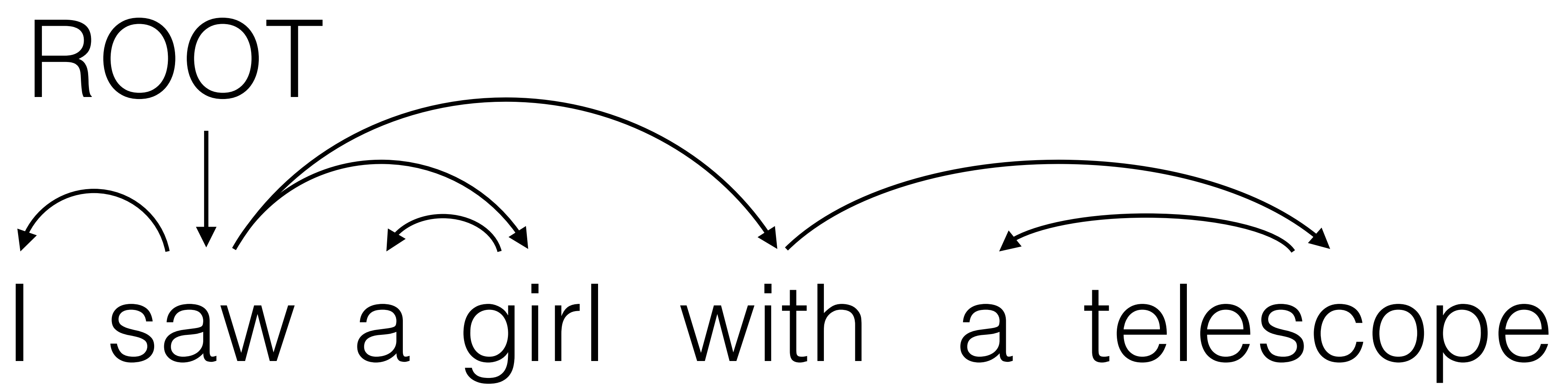
[Using content from Graham Neubig and Emma Strubell et al.]

Dependency Parsing

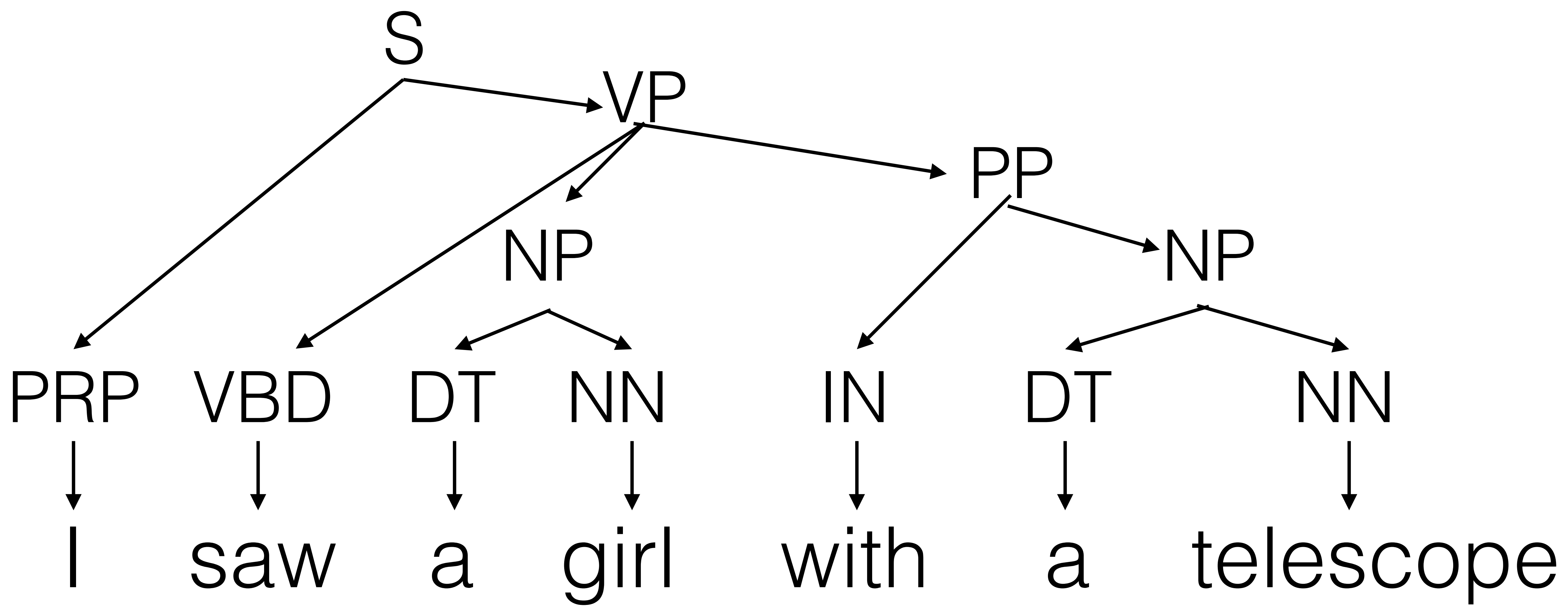
Tim Rocktäschel & **Sebastian Riedel**
COMP0087 Natural Language Processing

Two Types of Linguistic Structure

- **Dependency:** focus on relations between words

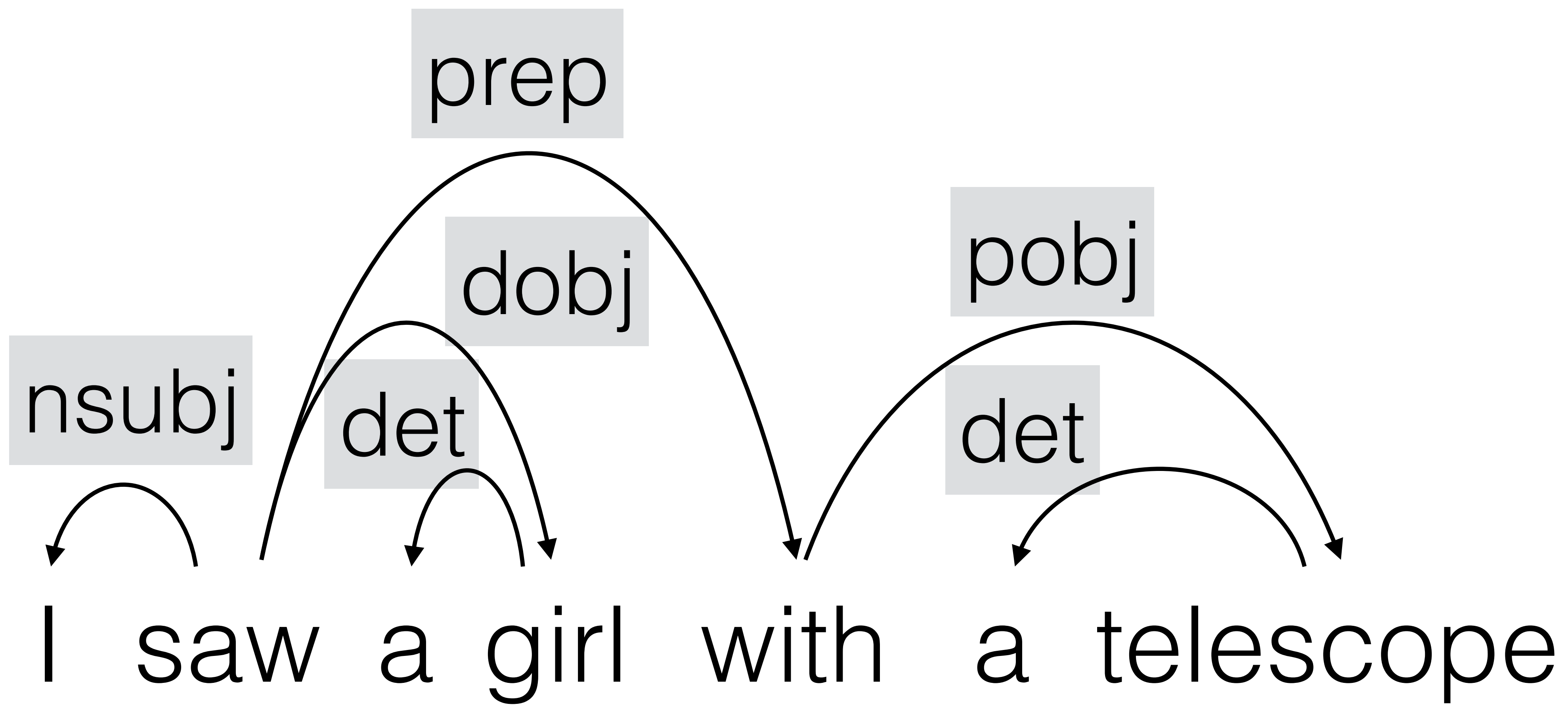


- **Phrase structure:** focus on the structure of the sentence



Why Dependencies?

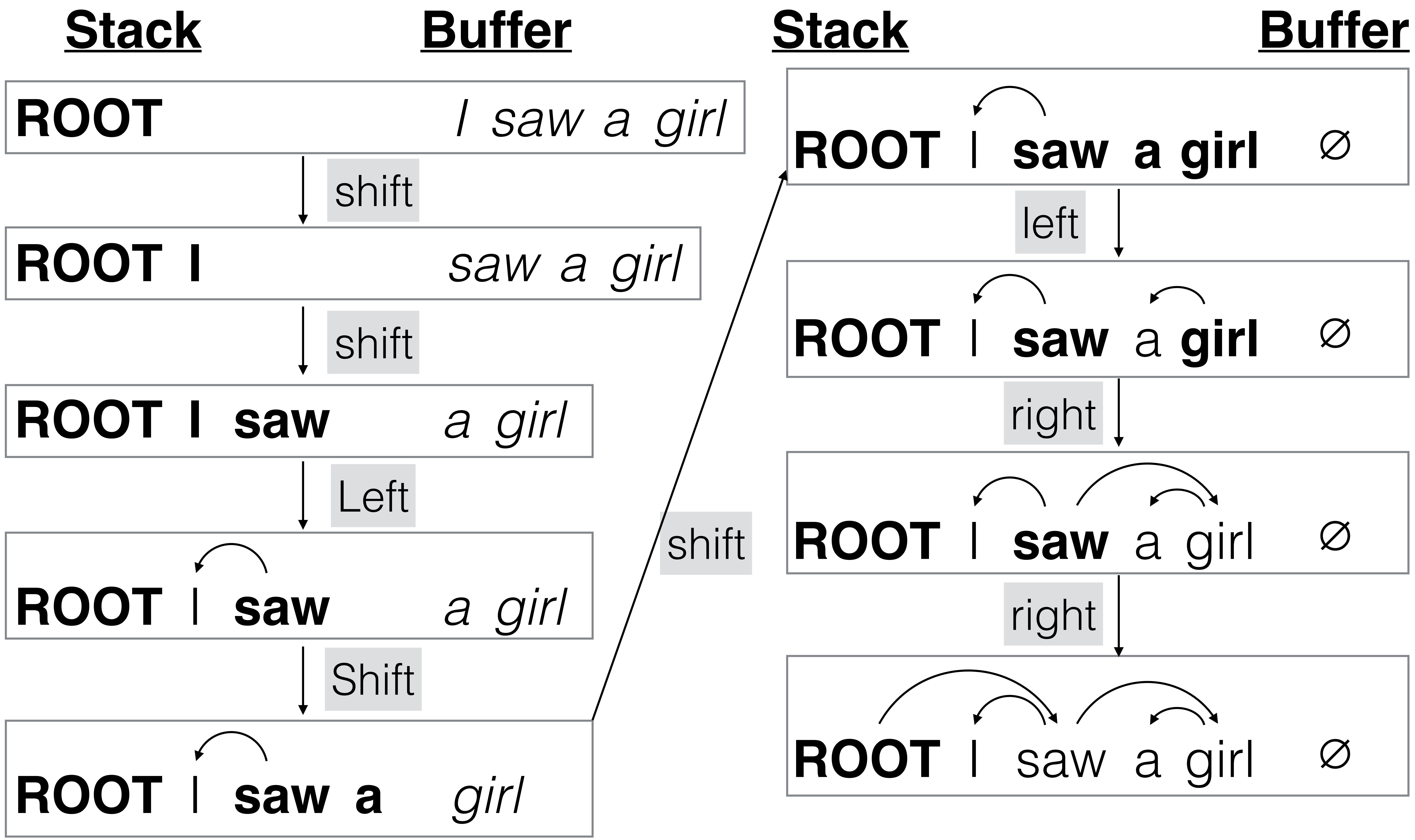
- Dependencies are often good for semantic tasks, as related words are close in the tree
- It is also possible to create labeled dependencies, that explicitly show the relationship between words
- Easier to annotate and understand



Arc Standard Shift-Reduce Parsing
(Yamada & Matsumoto 2003, Nivre 2003)

- Process words one-by-one left-to-right
- Two data structures
  - **Buffer:** of unprocessed words
  - **Stack:** of partially processed words
- At each point choose
  - **shift:** move one word from queue to stack
  - **reduce left:** top word on stack is head of second word
  - **reduce right:** second word on stack is head of top word
- Learn how to choose each action with a classifier

Shift Reduce Example



Classification for Shift-reduce

- Given a **configuration**

<u>Stack</u> **<u>Buffer</u>**

ROOT I saw a *girl*

- Which **action** do we choose?

shift

ROOT I saw a girl ∅

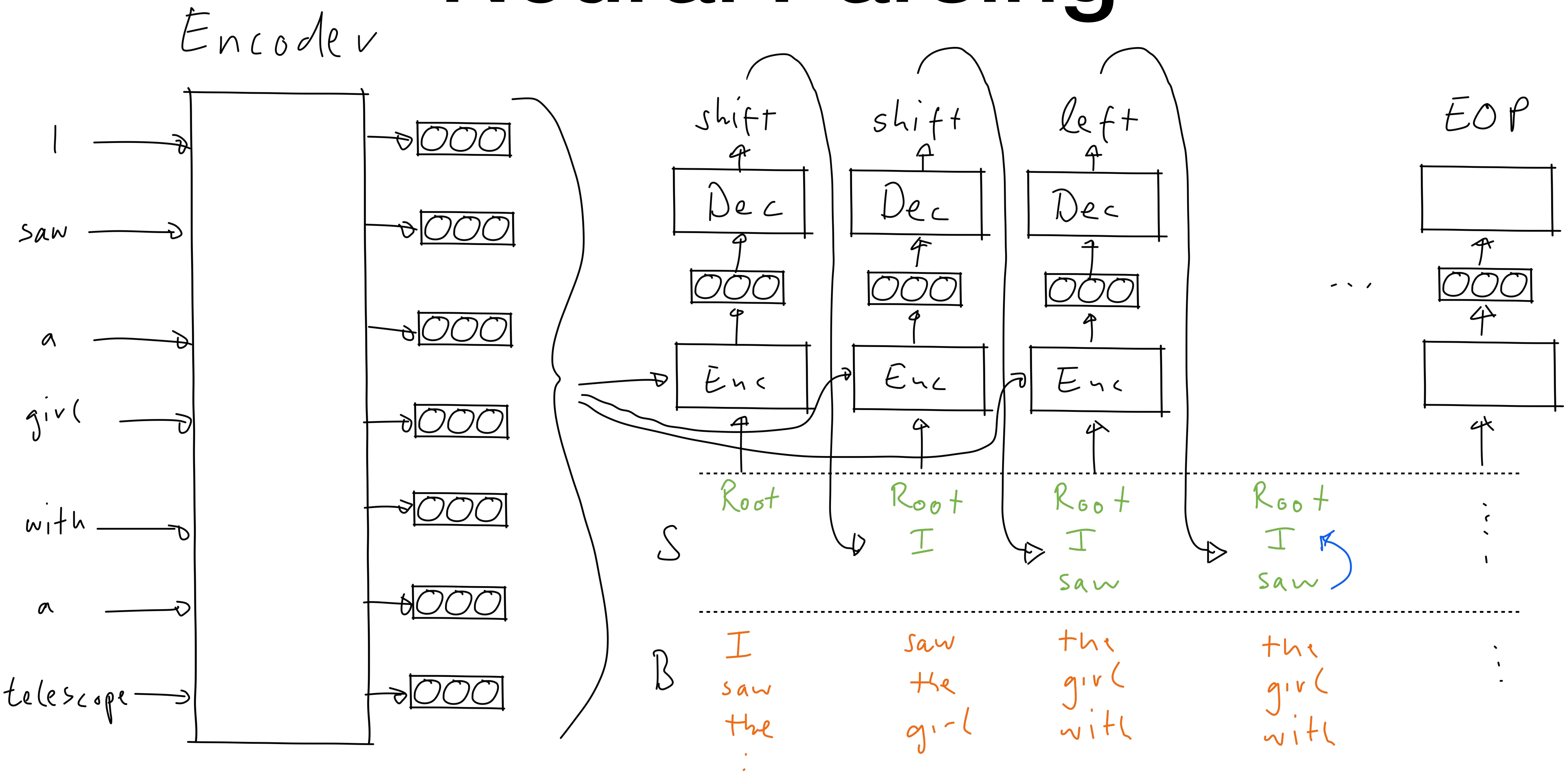
right

ROOT I saw a *girl*

left

ROOT I saw a *girl*

Neural Parsing



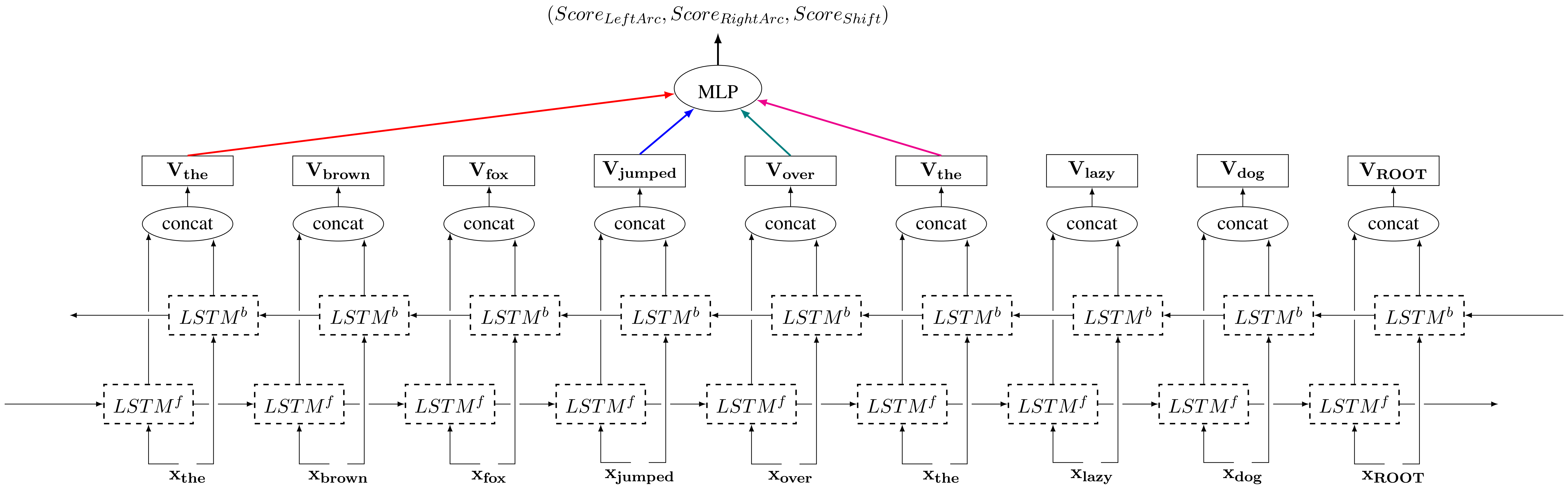
Neural Transition Based Parser

Figure here

Action Encoder-Decoder

Configuration:

Scoring:



LISA

Linguistically-Informed Self-Attention for Semantic Role Labeling



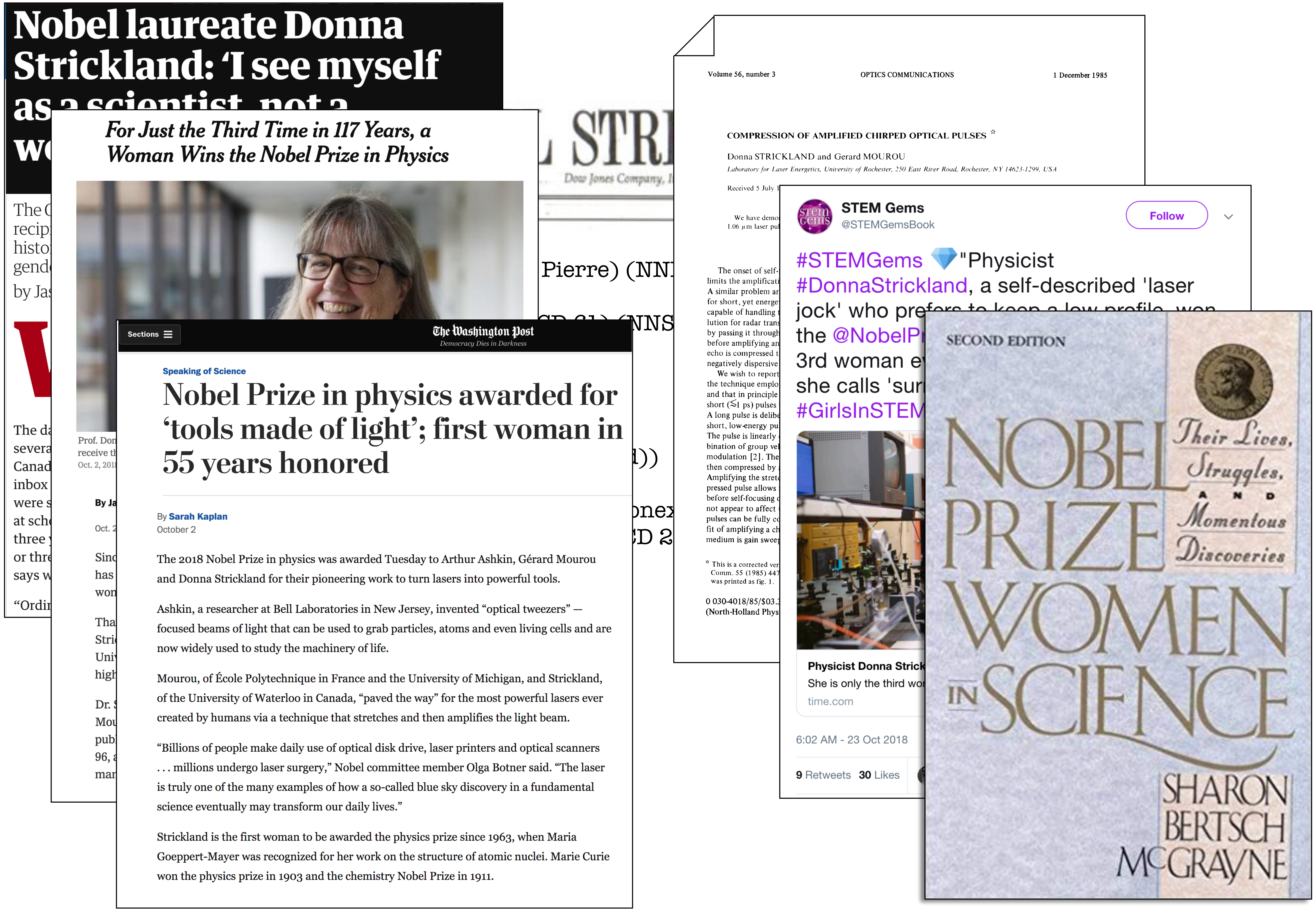


Emma Strubell[1] Patrick Verga[1] Daniel Andor[2] David Weiss[2] Andrew McCallum[1]

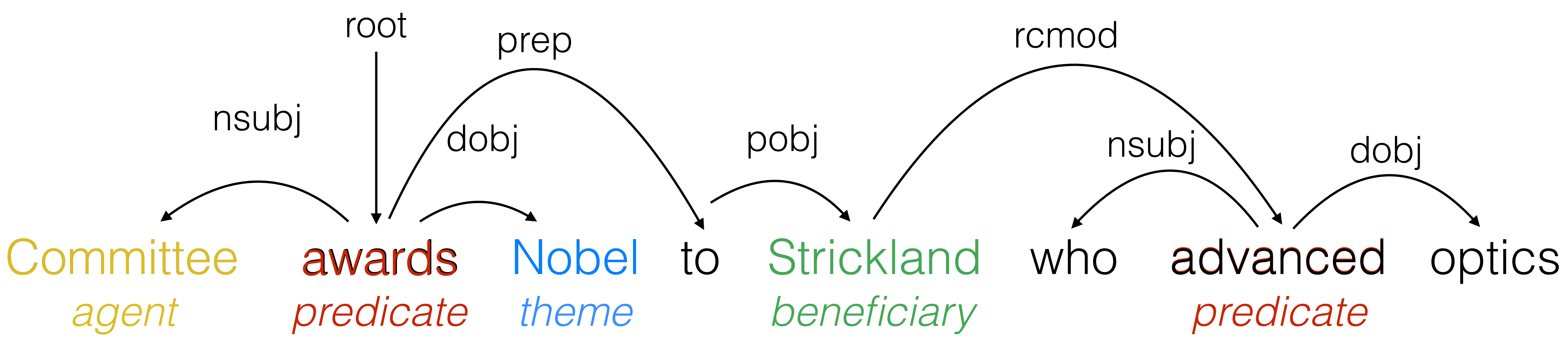
[1] UMassAmherst College of Information and Computer Sciences

[2] Google AI

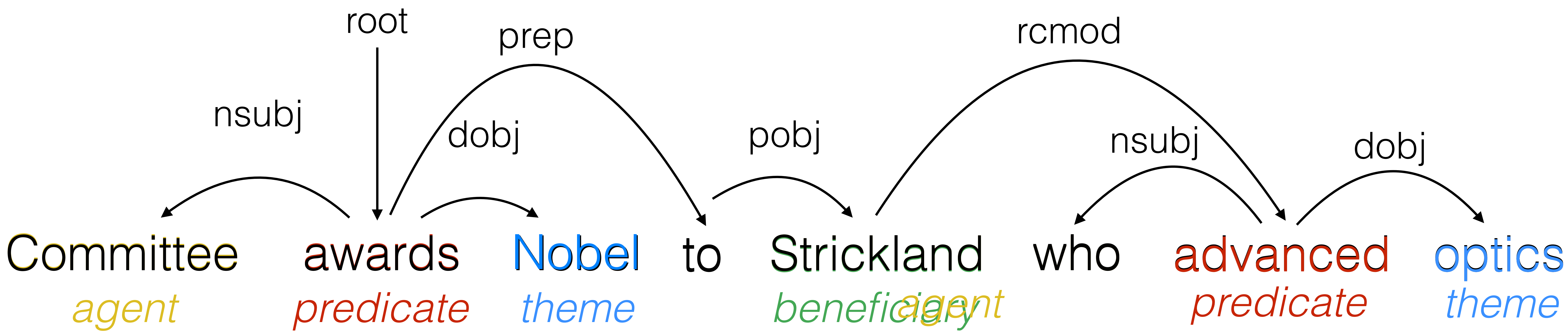
Want fast, accurate, robust NLP



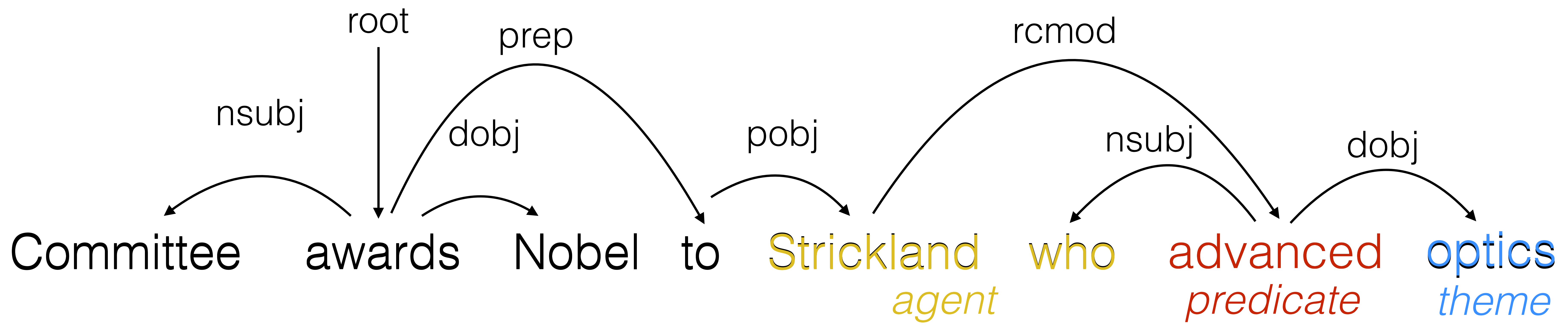
SRL: Who did what to whom?



SRL: Who did what to whom?



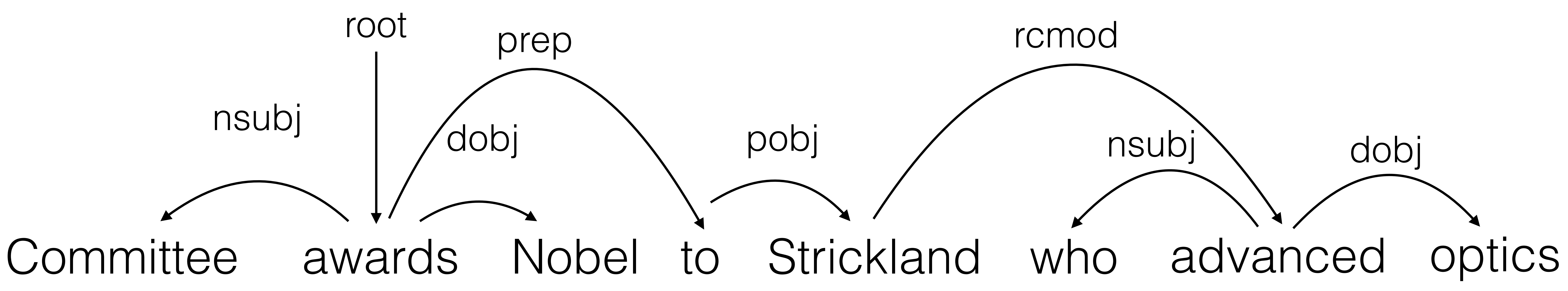
SRL: Who did what to whom?



Committee awards Nobel to Strickland who advanced optics

agent *predicate* *theme* *beneficiary*

PropBank SRL: Who did what to whom?



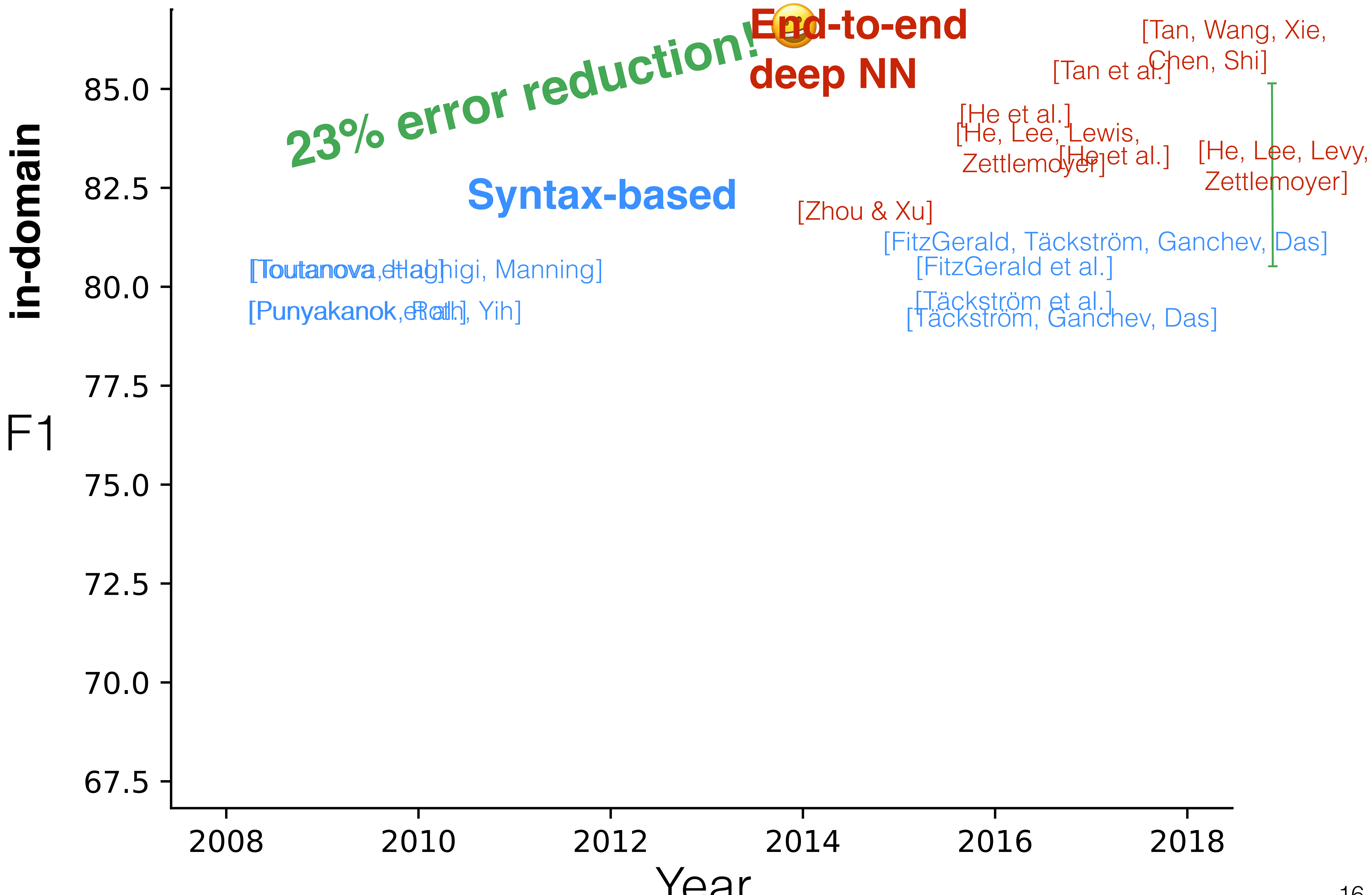
Committee awards Nobel to Strickland who advanced optics

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

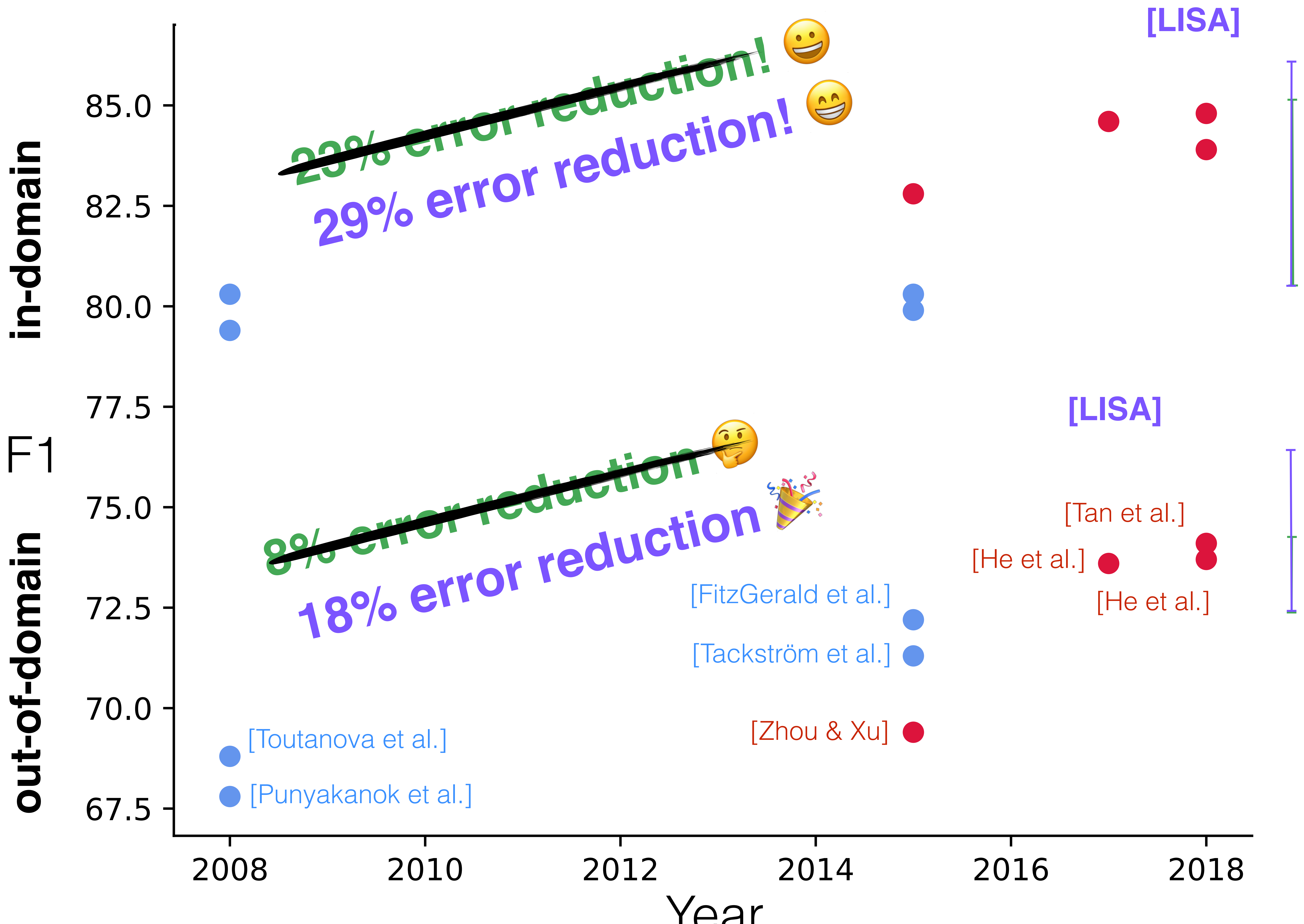
Committee awards Nobel to Strickland who advanced optics

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

10 years of PropBank SRL

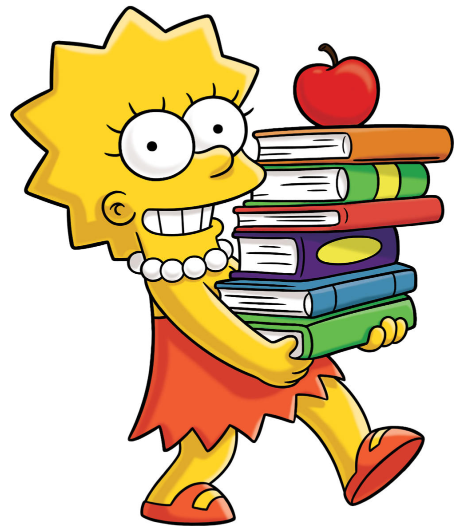


10 years of PropBank SRL



Linguistically-Informed Self-Attention

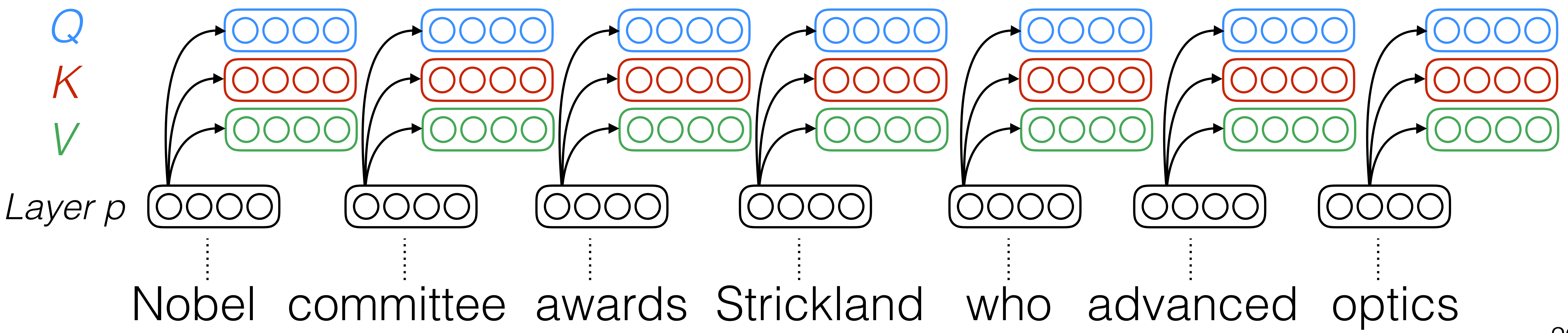
- **Multi-task learning**,
  - Part-of-speech tagging
  - Labeled dependency parsing
  - Predicate detection
  - Semantic role spans & labeling
- **Syntactically-informed self-attention**
  - Multi-head self-attention supervised by syntax



Outline

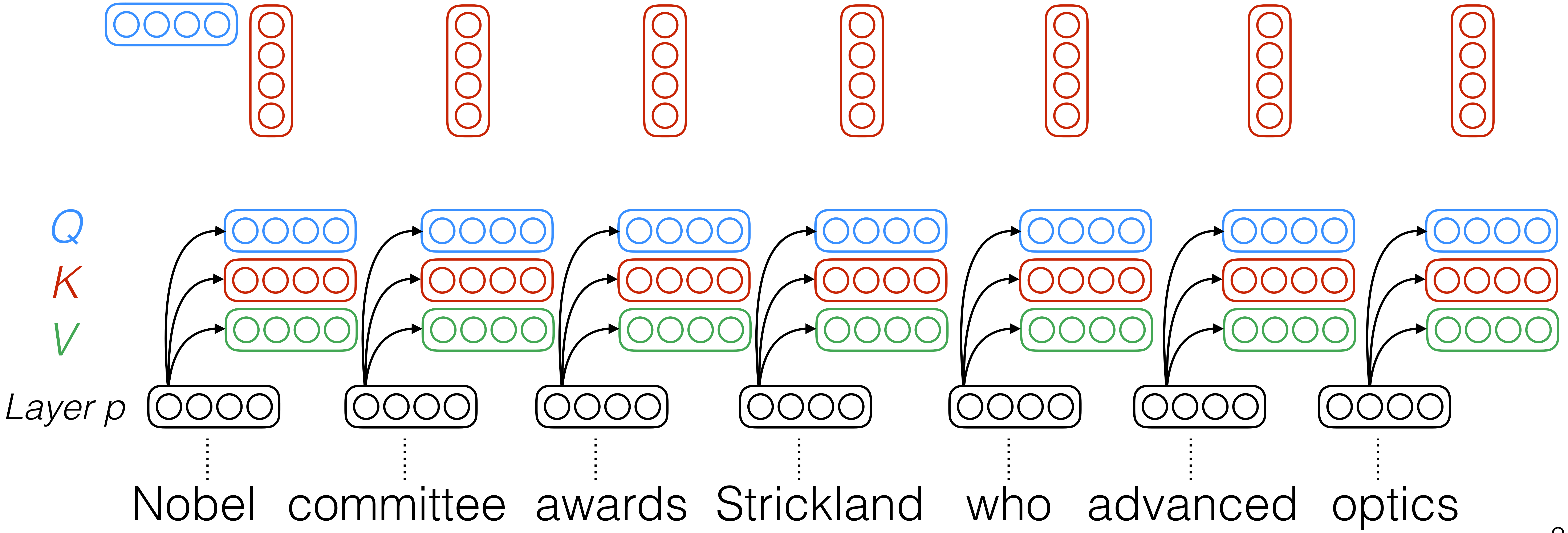
- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
    - Multi-head self-attention [Vaswani et al. 2017]
    - Syntactically-informed self-attention
    - Multi-task learning, single-pass inference
- Experimental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

Self-attention



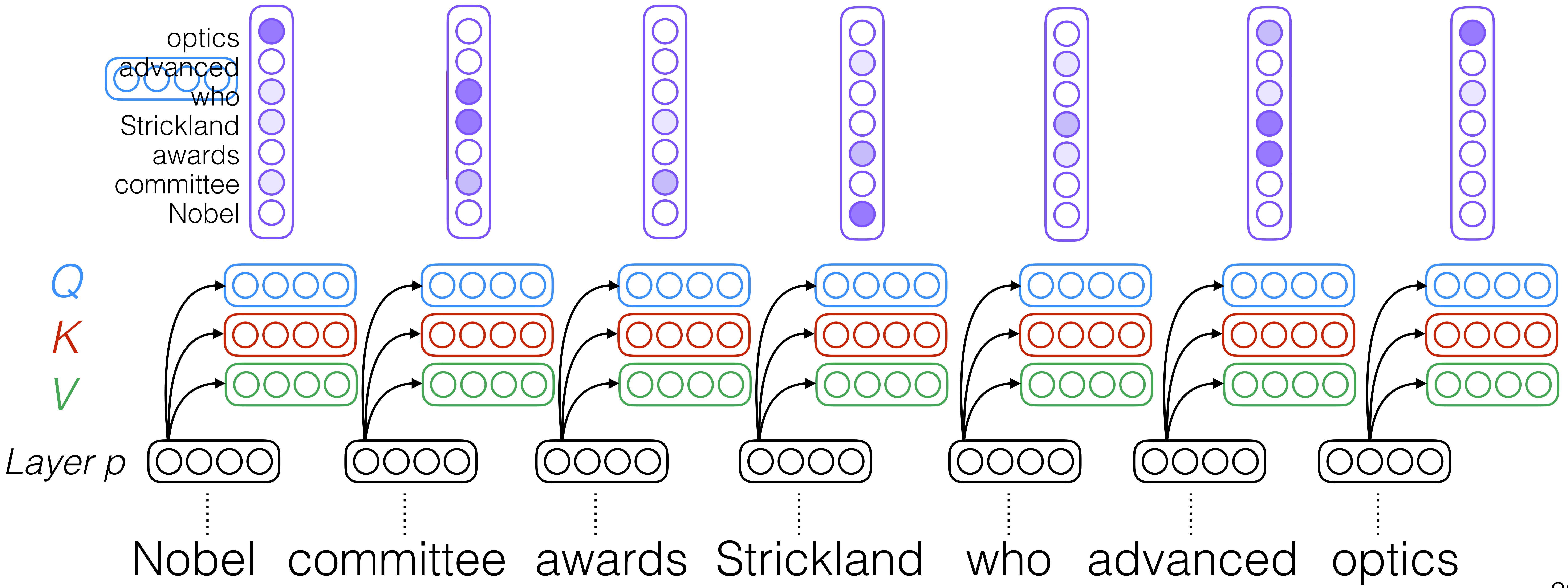
[Vaswani et al. 2017]

Self-attention



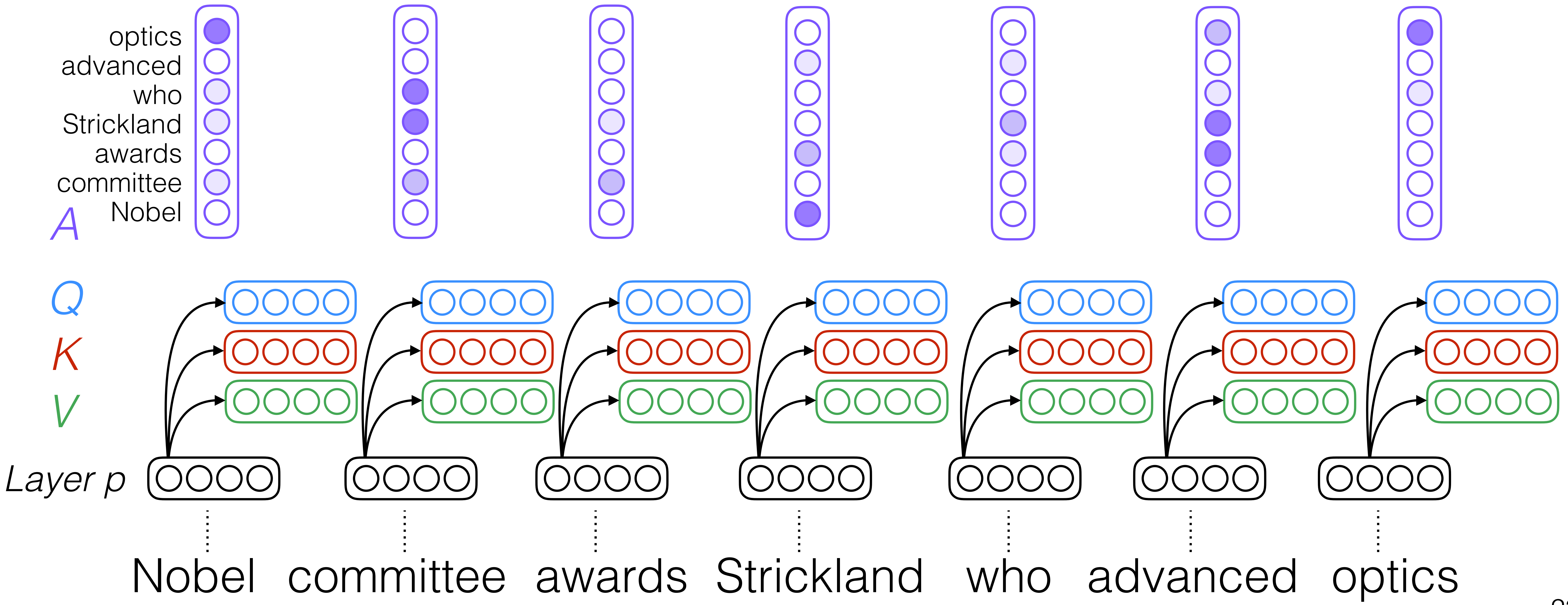
[Vaswani et al. 2017]

Self-attention



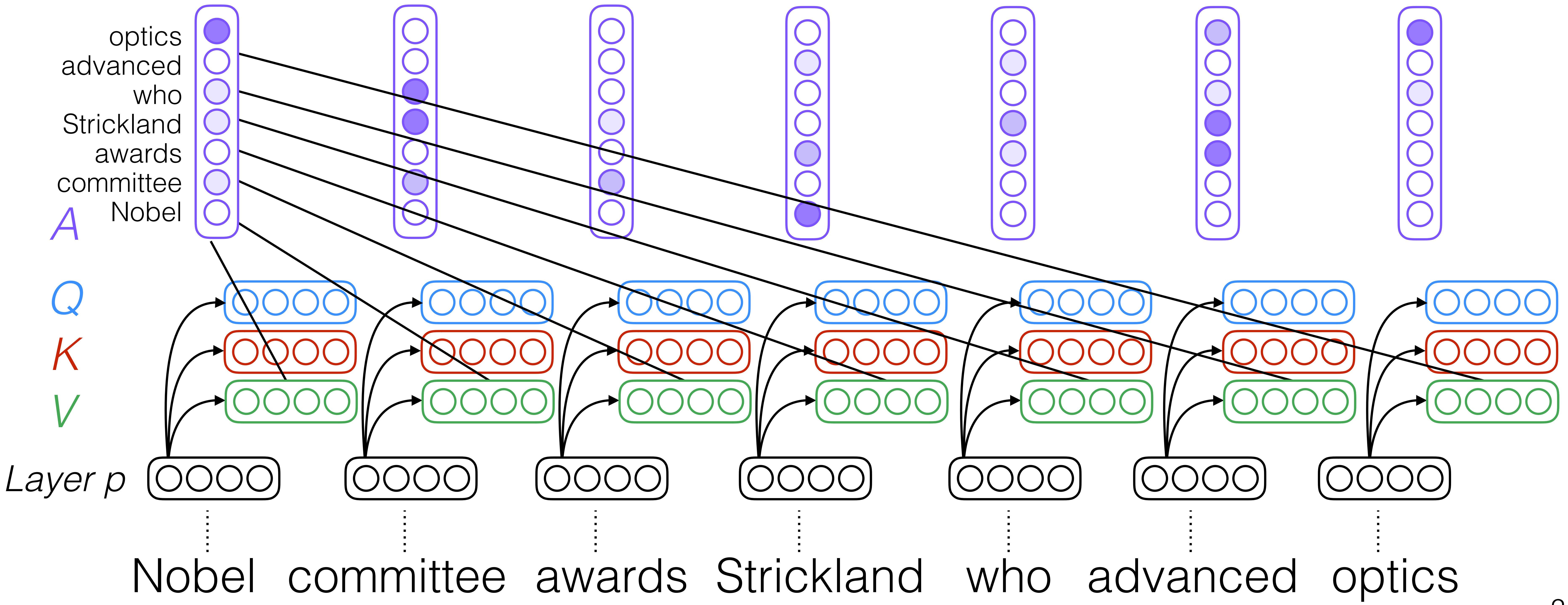
Self-attention

[Vaswani et al. 2017]



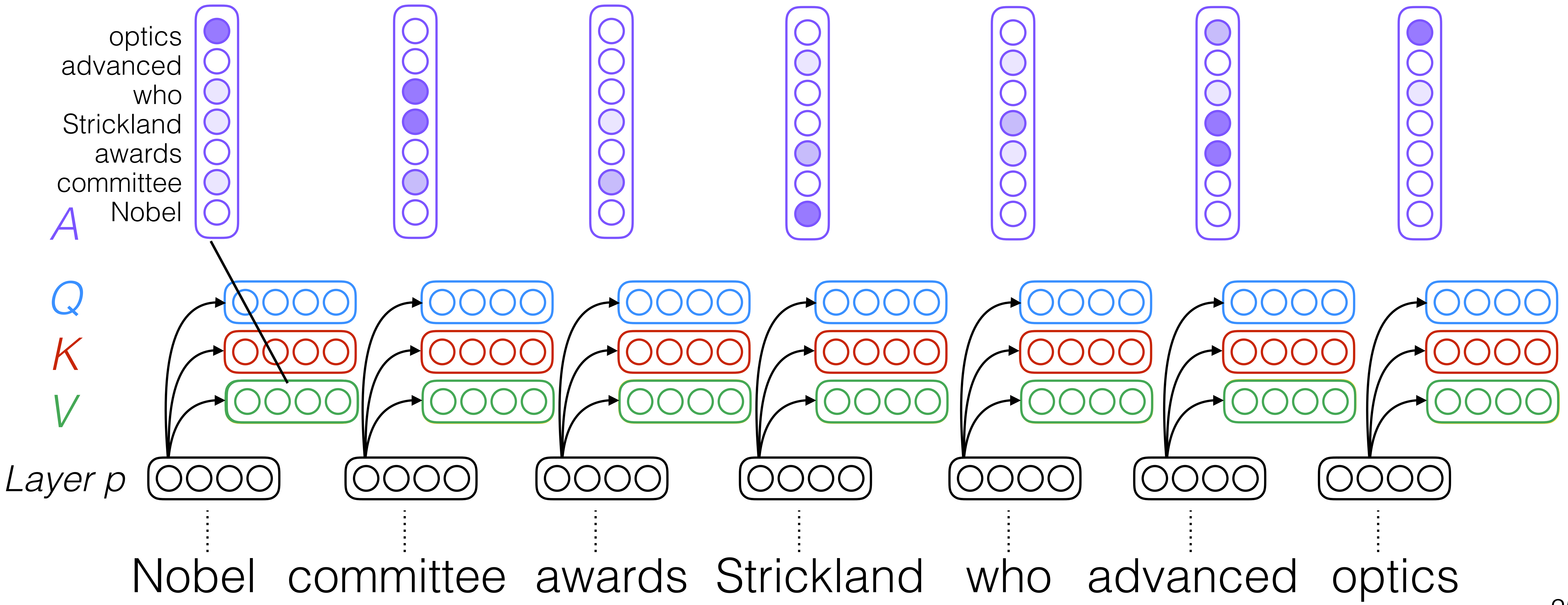
Self-attention

[Vaswani et al. 2017]



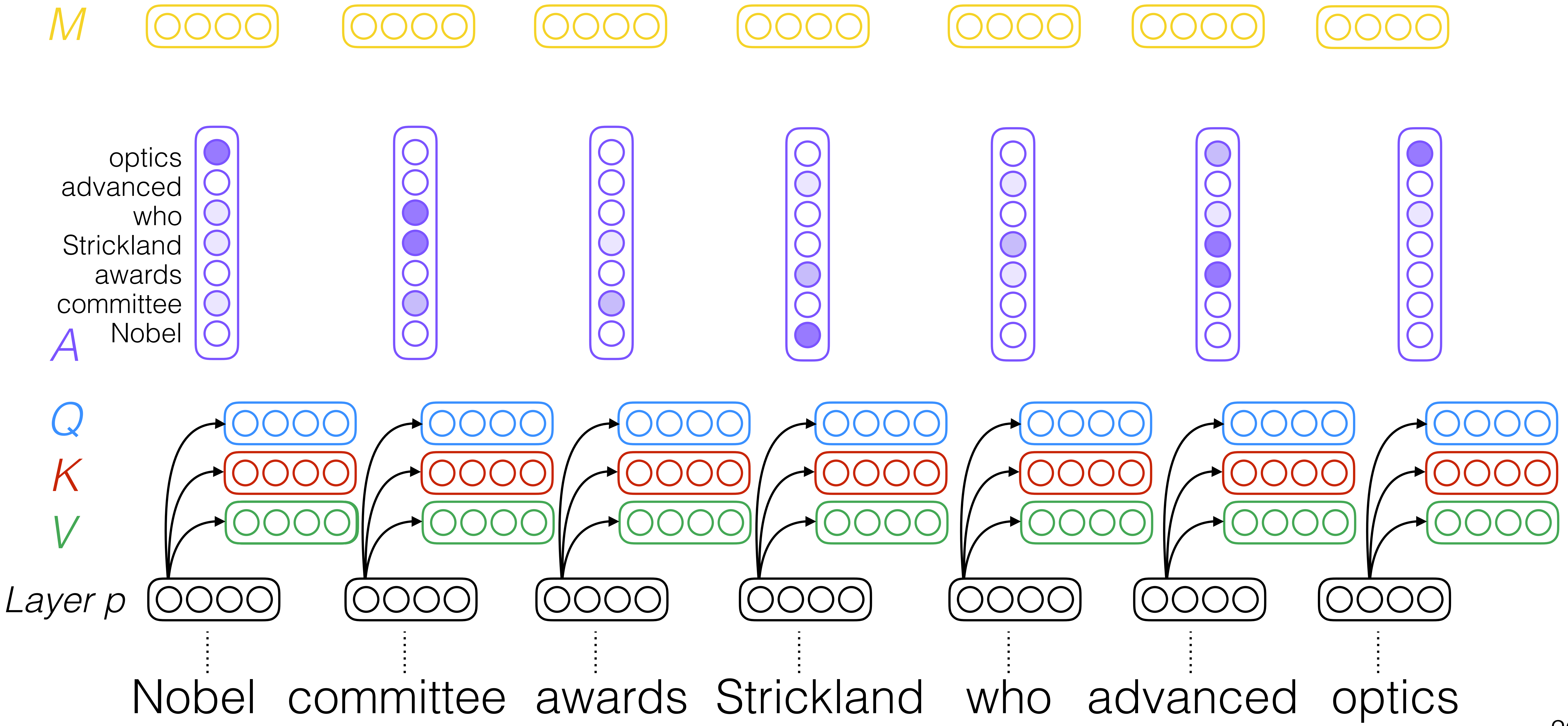
[Vaswani et al. 2017]

Self-attention



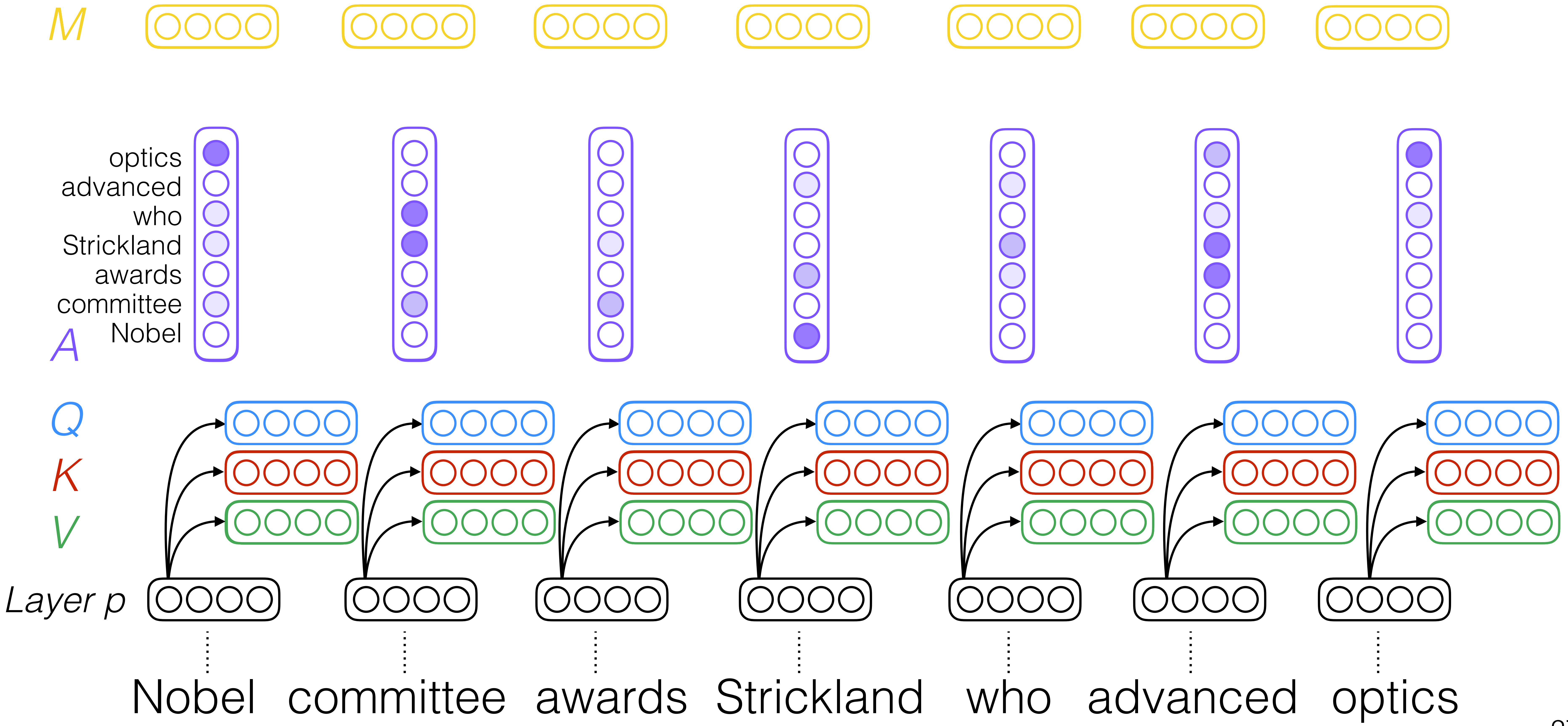
Self-attention

[Vaswani et al. 2017]



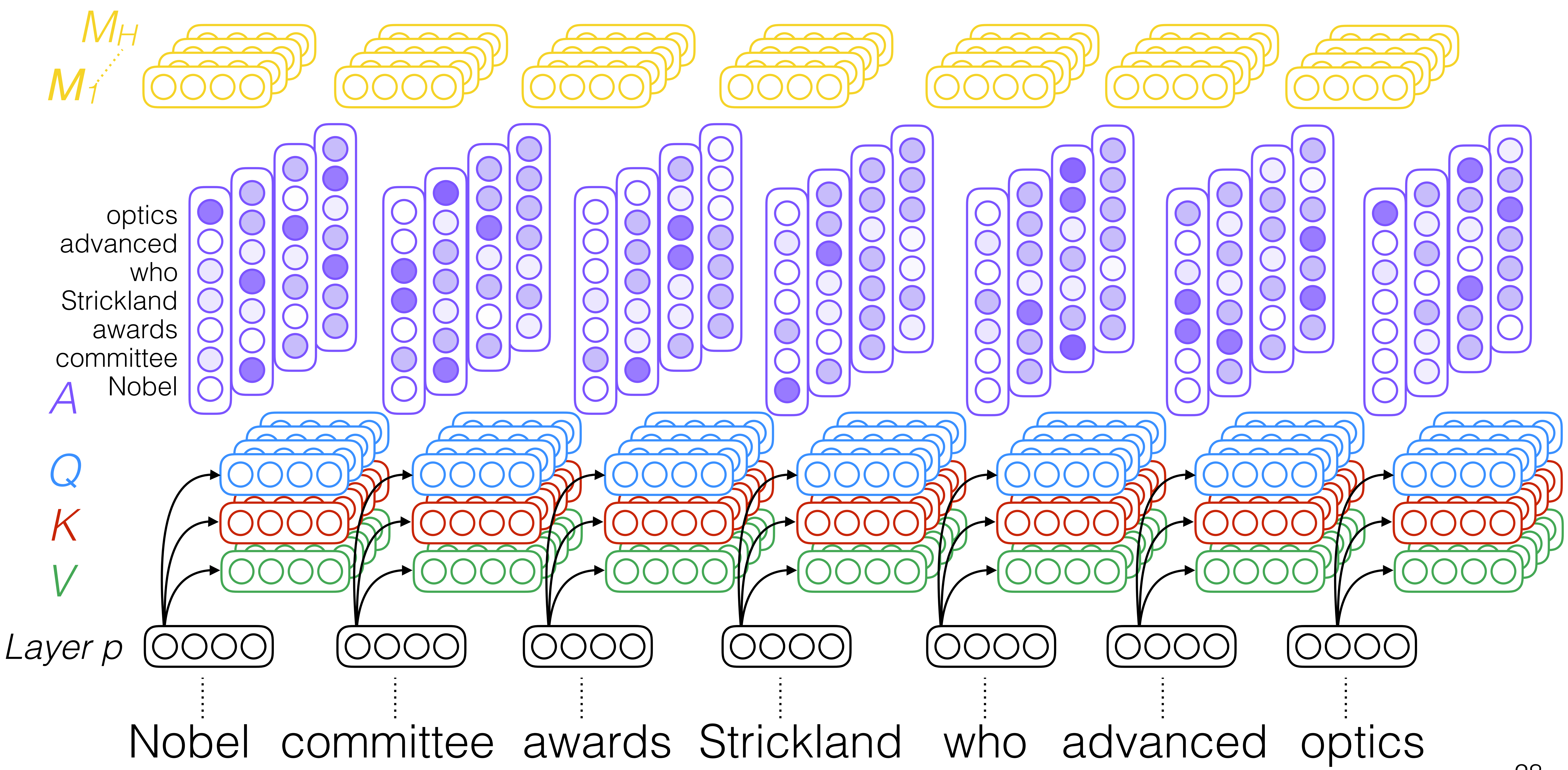
[Vaswani et al. 2017]

Self-attention



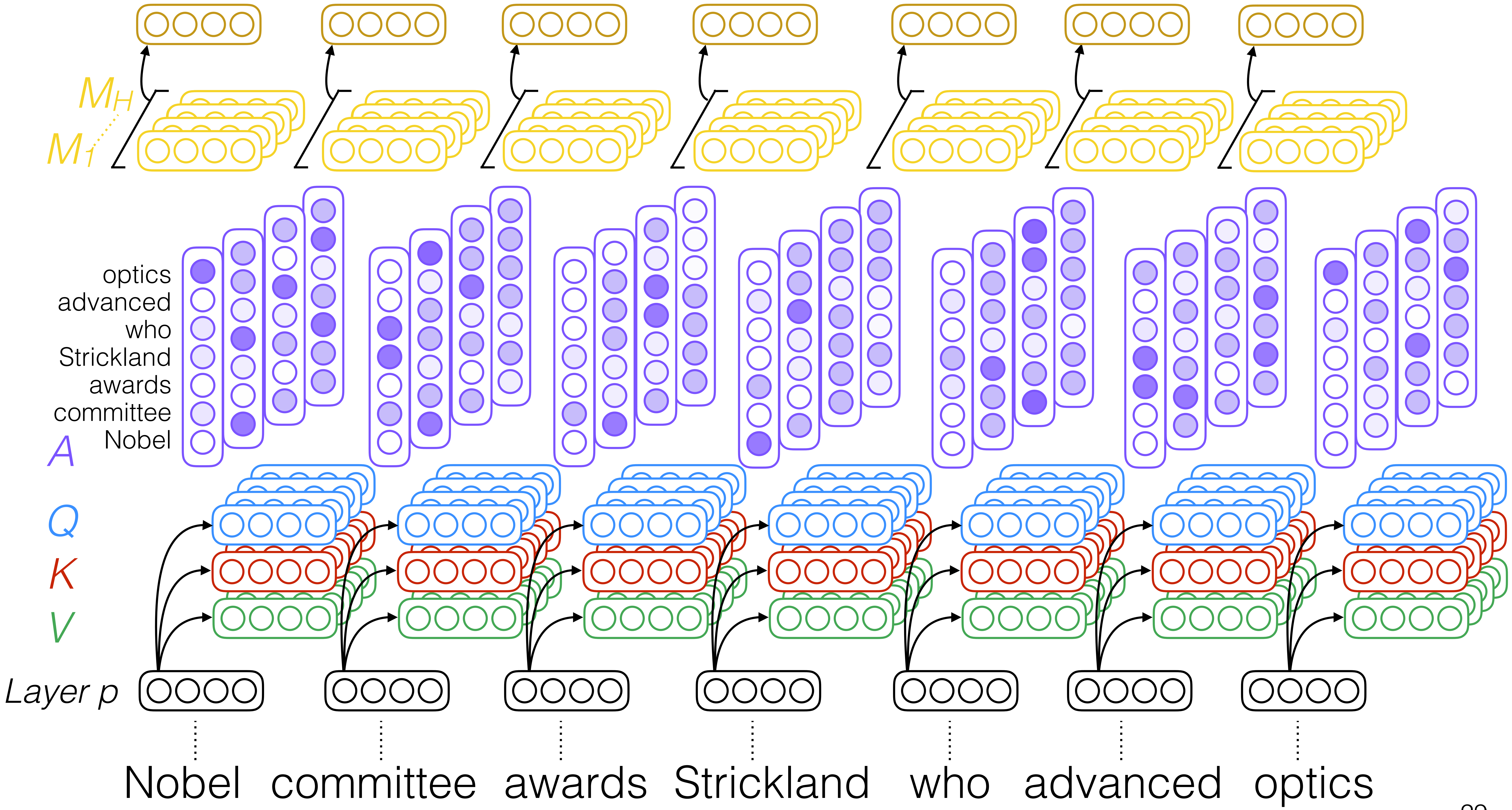
[Vaswani et al. 2017]

Multi-head self-attention



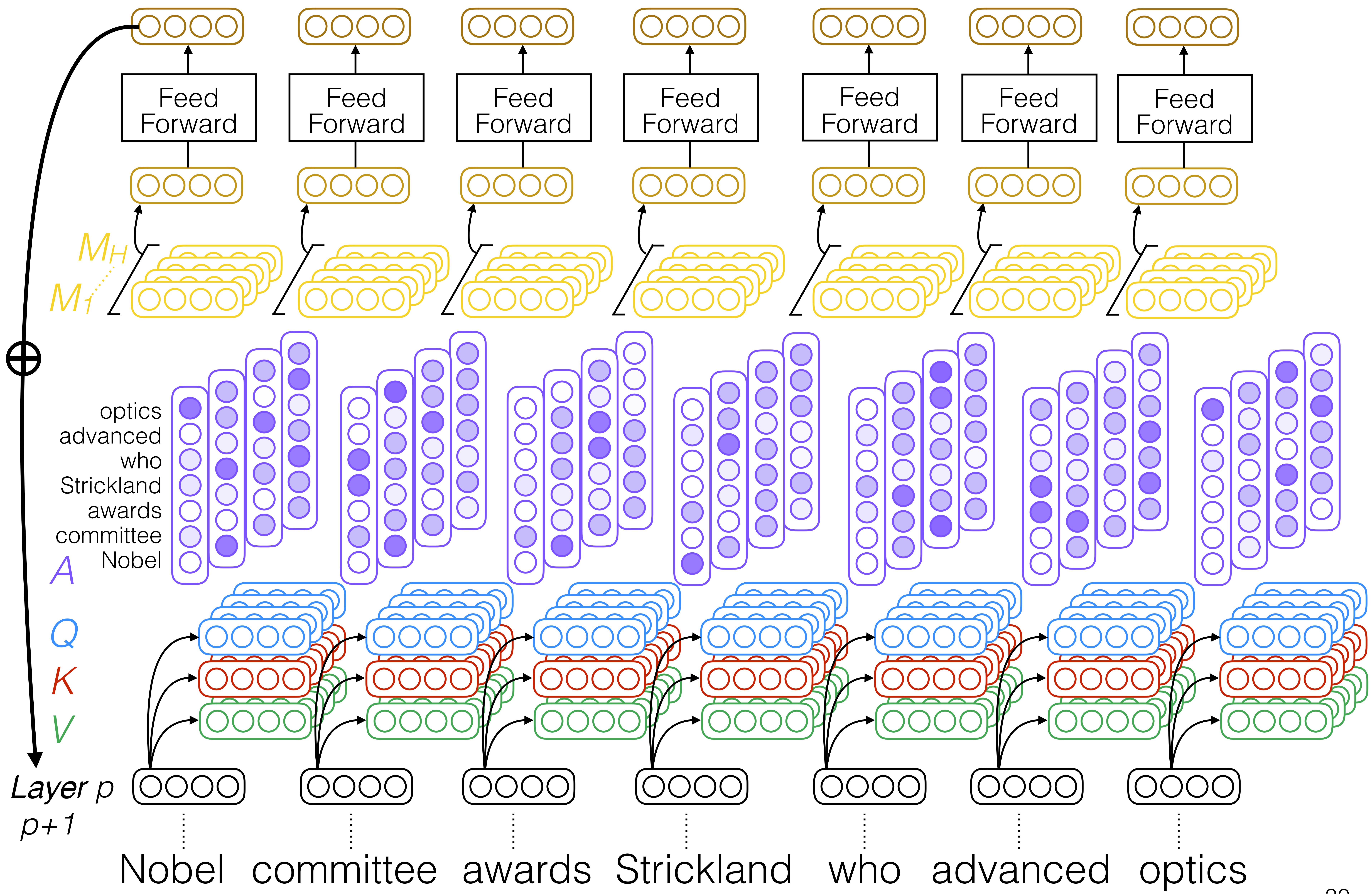
[Vaswani et al. 2017]

Multi-head self-attention

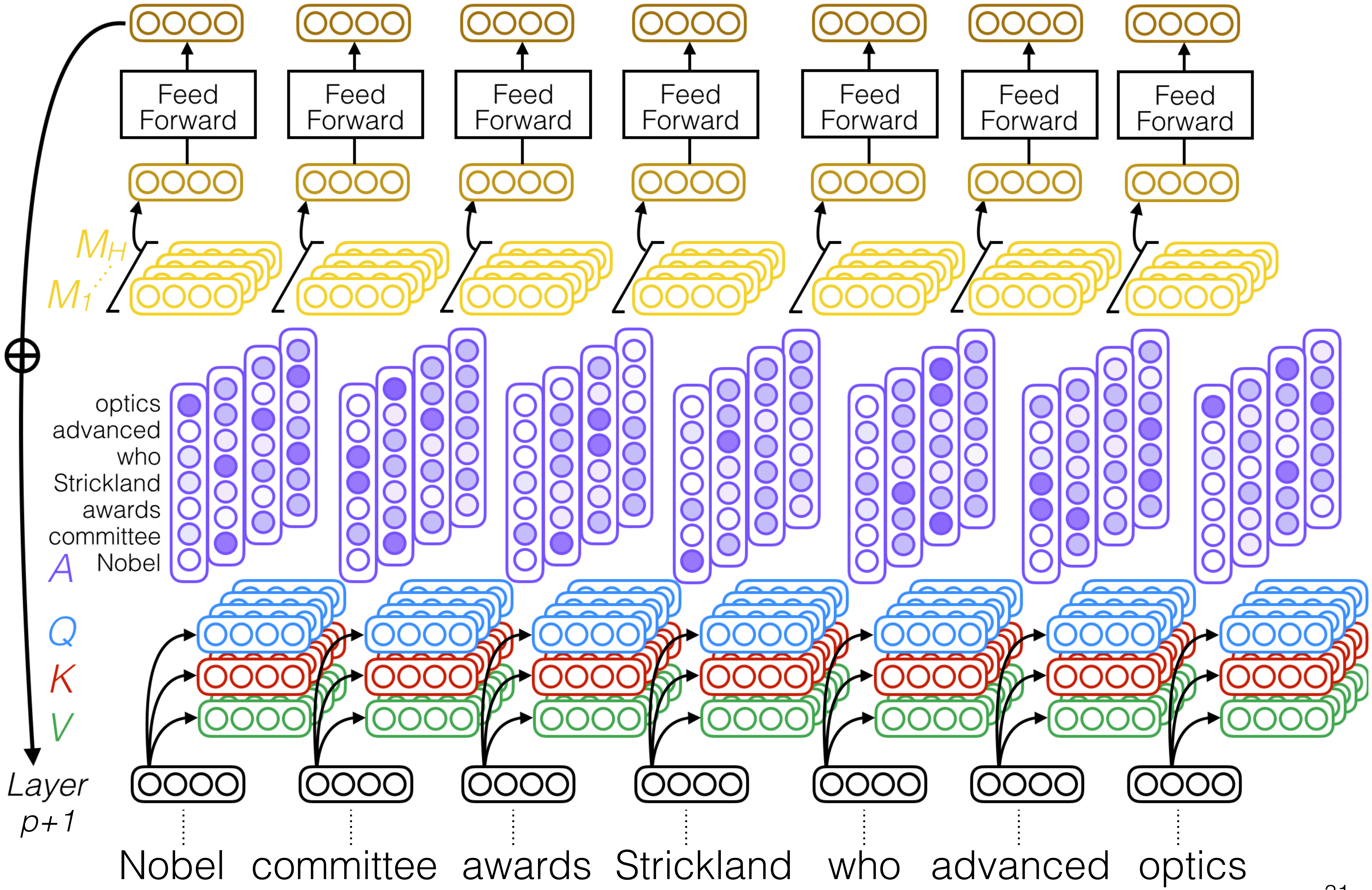


[Vaswani et al. 2017]

Multi-head self-attention

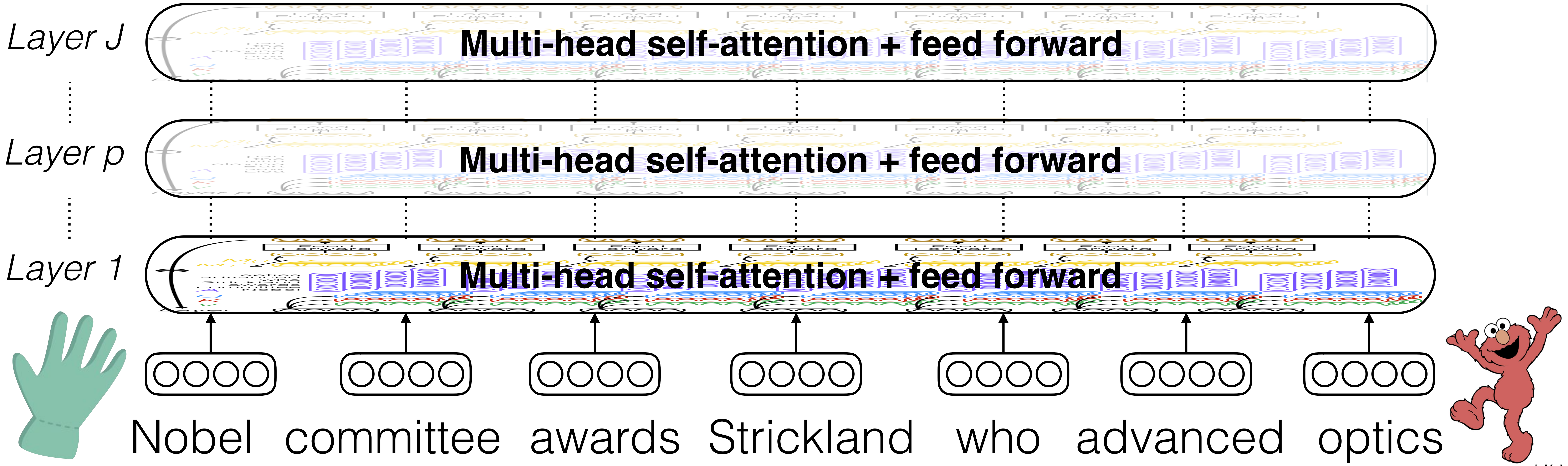


Multi-head self-attention



Multi-head self-attention

[Vaswani et al. 2017]



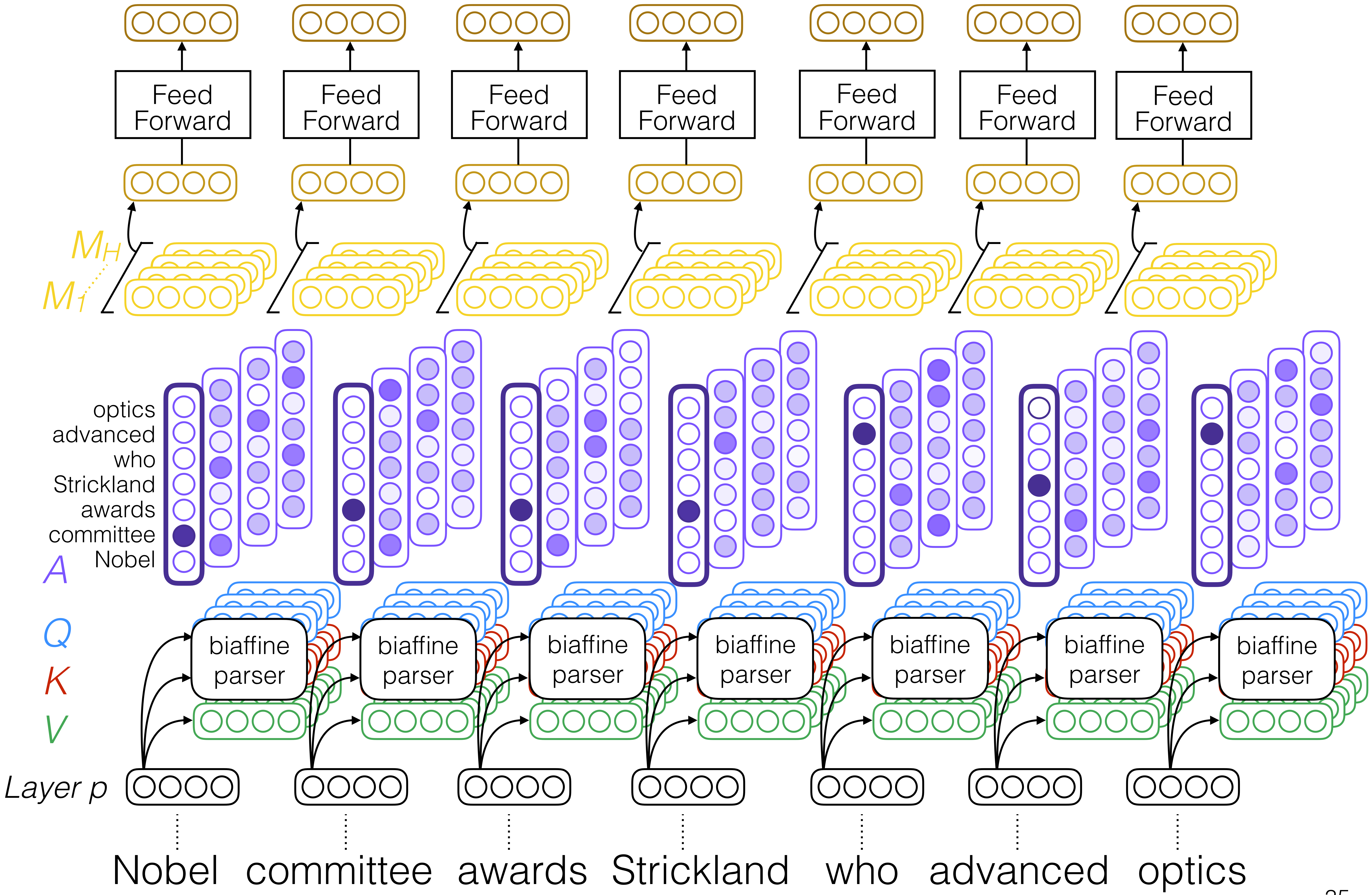
Outline

- 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?

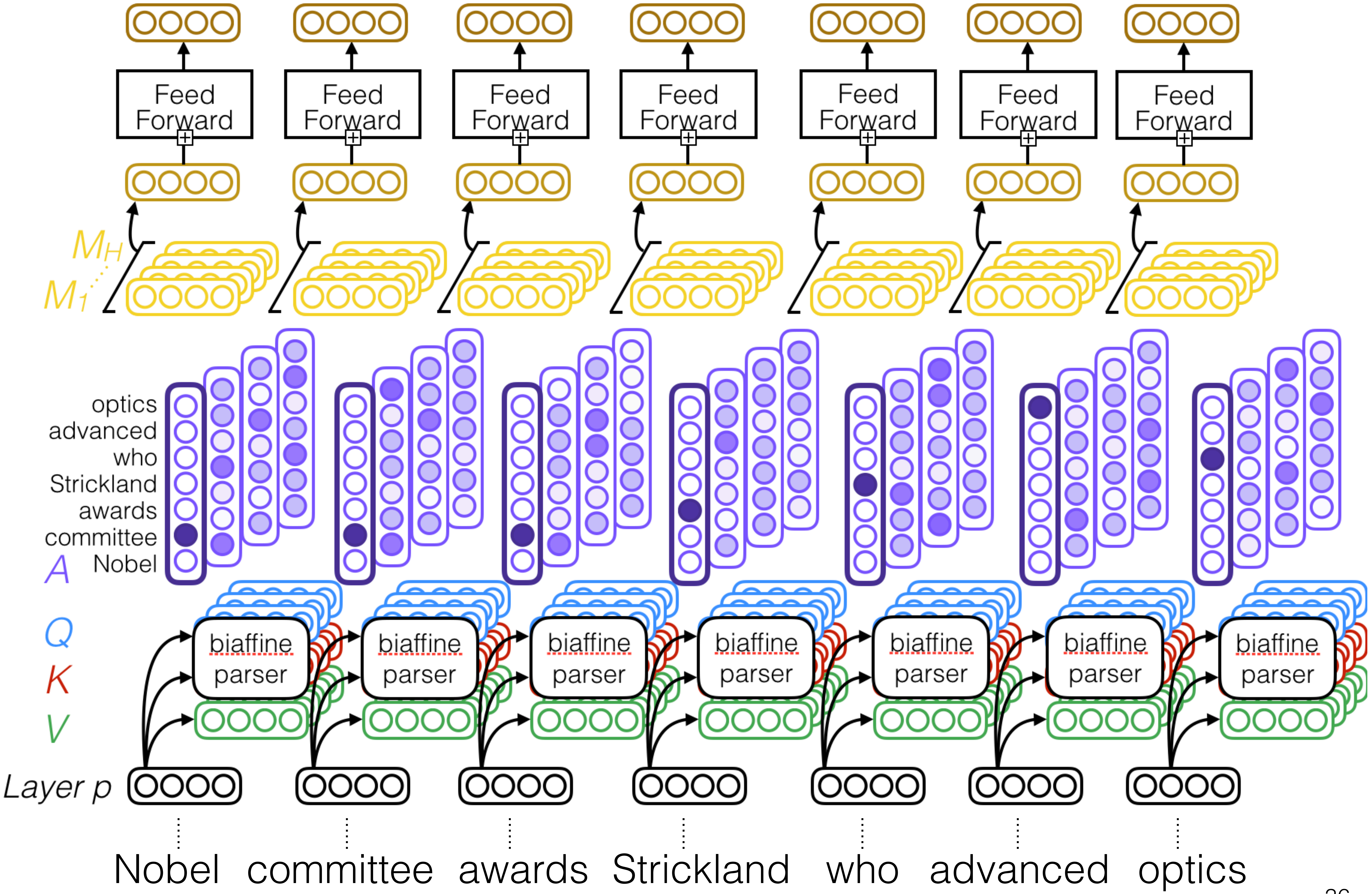
- Multi-task learning [Caruana 1993; Collobert et al. 2011]:
  - Overfits to training domain like single-task end-to-end NN.
  - Must re-train SRL model to leverage new (improved) syntax.
- Dependency path embeddings [Roth & Lapata 2016]; Graph CNN over parse [Marcheggiani & Titov 2017]
  - Restricted 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

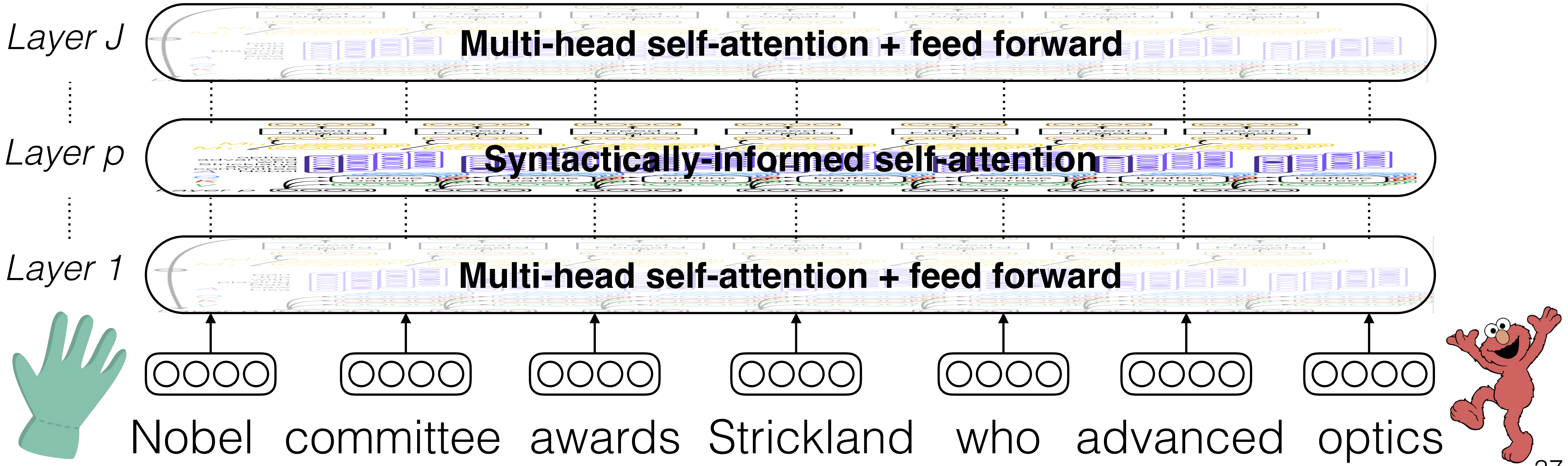


[Dozat and Manning 2017]

Syntactically-informed self-attention



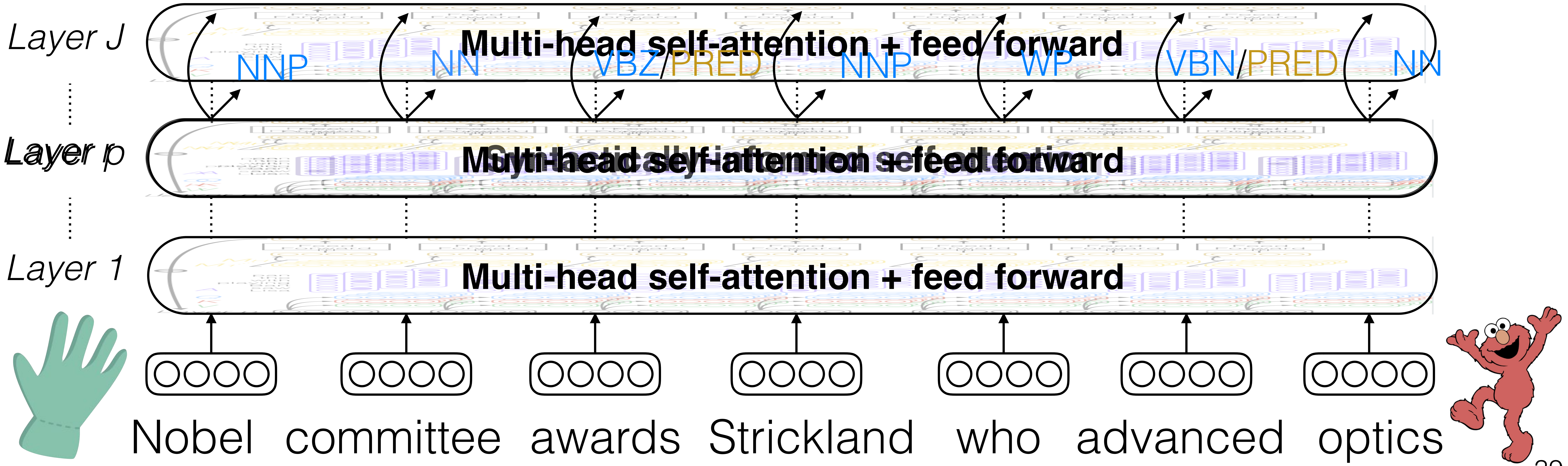
Syntactically-informed self-attention



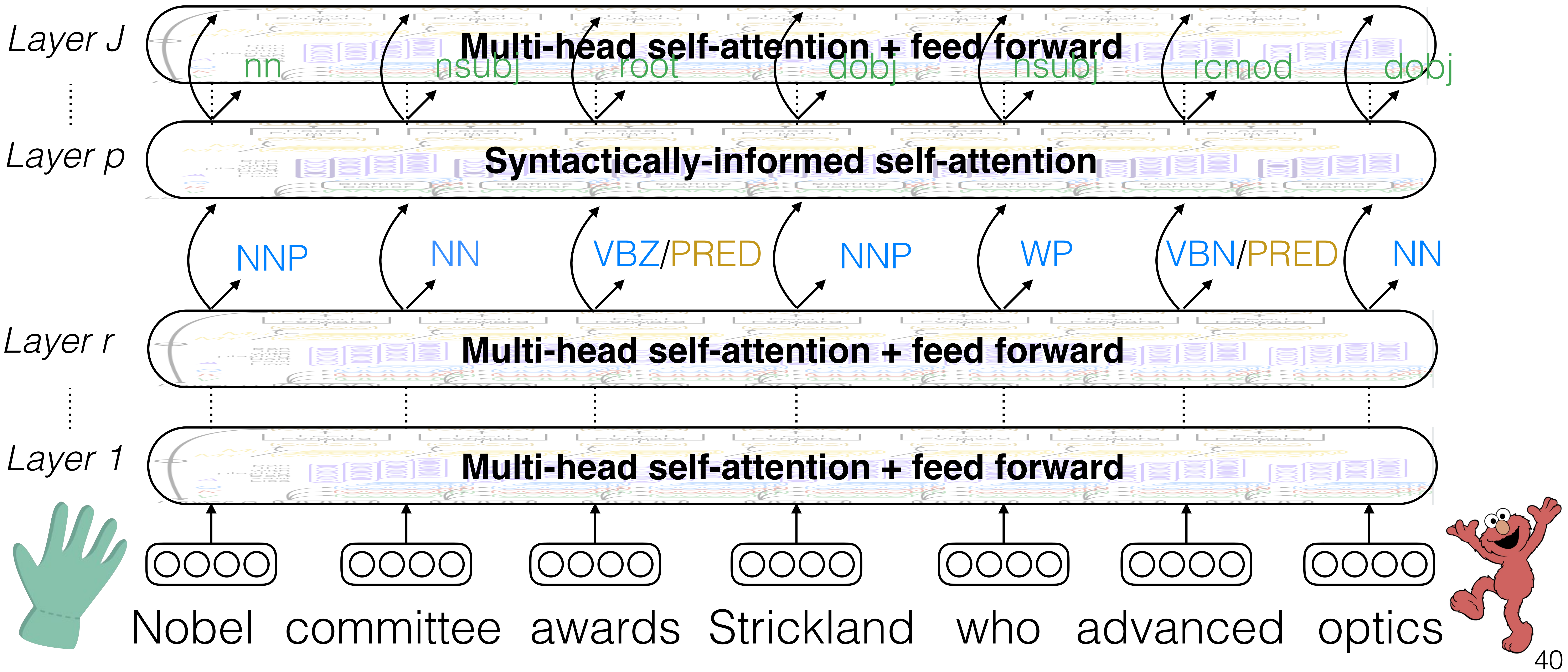
Outline

- 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

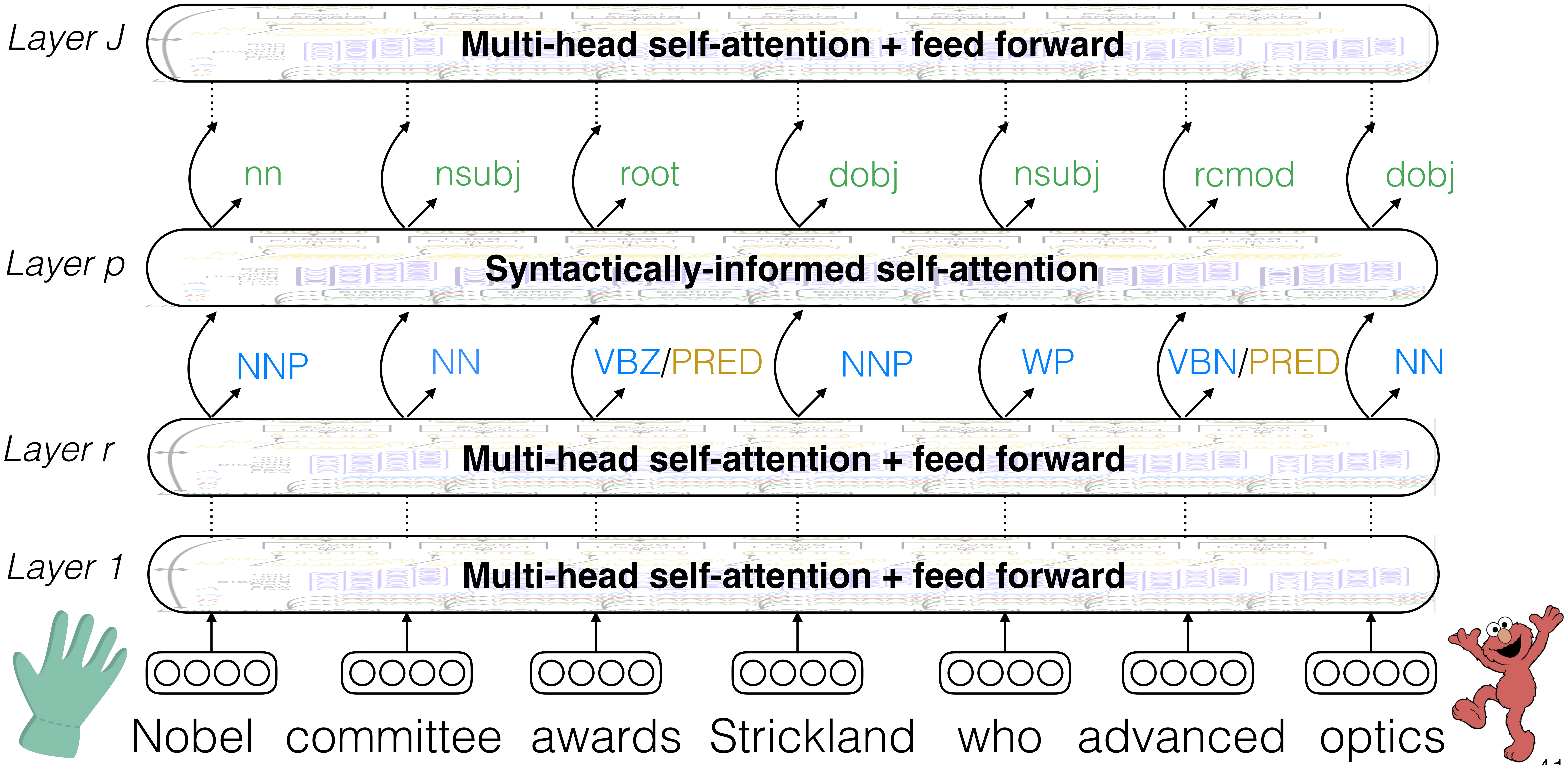
LISA: Linguistically-Informed Self-Attention



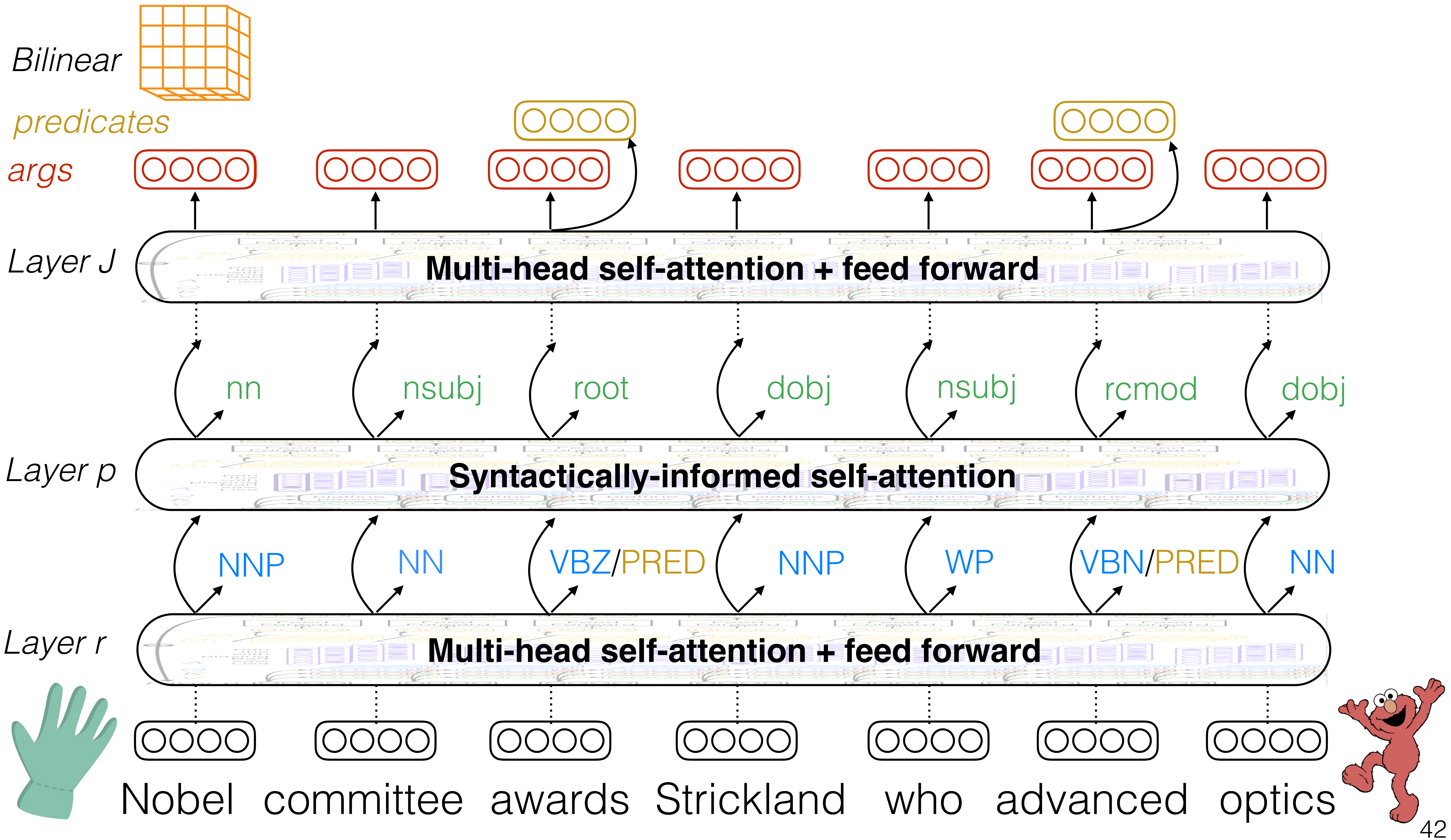
LISA: Linguistically-Informed Self-Attention



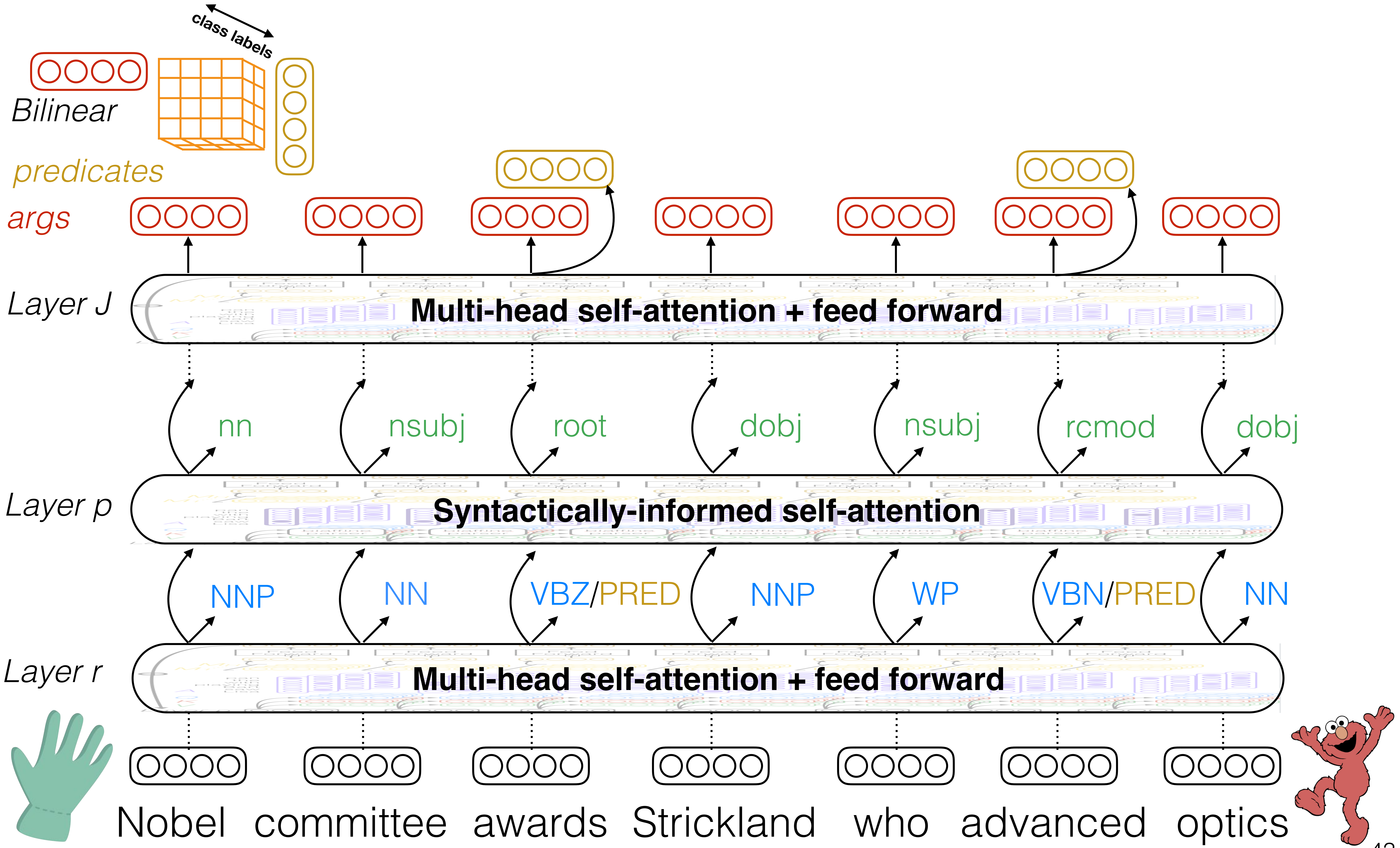
LISA: Linguistically-Informed Self-Attention



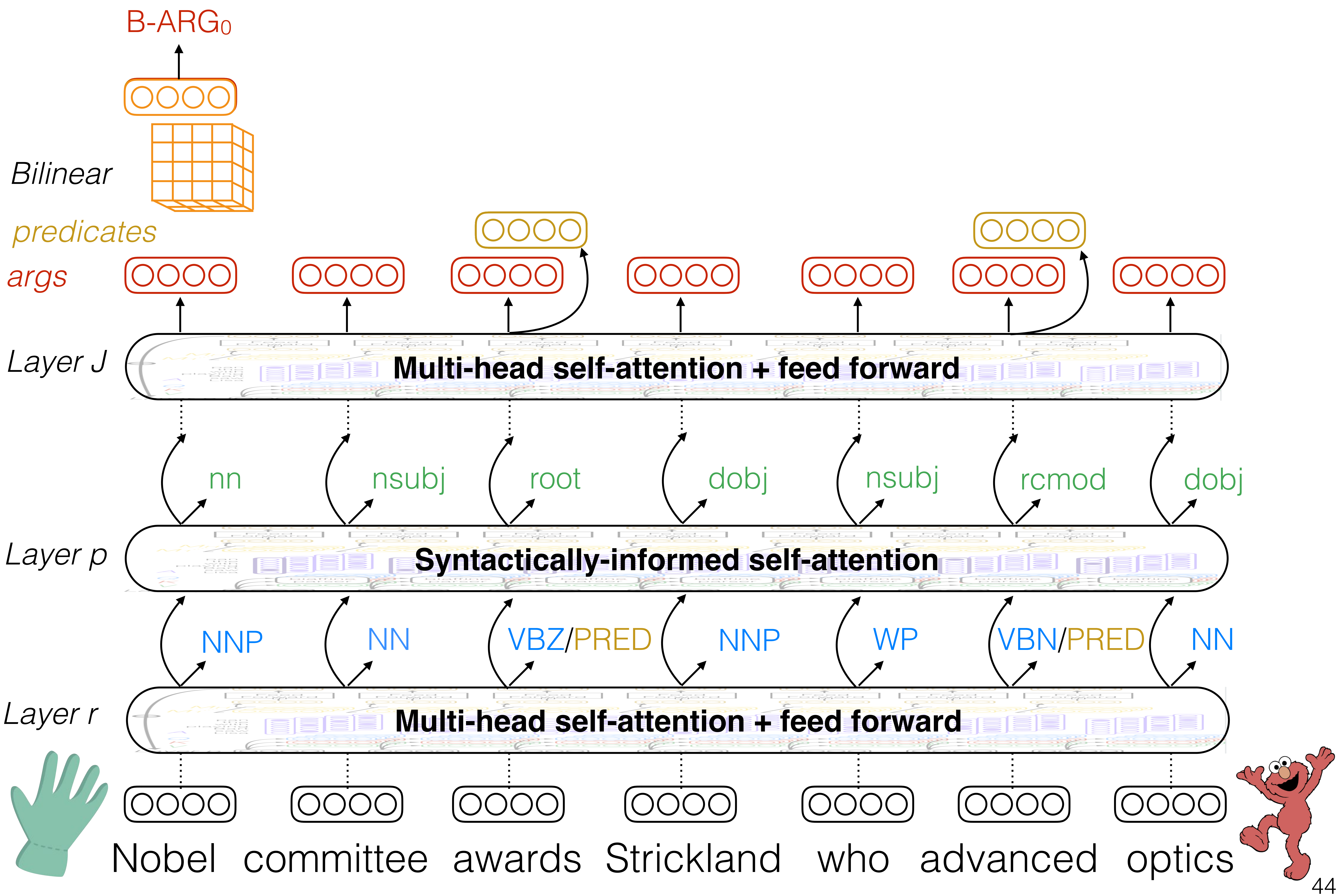
LISA: Linguistically-Informed Self-Attention



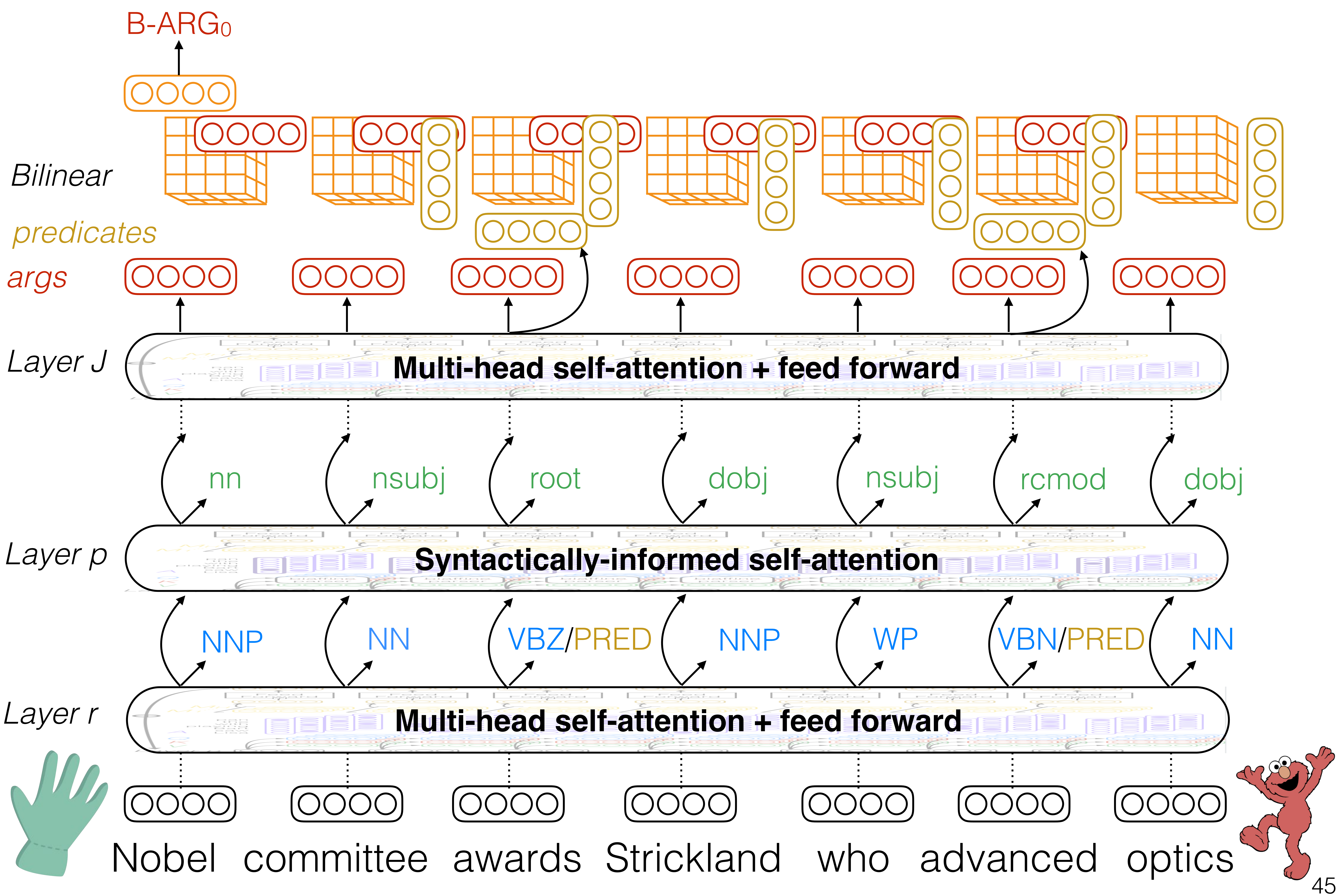
LISA: Linguistically-Informed Self-Attention



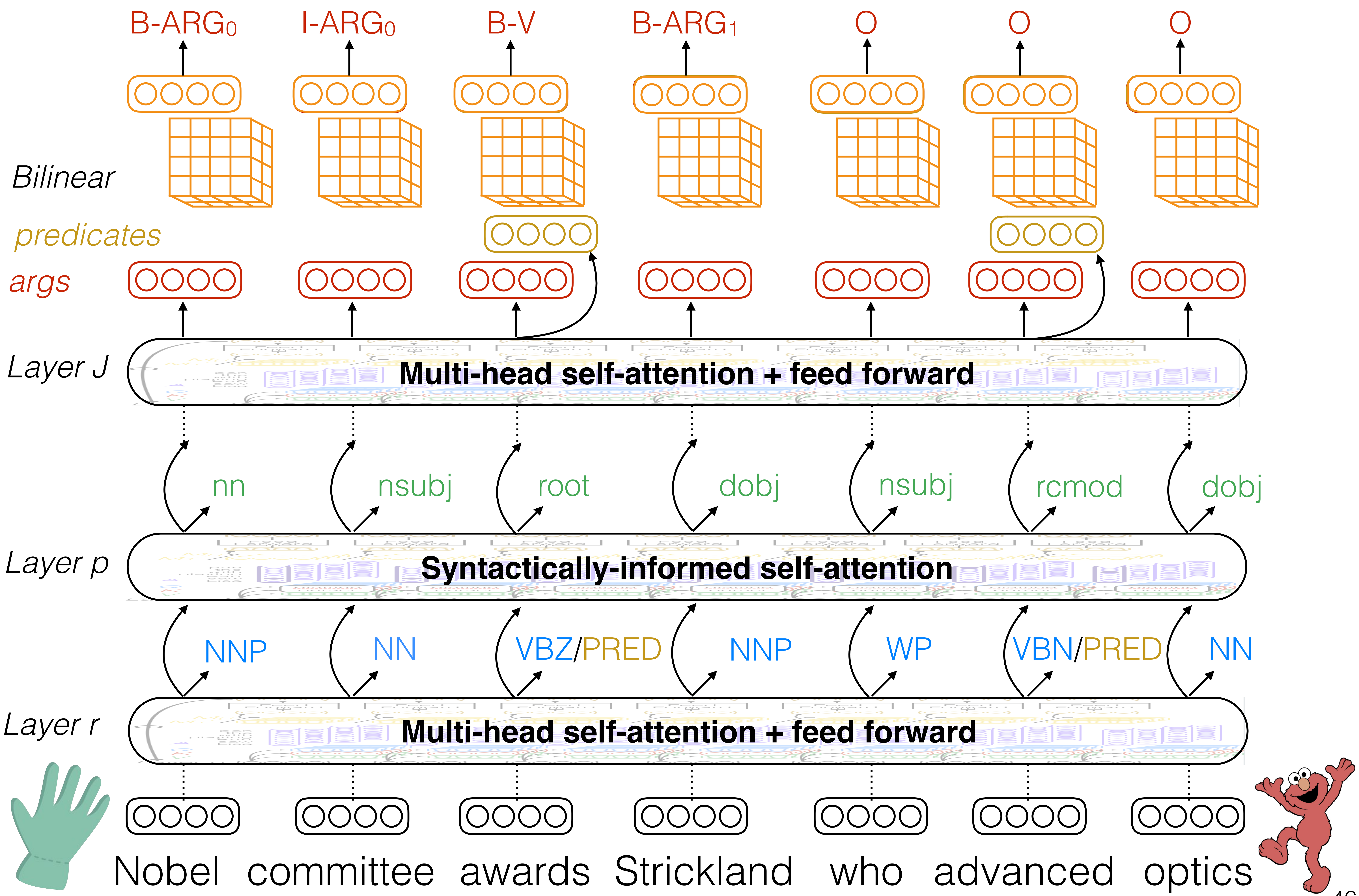
LISA: Linguistically-Informed Self-Attention



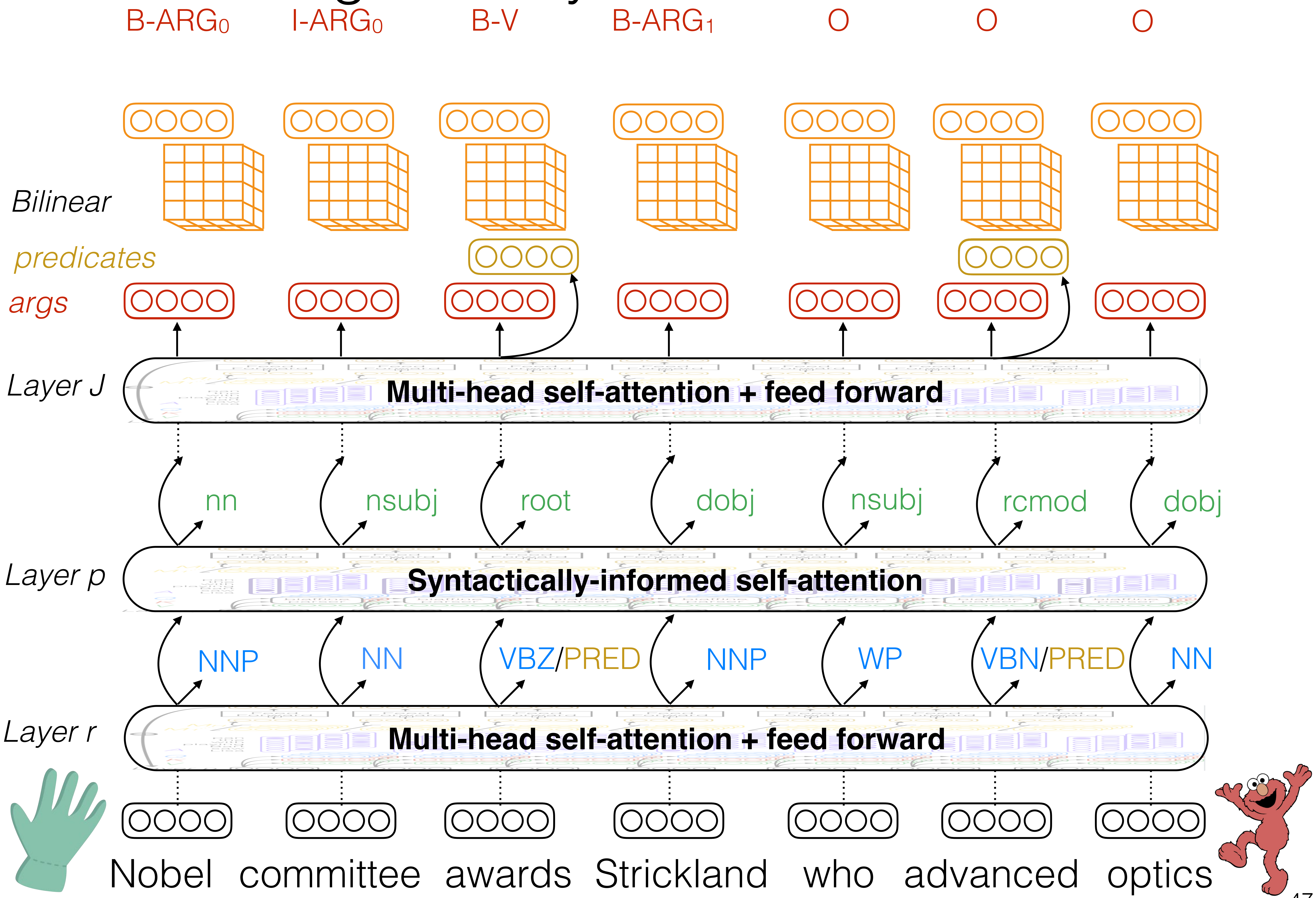
LISA: Linguistically-Informed Self-Attention



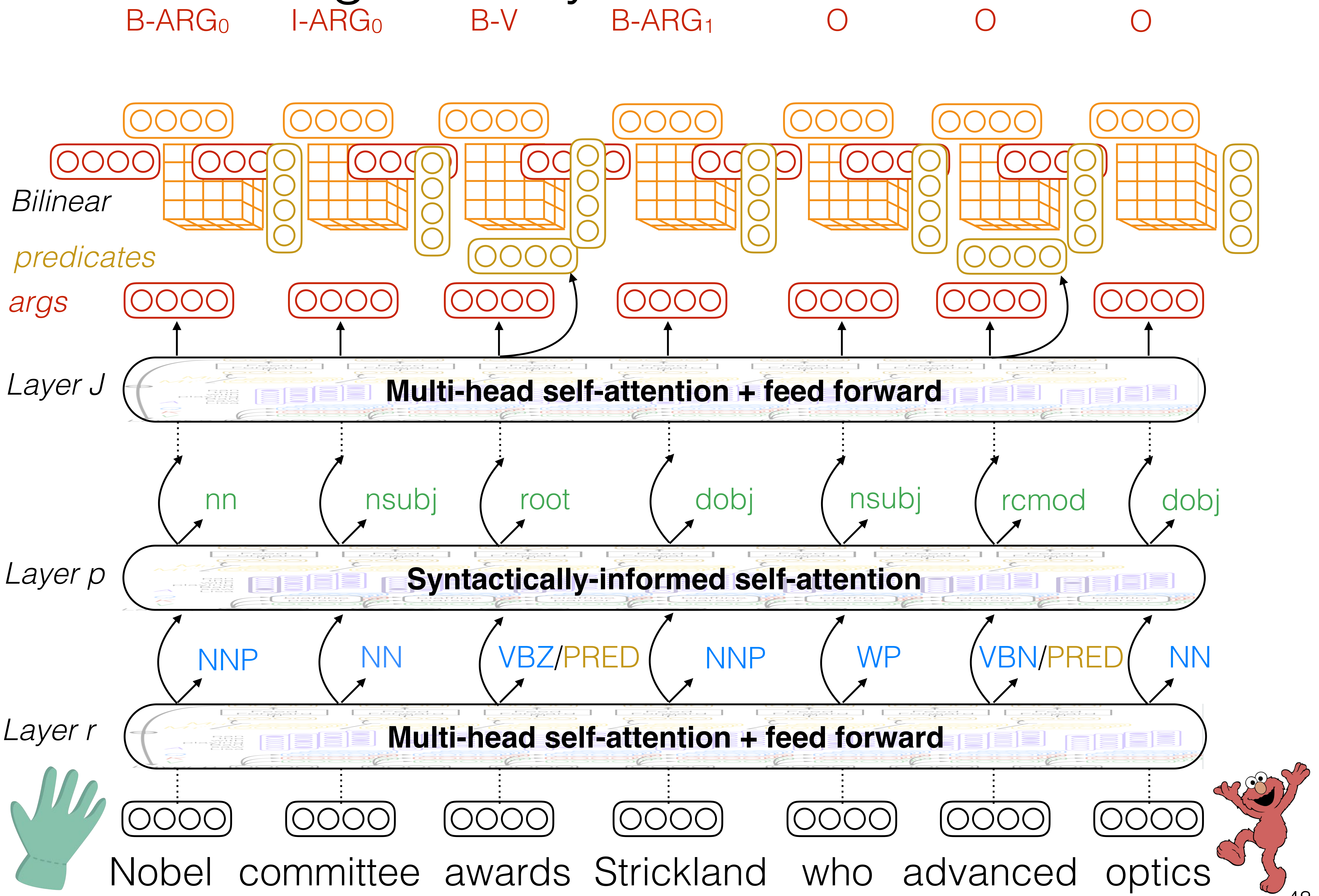
LISA: Linguistically-Informed Self-Attention



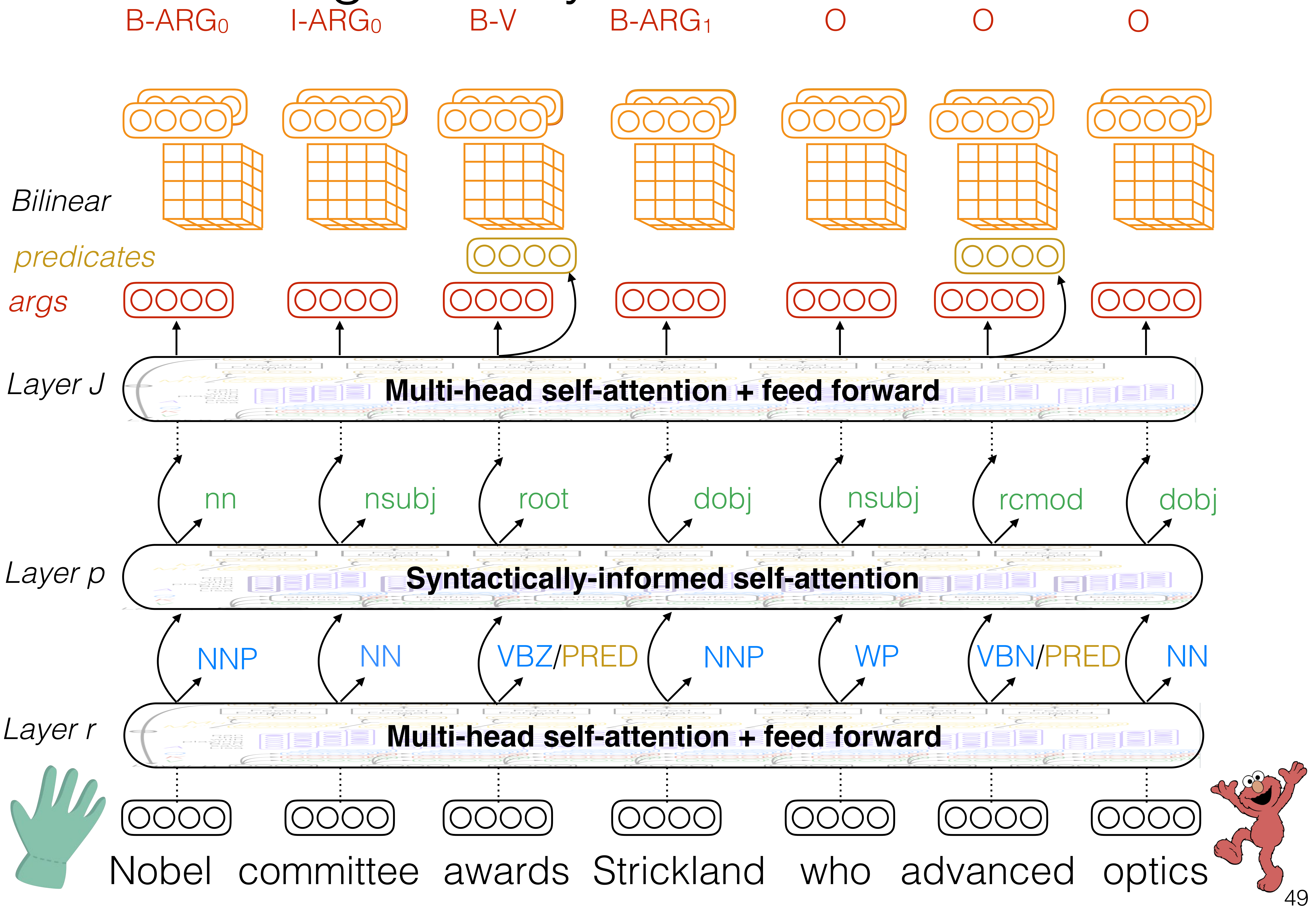
LISA: Linguistically-Informed Self-Attention



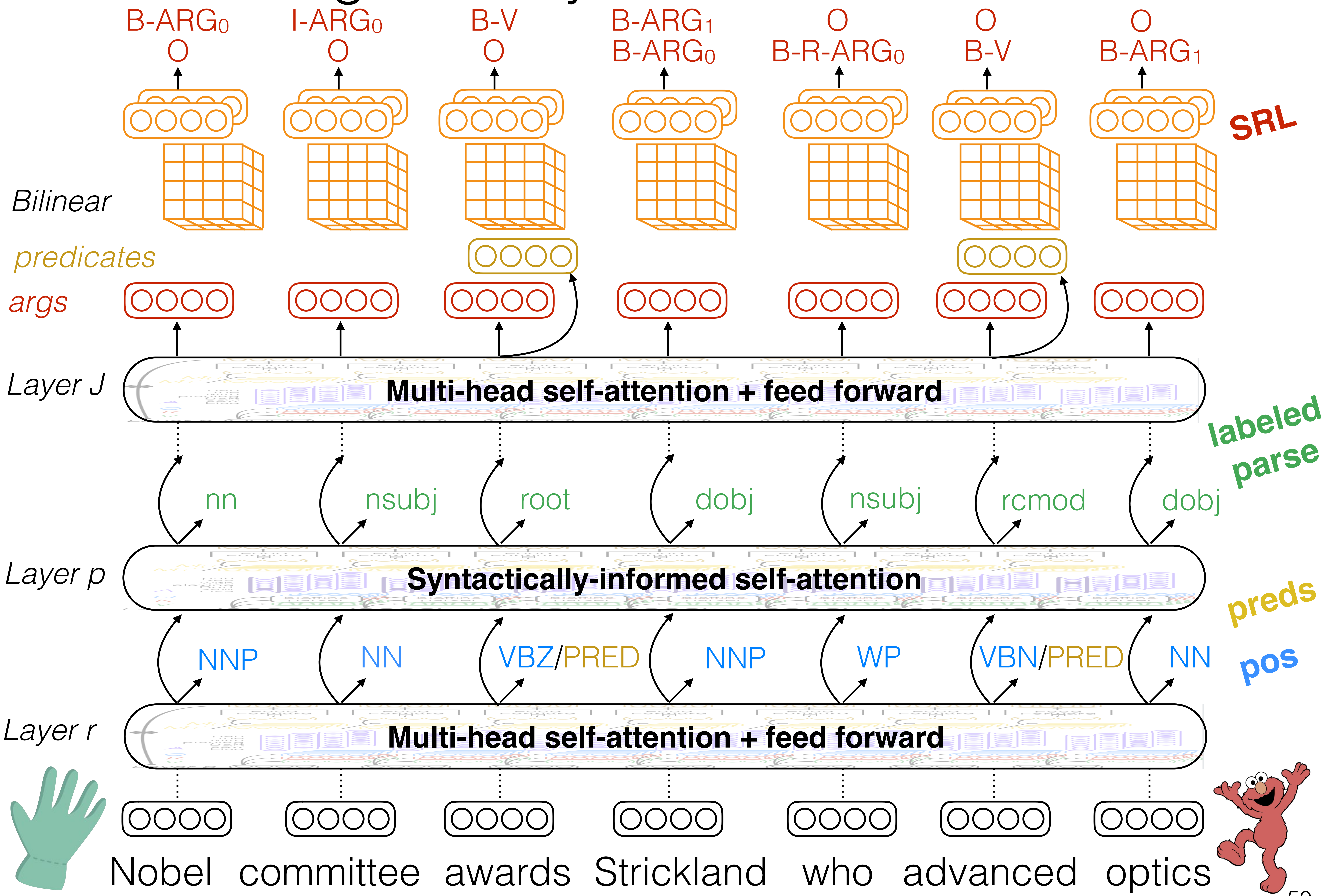
LISA: Linguistically-Informed Self-Attention



LISA: Linguistically-Informed Self-Attention



LISA: Linguistically-Informed Self-Attention



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

Experimental results: CoNLL-2005

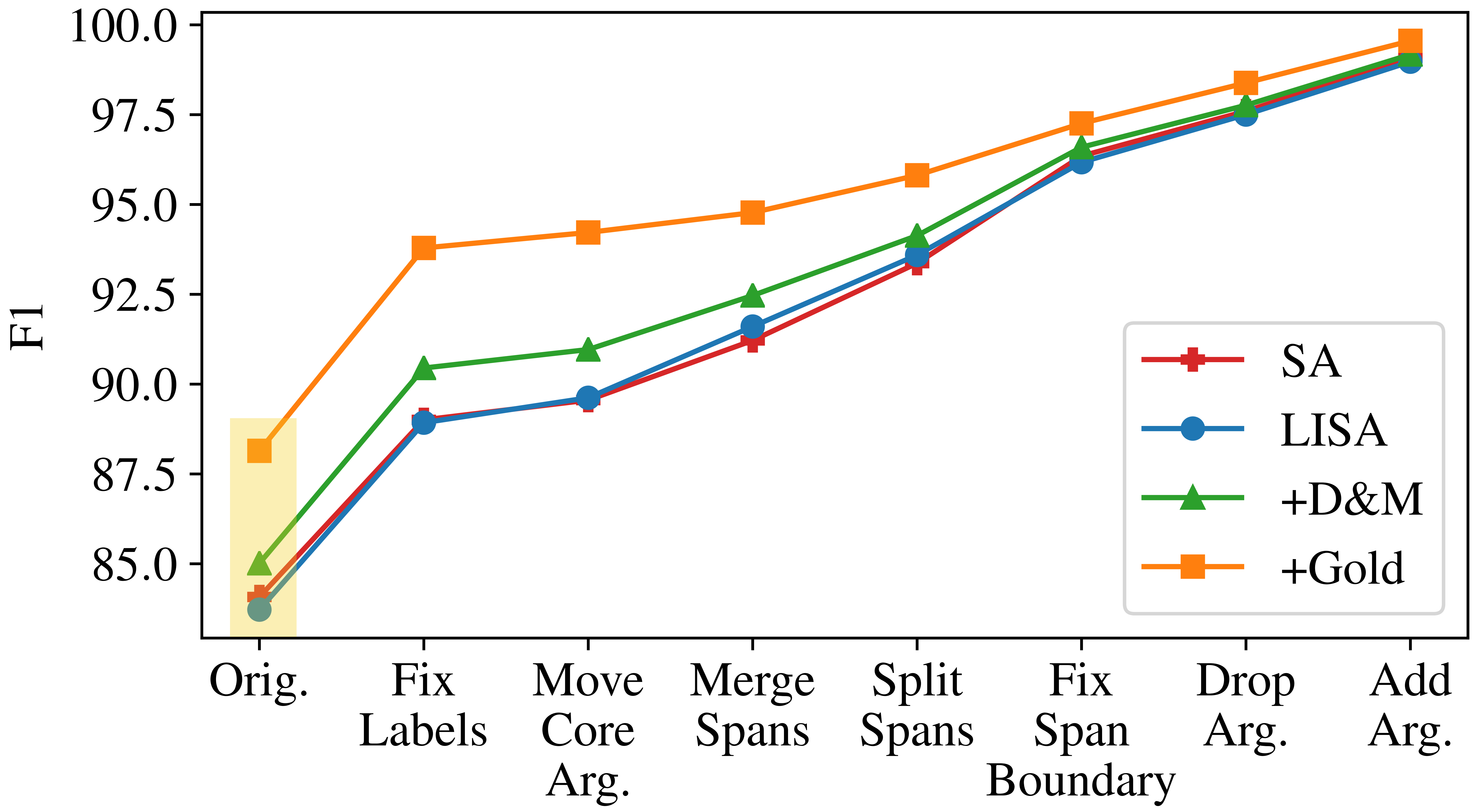
	GloVe		ELMo	
	in-domain	**out-of-domain**	**in-domain**	**out-of-domain**
He et al. 2017	82.7	70.1	---	---
He et al. 2018	82.5	70.8	86.0	76.1
SA	83.72	71.51	86.09	76.35
LISA	83.61	71.91	86.55	78.05
+D&M	84.99	74.66	86.90	78.25
	94.9 UAS	90.3 UAS	96.3 UAS	93.4 UAS
	+2.49 F1	+3.86 F1	+0.9 F1	+2.15 F1
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

Experimental results: CoNLL-2005

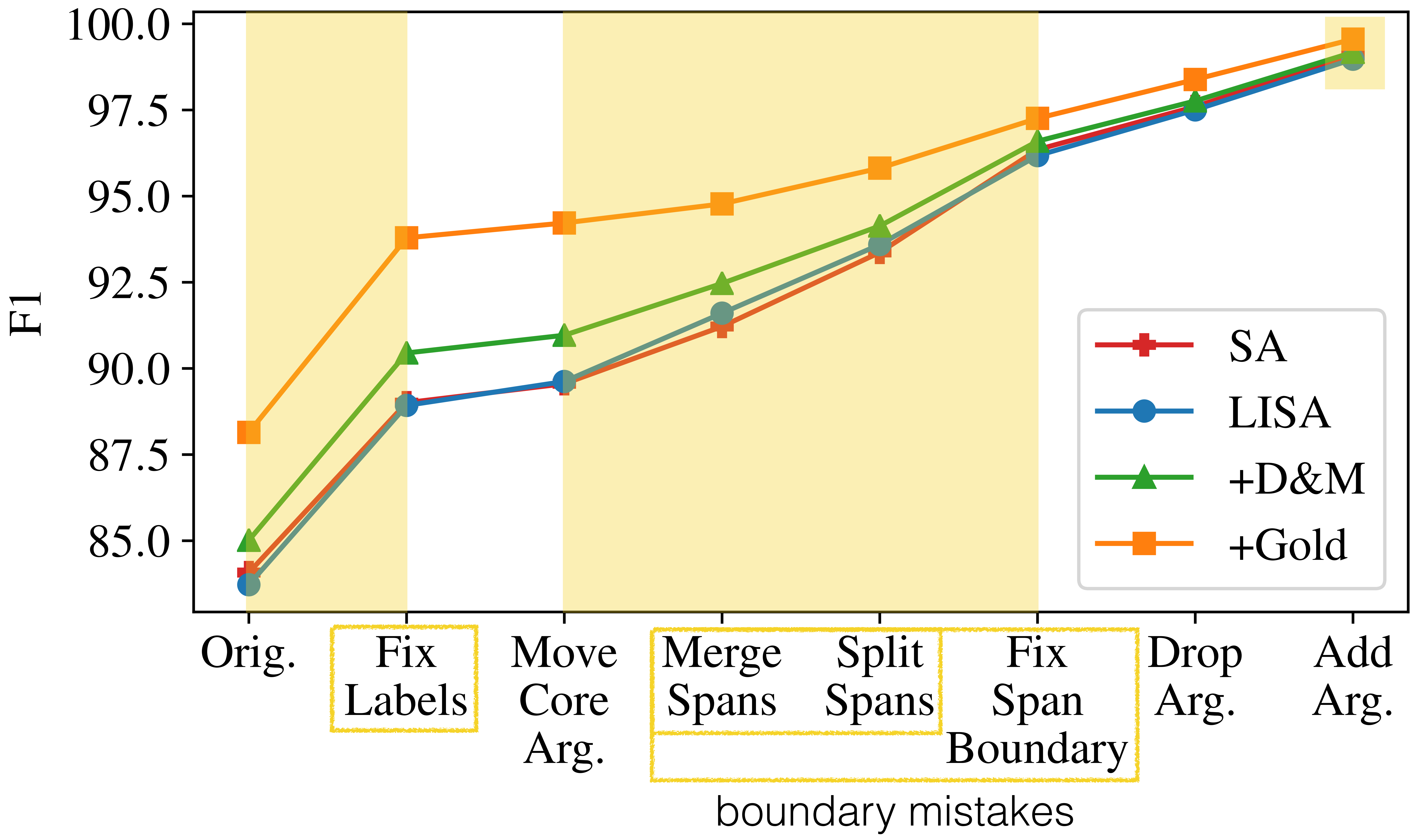
	GloVe	ELMo 
	in-domain (dev)	in-domain (dev)

96.5 UAS!

Experimental results: Analysis



Experimental results: Analysis

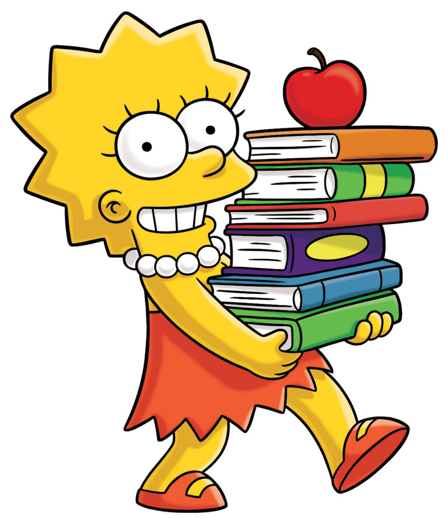


Summary

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!