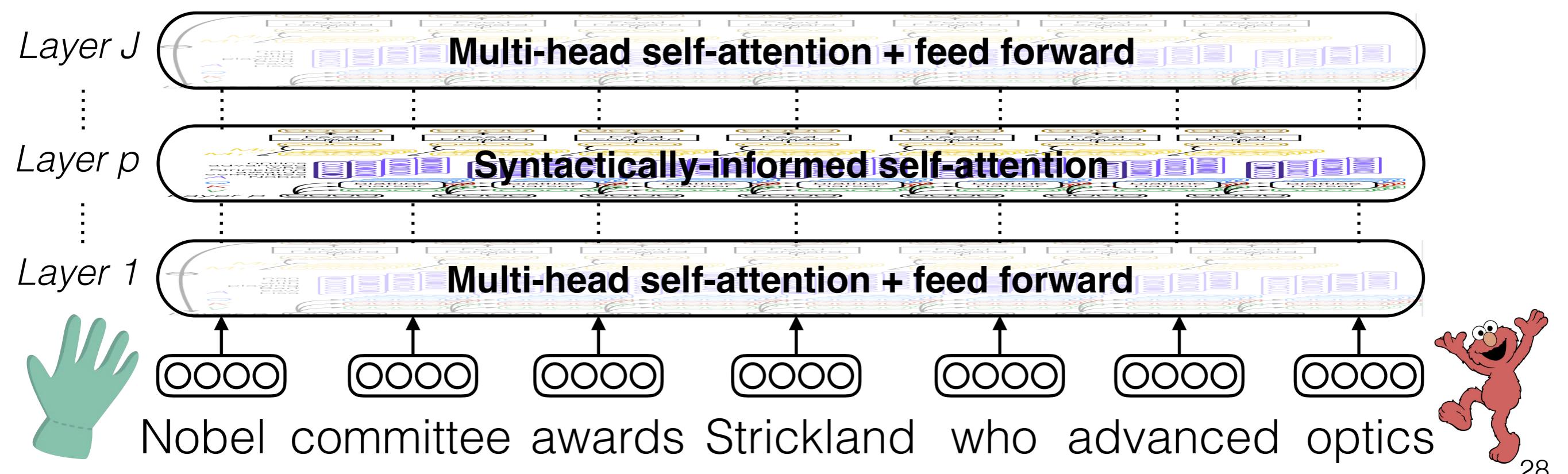
# Syntactically-informed self-attention



# Syntactically-informed self-attention



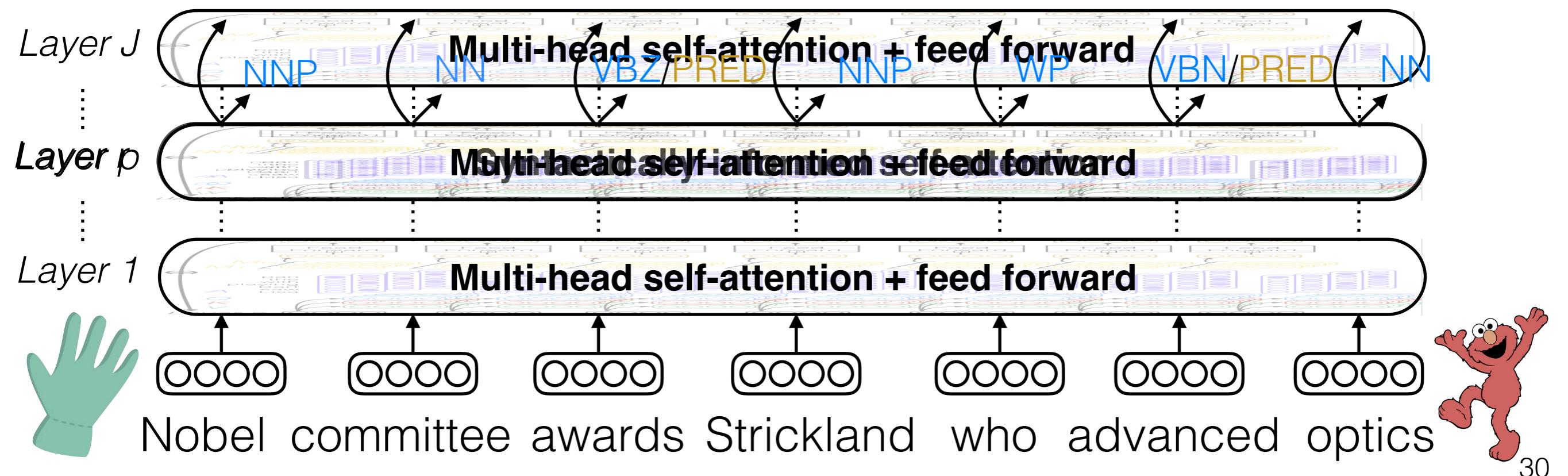
# Syntactically-informed self-attention



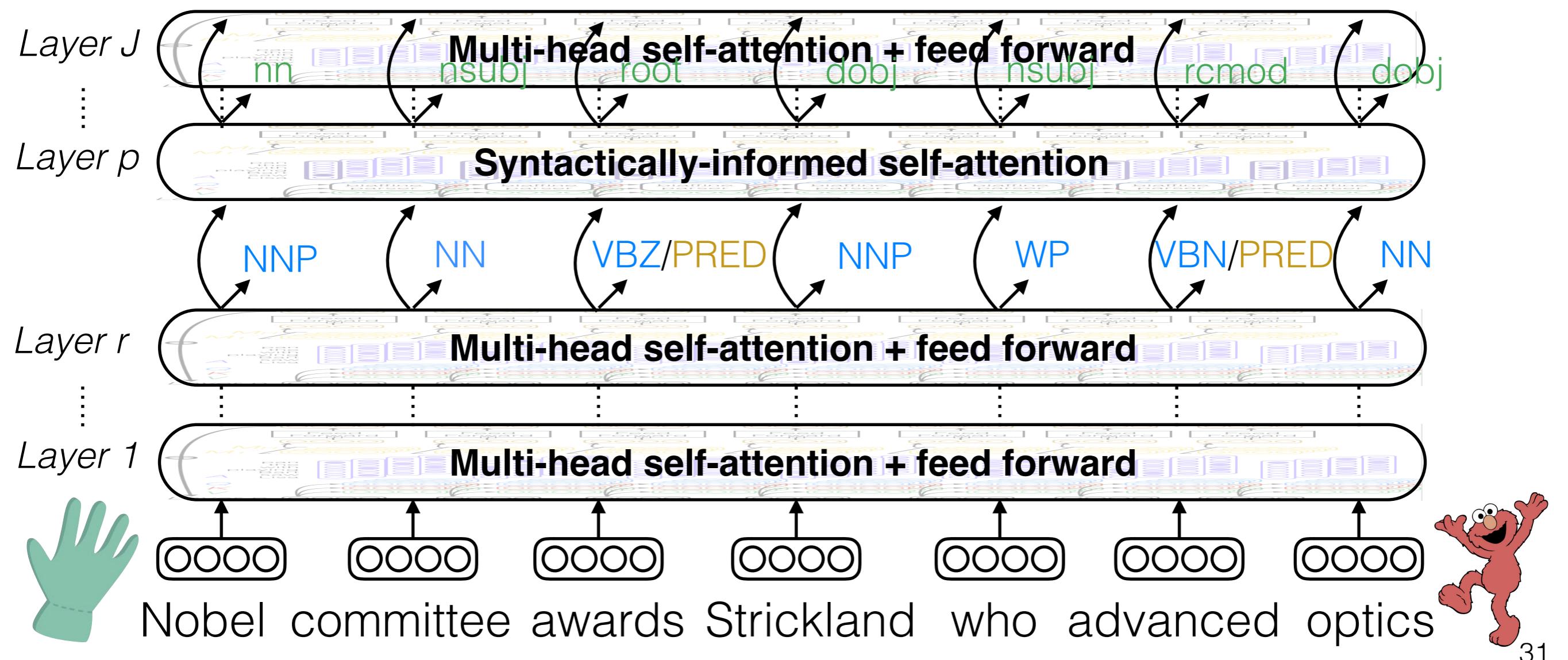
# Outline

- 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

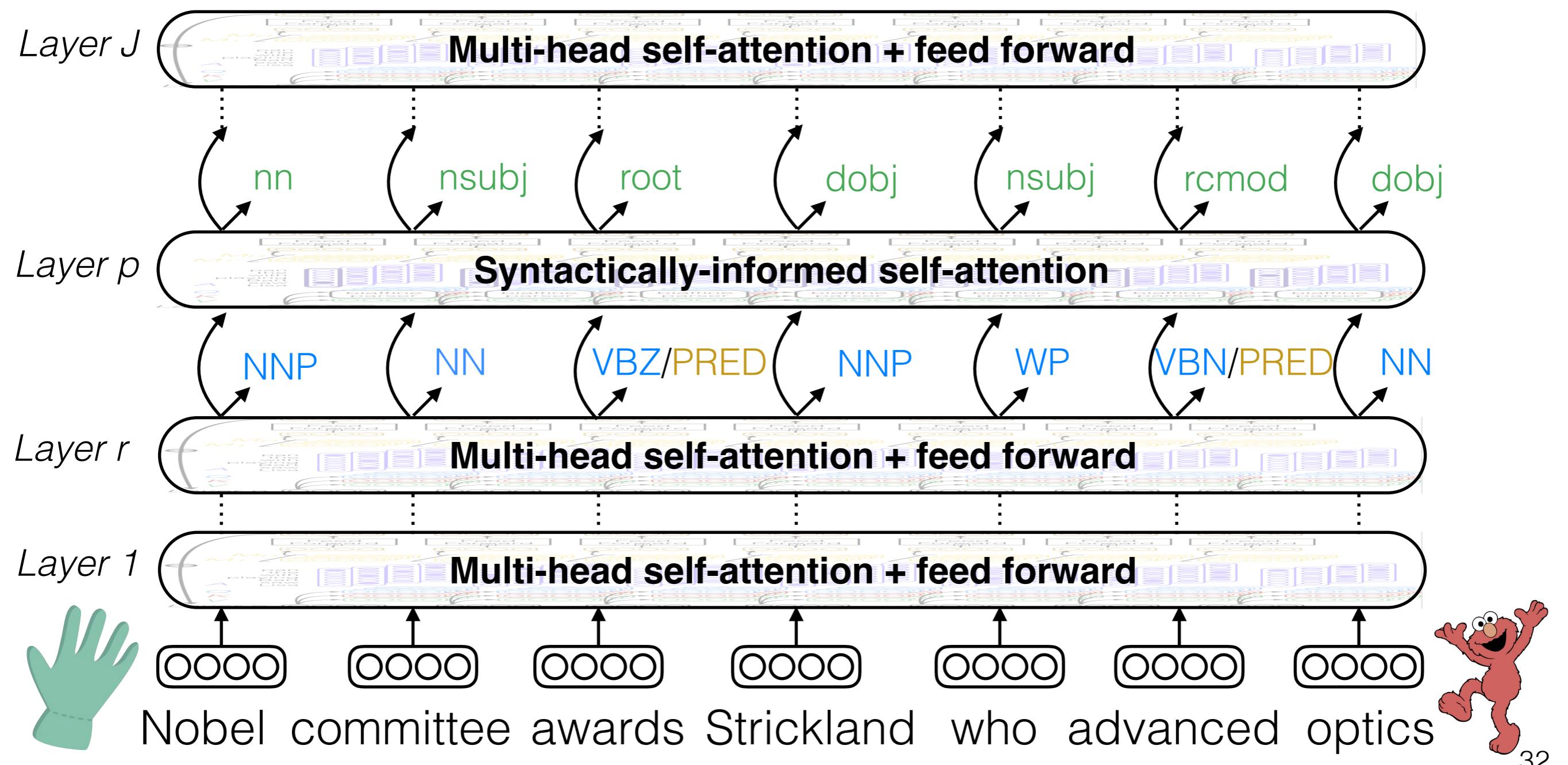
# LISA: Linguistically-Informed Self-Attention



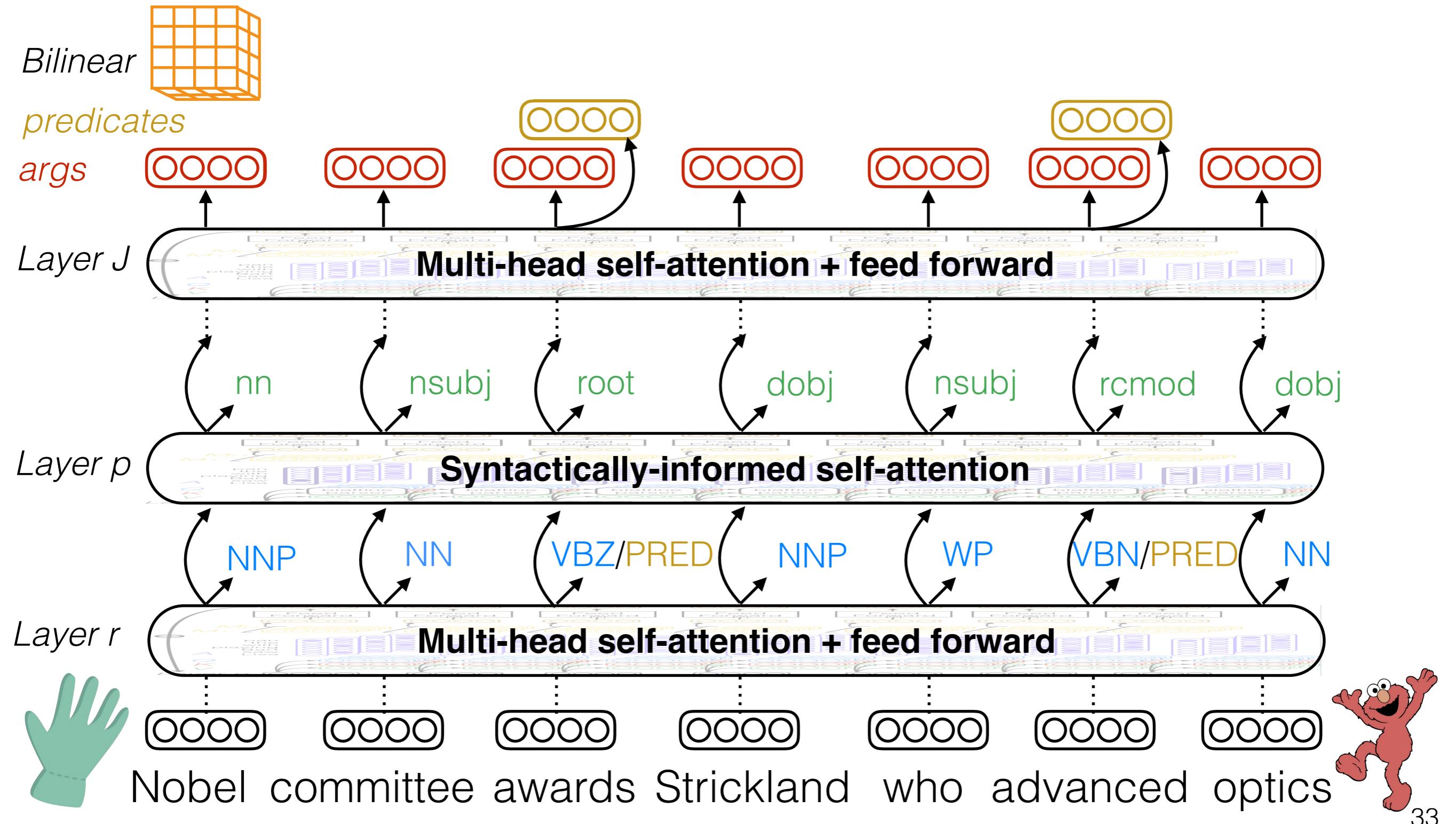
# LISA: Linguistically-Informed Self-Attention



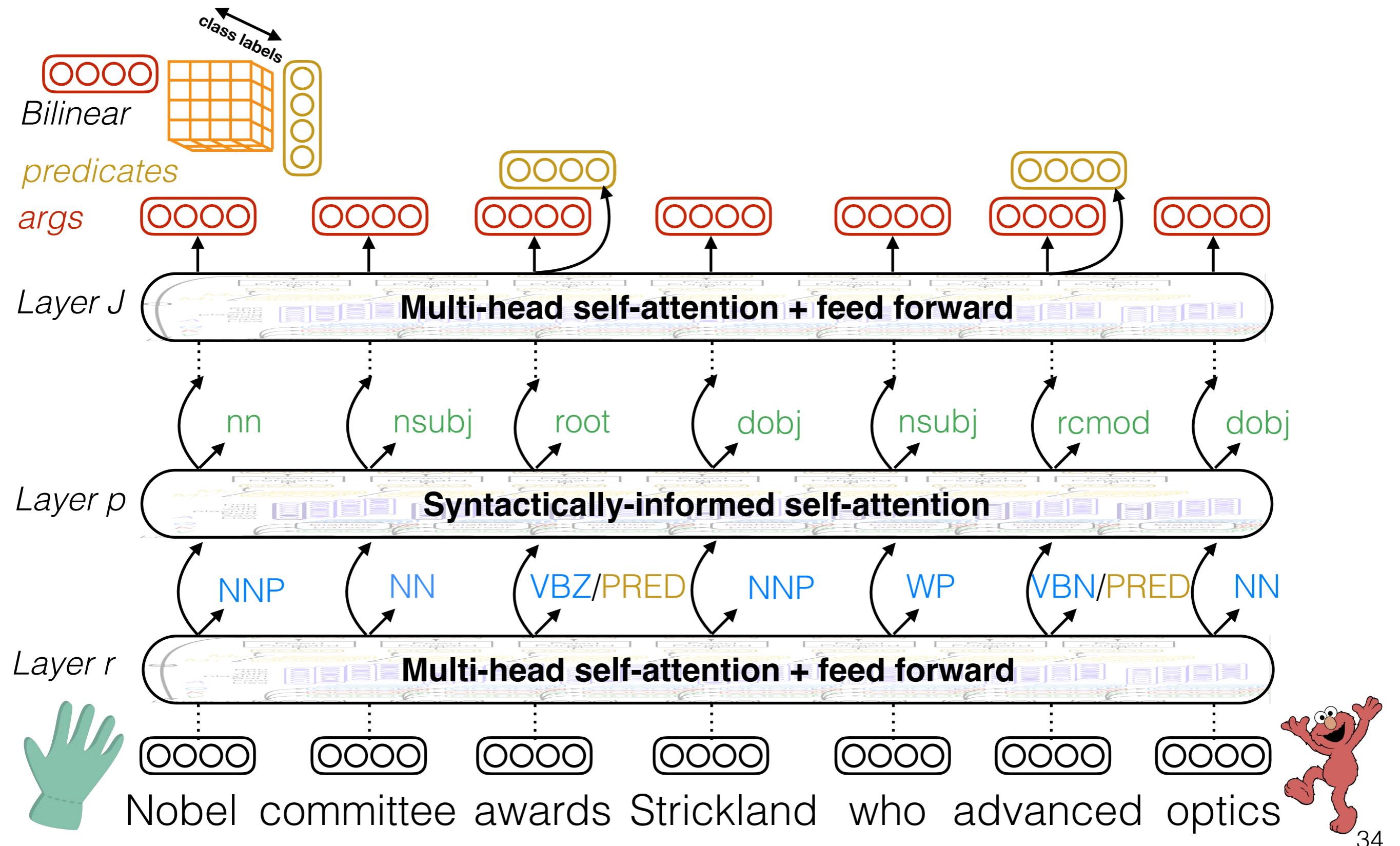
# LISA: Linguistically-Informed Self-Attention



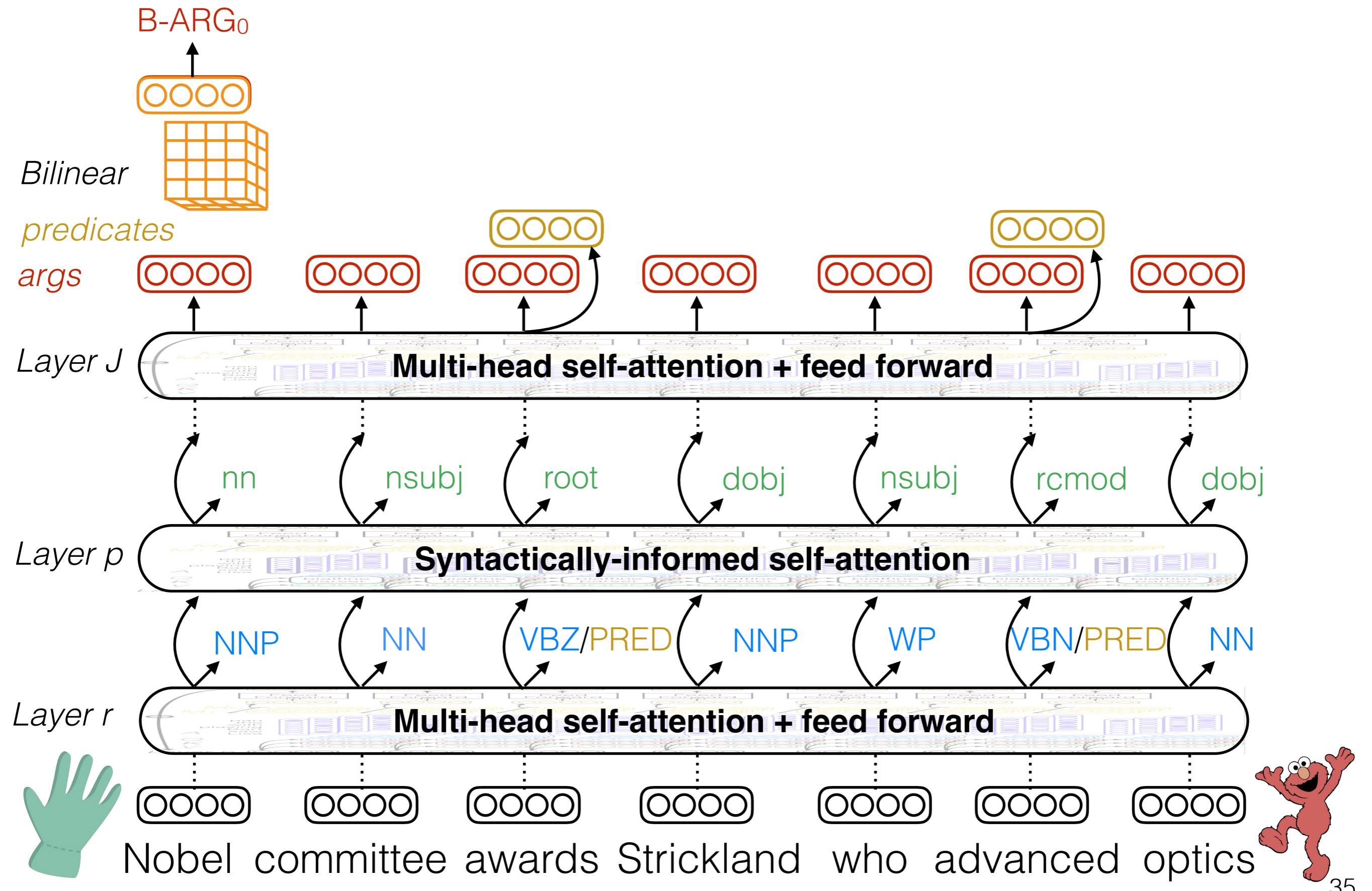
# LISA: Linguistically-Informed Self-Attention



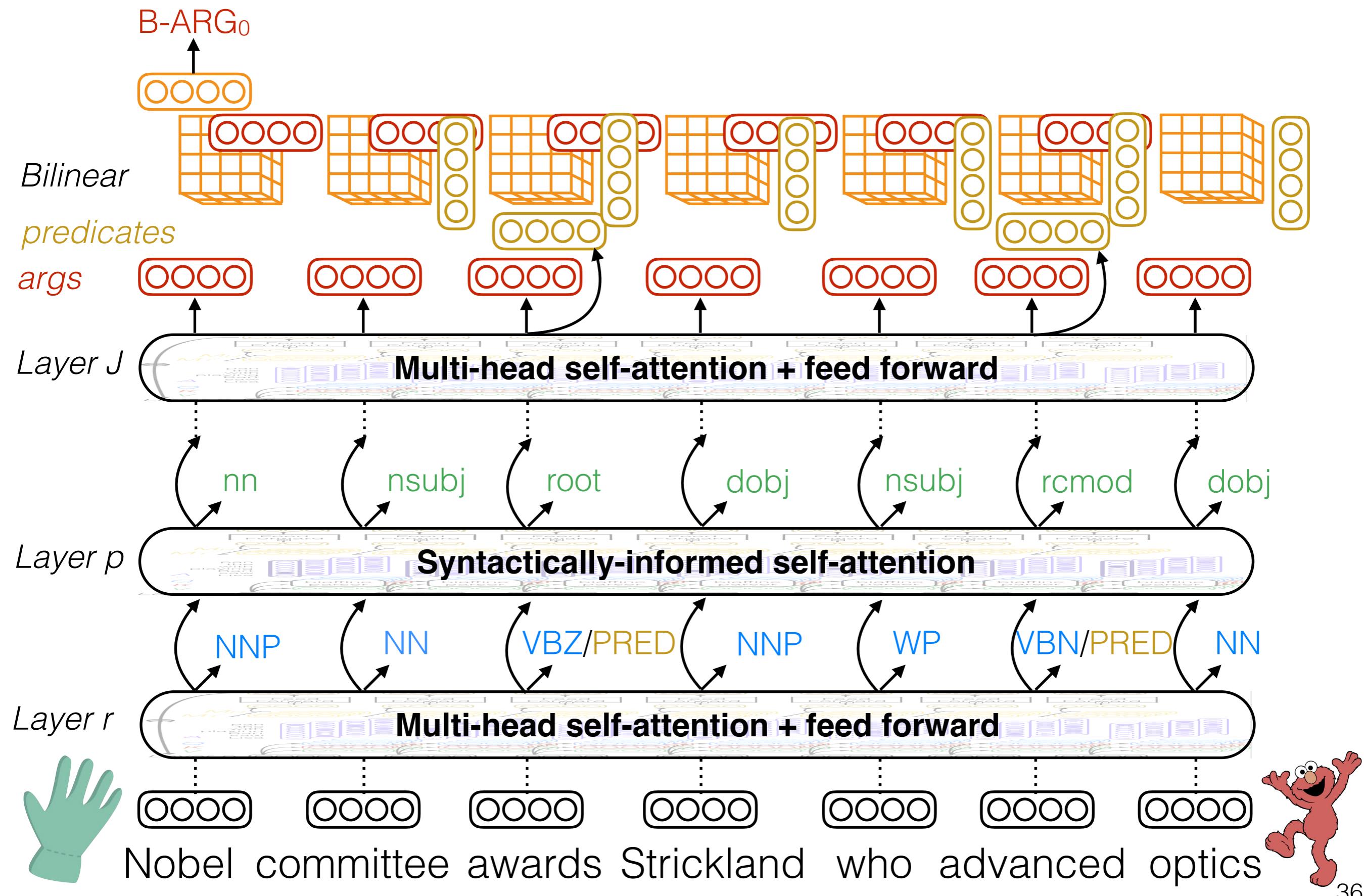
# LISA: Linguistically-Informed Self-Attention



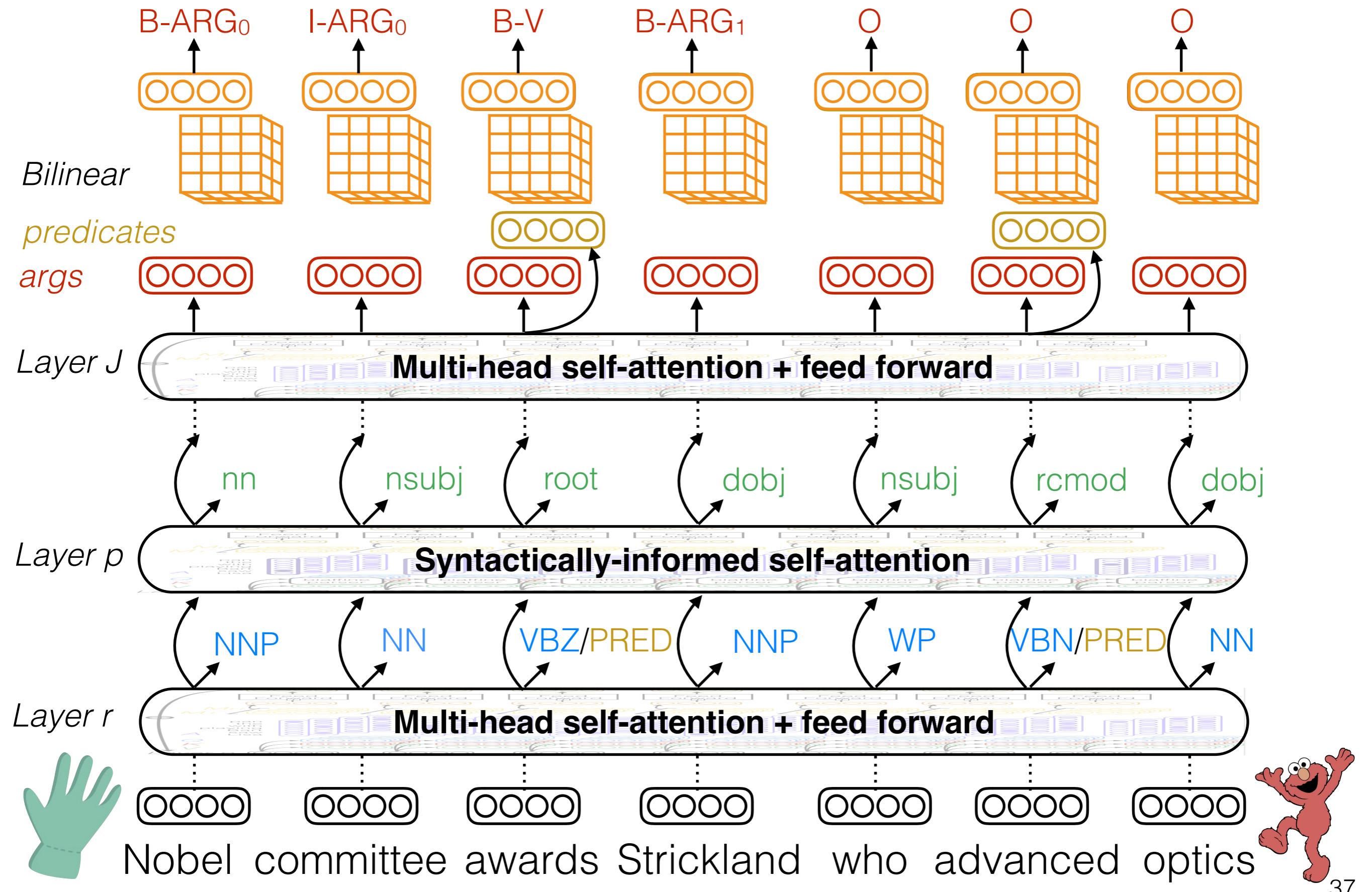
# LISA: Linguistically-Informed Self-Attention



# LISA: Linguistically-Informed Self-Attention

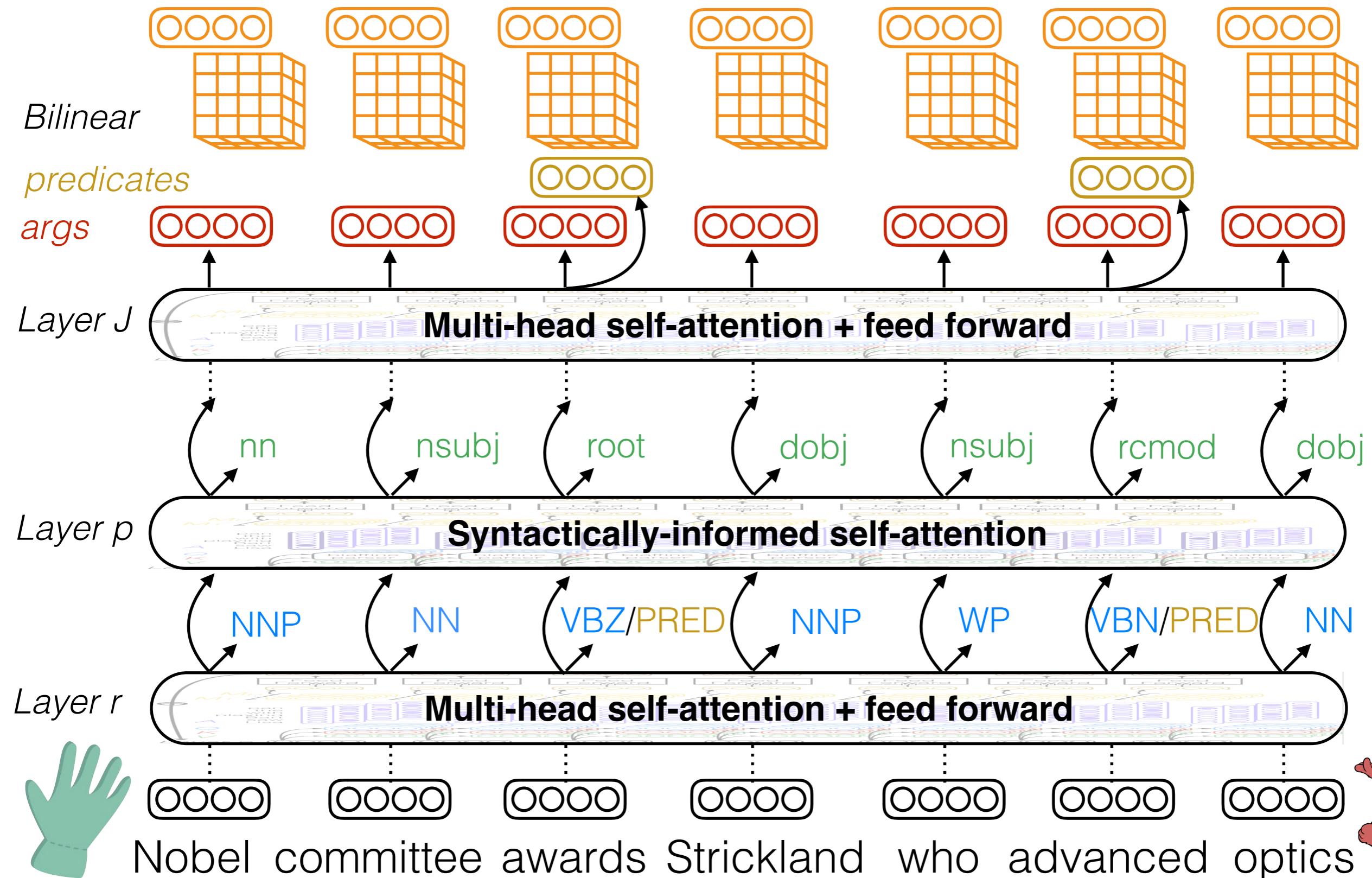


# LISA: Linguistically-Informed Self-Attention



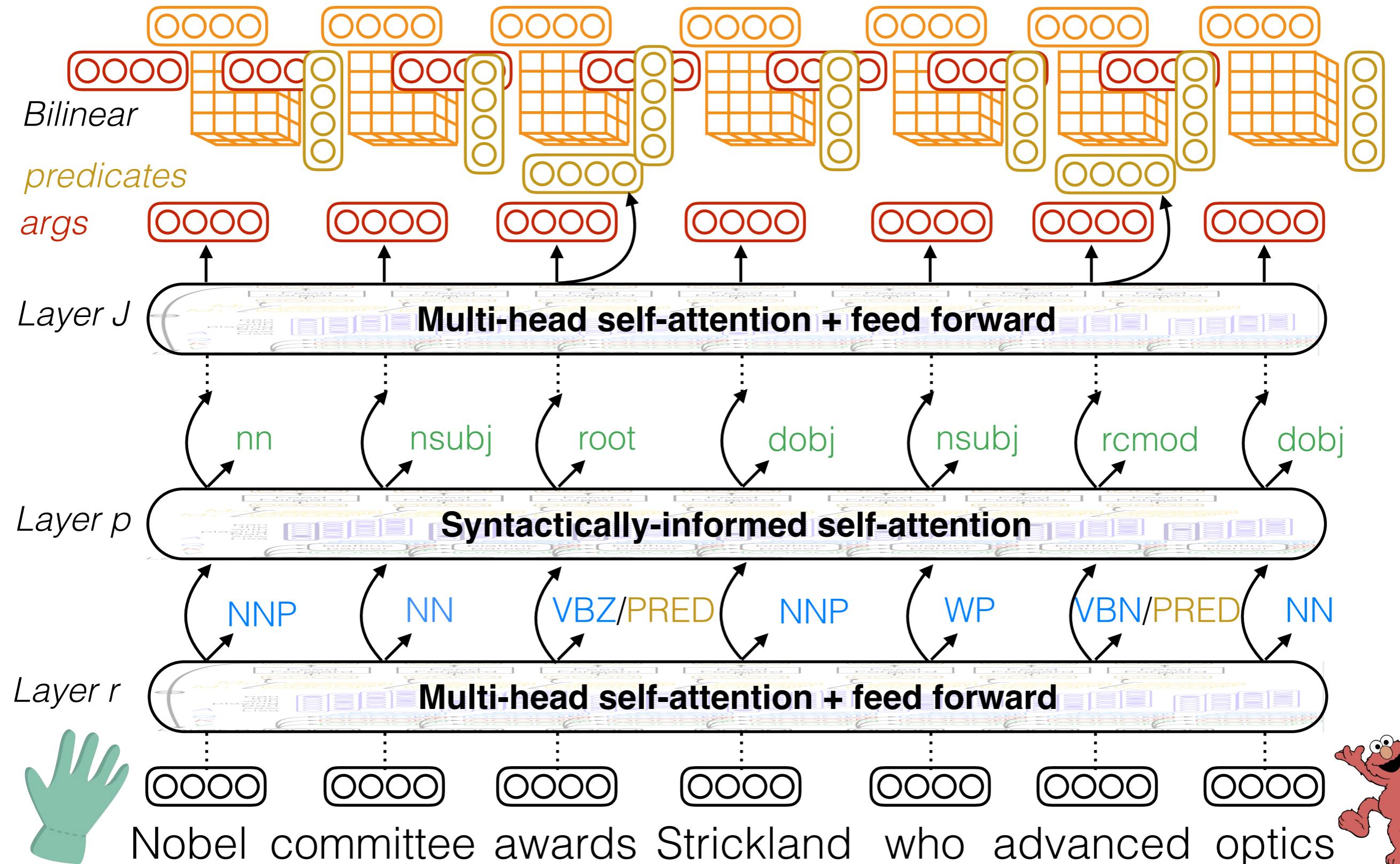
# LISA: Linguistically-Informed Self-Attention

B-ARG<sub>0</sub> I-ARG<sub>0</sub> B-V B-ARG<sub>1</sub> O O O



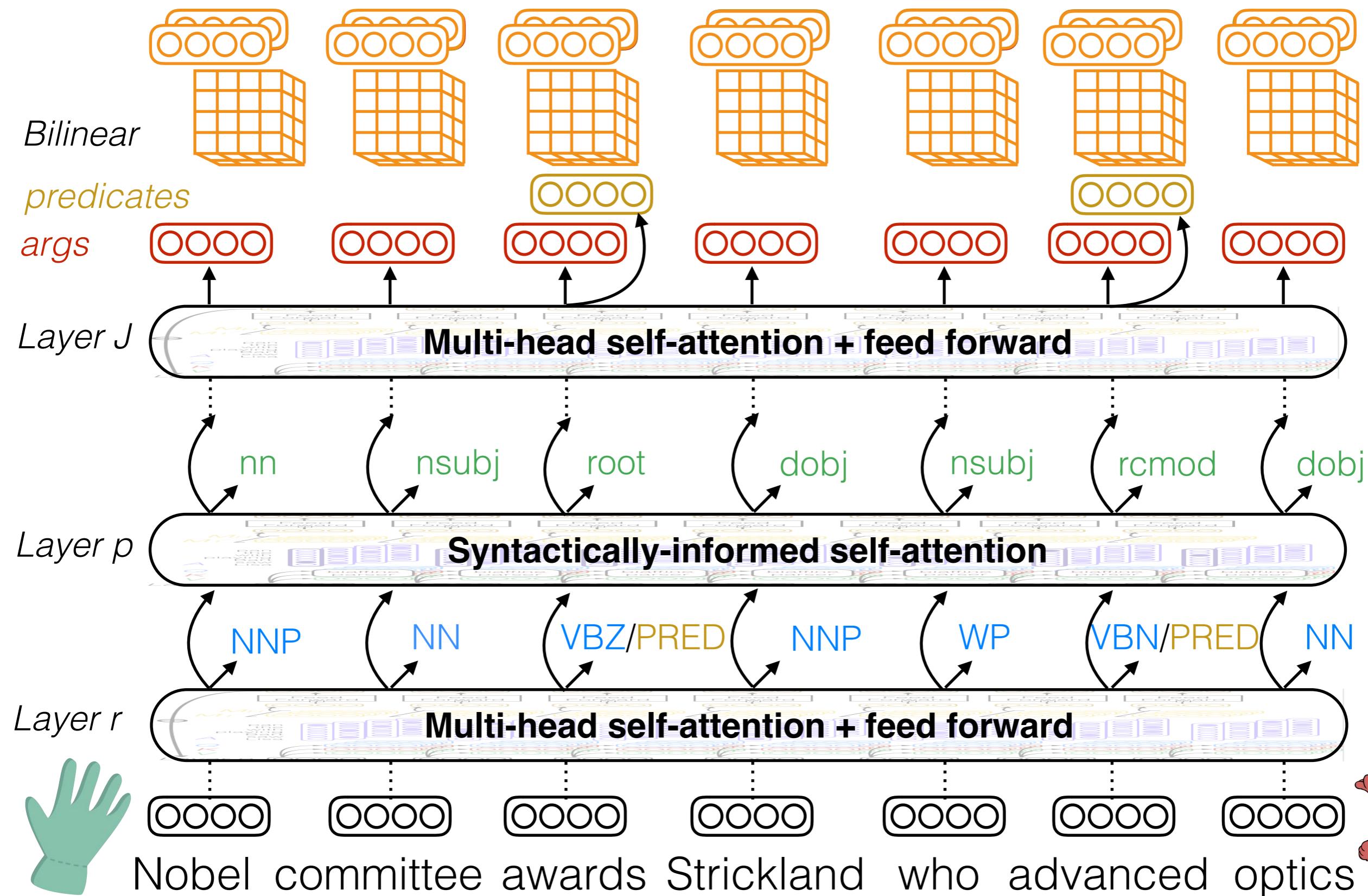
# LISA: Linguistically-Informed Self-Attention

B-ARG<sub>0</sub> I-ARG<sub>0</sub> B-V B-ARG<sub>1</sub> O O O

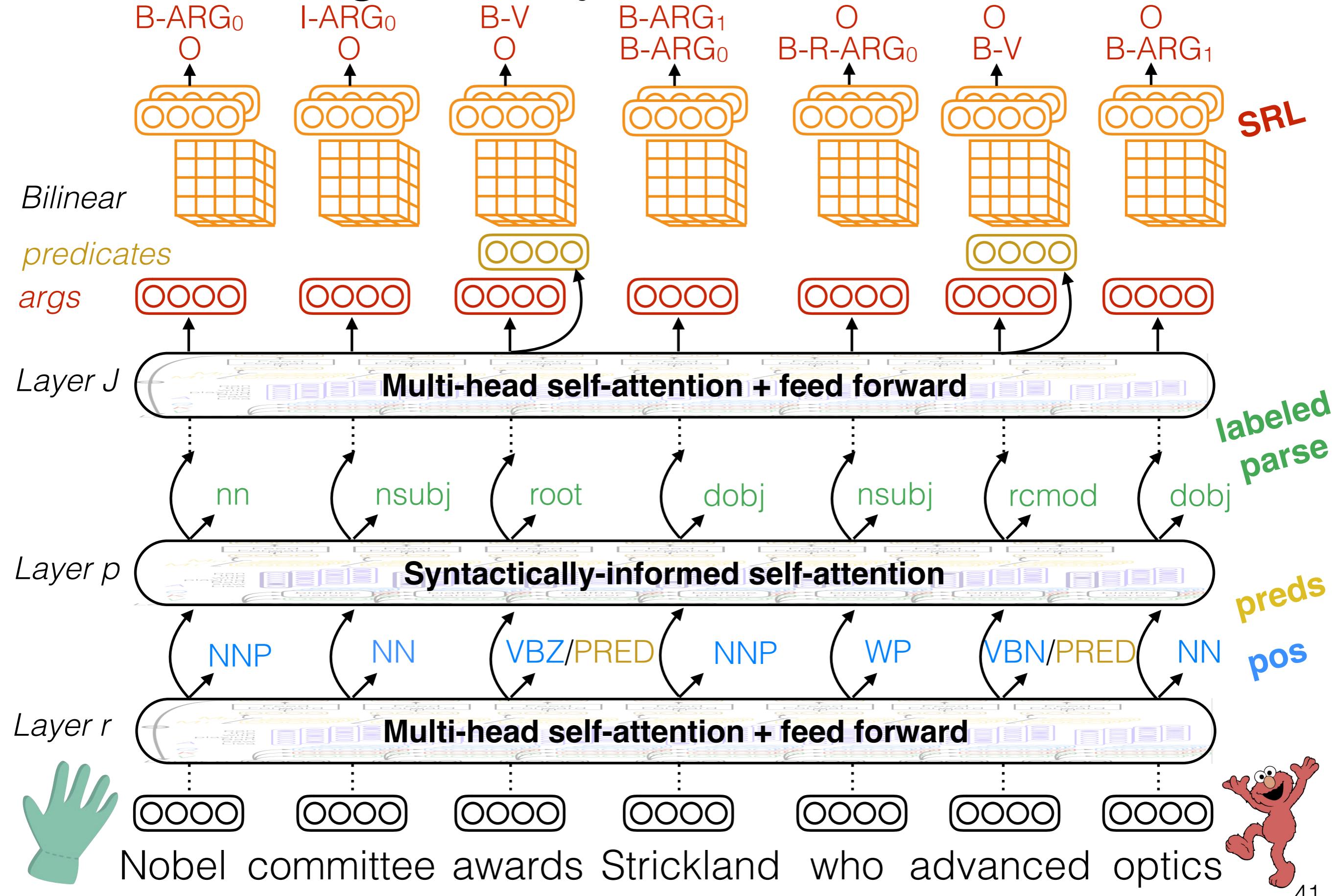


# LISA: Linguistically-Informed Self-Attention

B-ARG<sub>0</sub> I-ARG<sub>0</sub> B-V B-ARG<sub>1</sub> 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
- Experimental results & error analysis

# Experimental results

	<b>CoNLL-2005</b>	<b>CoNLL-2012</b>
<b>domains</b>	Train, dev: news Test: news, novels	Train, dev, test: 7 domains (news, telephone, bible, ...)
<b>word embeddings</b>	GloVe [Pennington et al. 2014] ELMo [Peters et al. 2018]	GloVe [Pennington et al. 2014] ELMo [Peters et al. 2018]
<b>predicates</b>	predicted; gold	predicted
<b>baselines</b>	He et al. 2017 He et al. 2018 Tan et al. 2018	He et al. 2018
<b>our models</b>	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 UAS 84.99	90.3 UAS 74.66	96.6 UAS 86.90	93.4 UAS 82.25

*+2.49 F1      +3.86 F1      +0.9 F1      +2.15 F1*

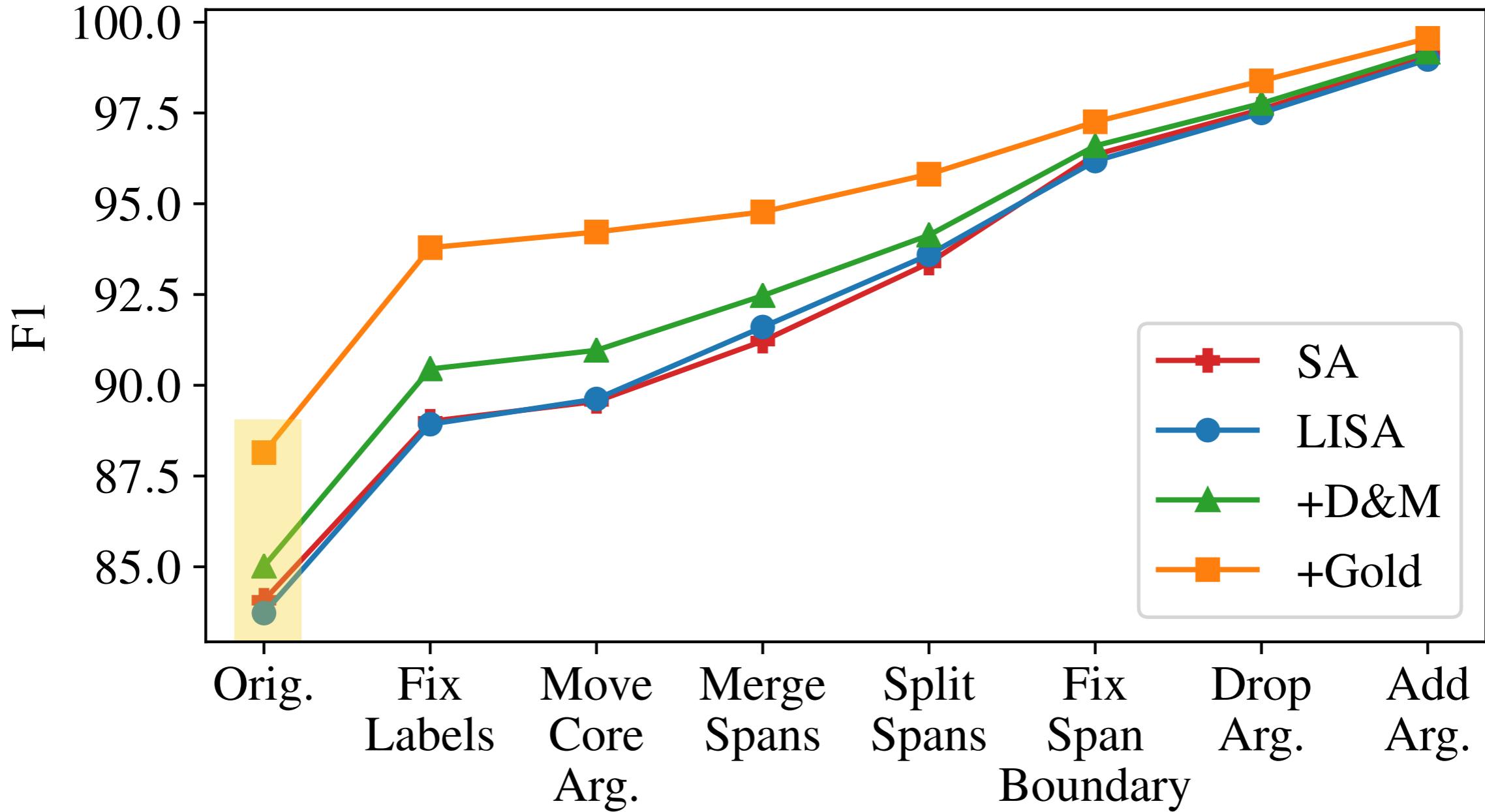
?

# Experimental results: CoNLL-2005

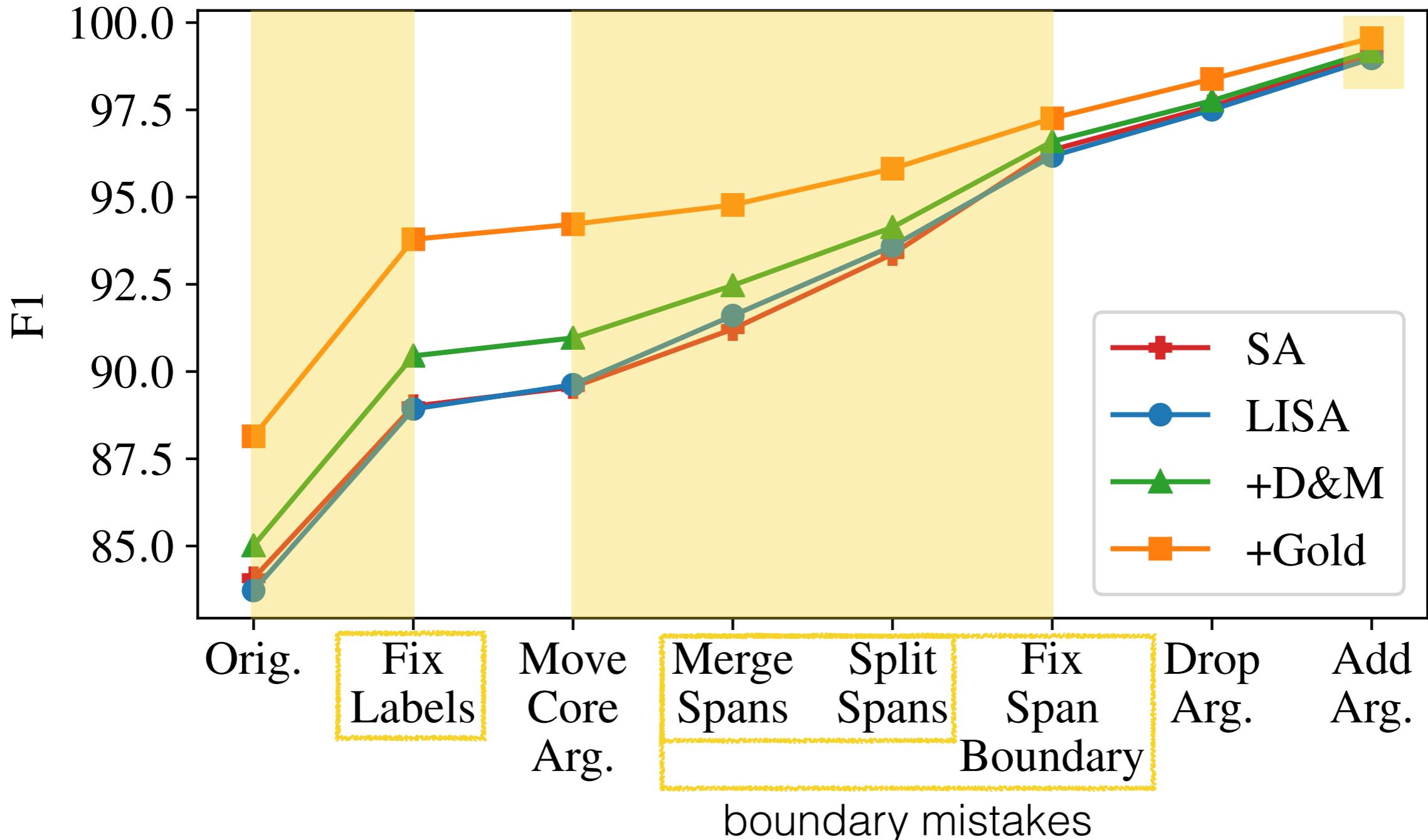


96.5 UAS!

# Experimental results: Analysis



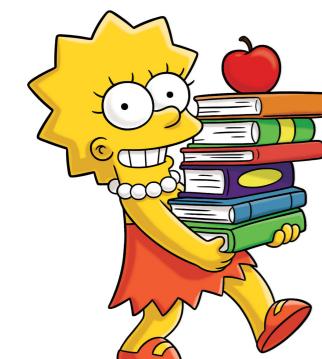
# Experimental results: Analysis



# 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>