Extractive Summarization with SWAP-NET: Sentences and Words from Alternating Pointer Networks

Aishwarya Jadhav

Indian Institute of Science

Bangalore, India

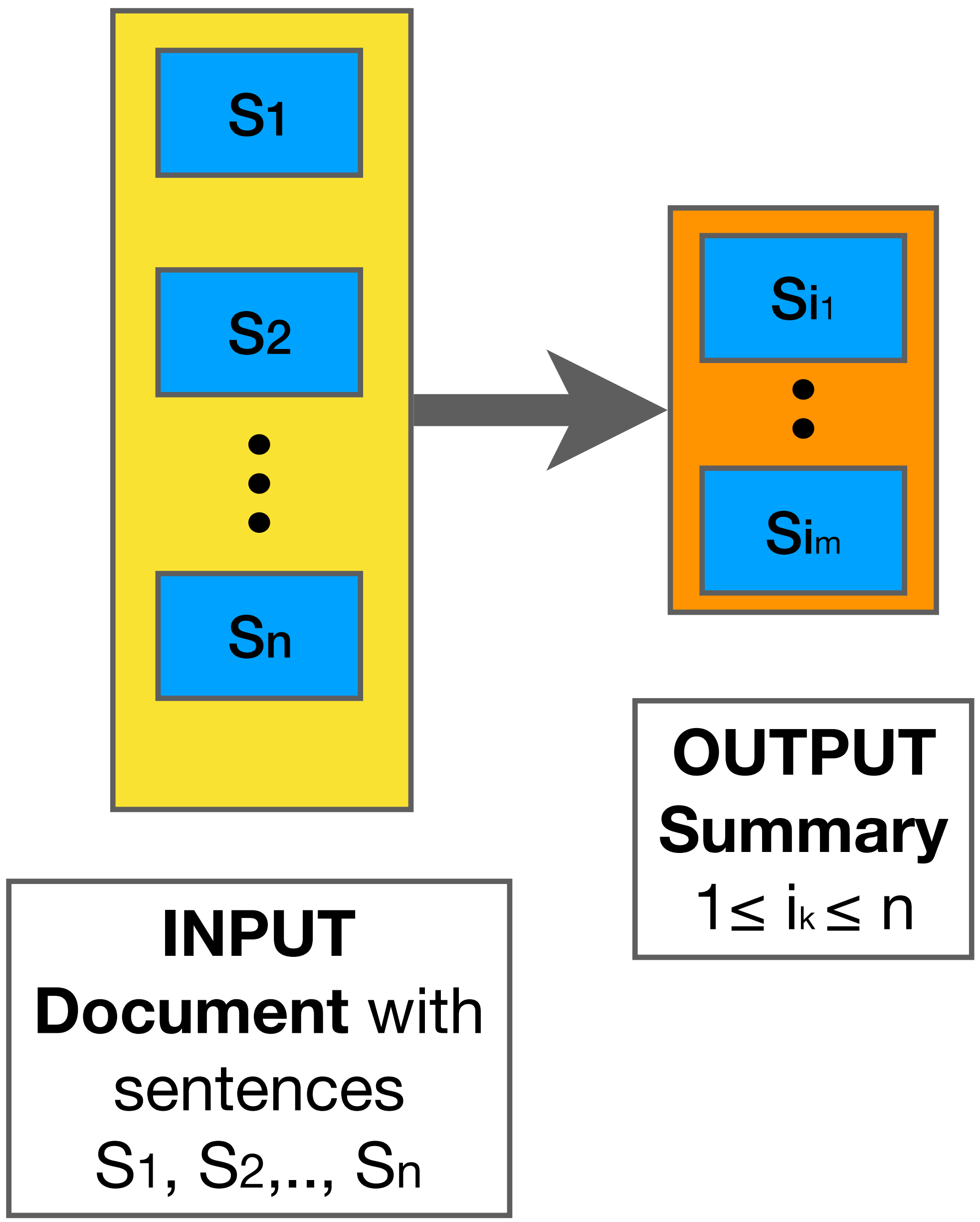
Vaibhav Rajan

School of Computing

National University of Singapore

Extractive Summarization

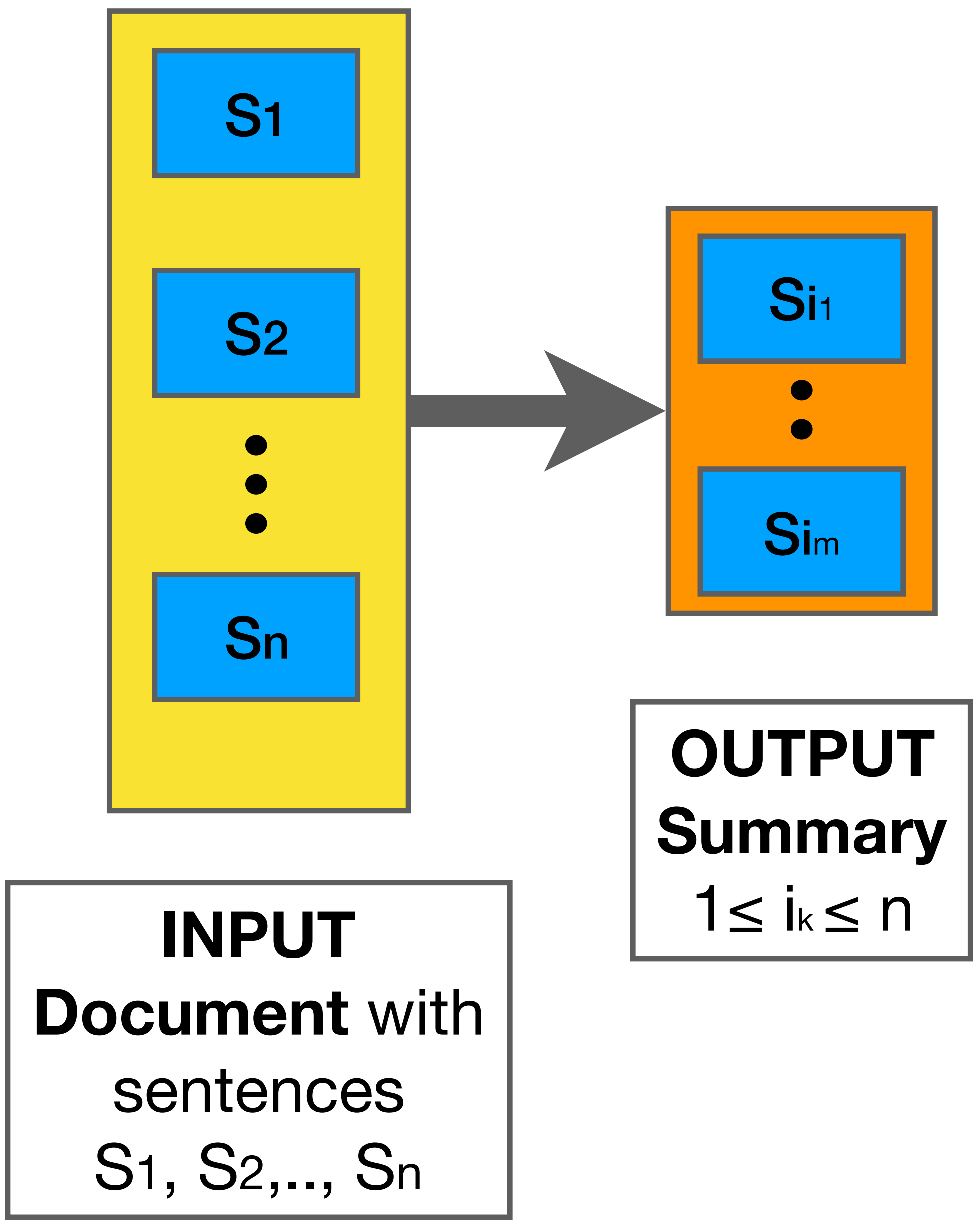
Select *salient sentences* from *input document* to create a summary



- **Supervised** extractive summarization for **single document** inputs

Our Contribution

A Deep Learning Architecture for training an extractive summarizer: **SWAP-NET**



- Unlike previous methods, SWAP-NET uses **keywords** for sentence selection
- Predicts both **important words and sentences** in document
- Two-level Encoder-Decoder Attention model
- **Outperform state of the art** extractive summarisers.

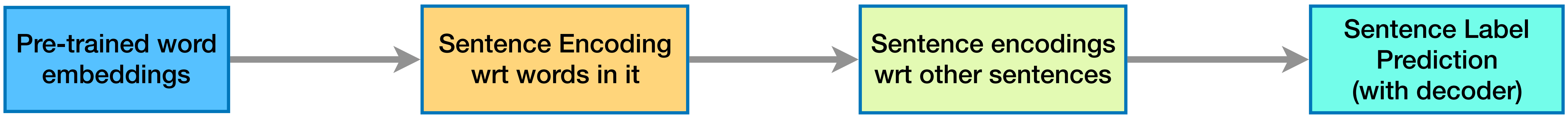
Extractive Summarization Methods

Recent extractive summarization methods

Extractive Summarization Methods

Recent extractive summarization methods

- **NN (Cheng and Lapata, 2016)**

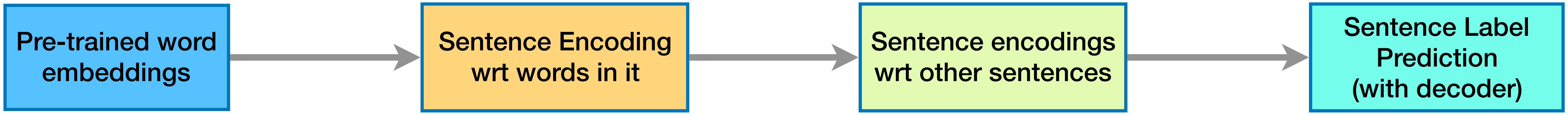


Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. 54th Annual Meeting of the Association for Computational Linguistics.

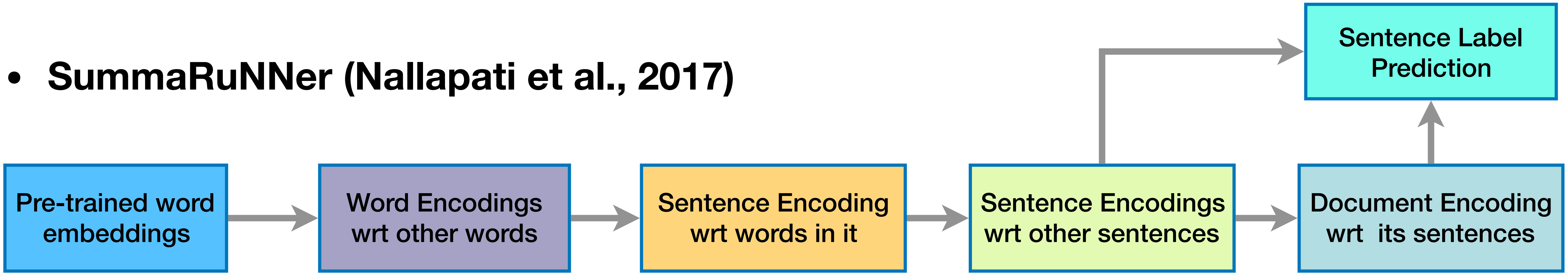
Extractive Summarization Methods

Recent extractive summarization methods

- **NN (Cheng and Lapata, 2016)**



- **SummaRuNNer (Nallapati et al., 2017)**



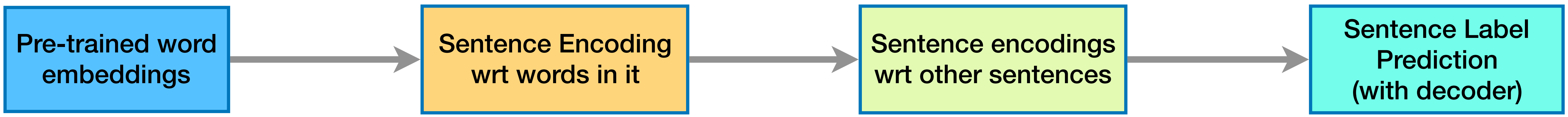
Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of docments. In Association for the Advancement of Artificial Intelligence, pages 3075–3081.

Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. 54th Annual Meeting of the Association for Computational Linguistics.

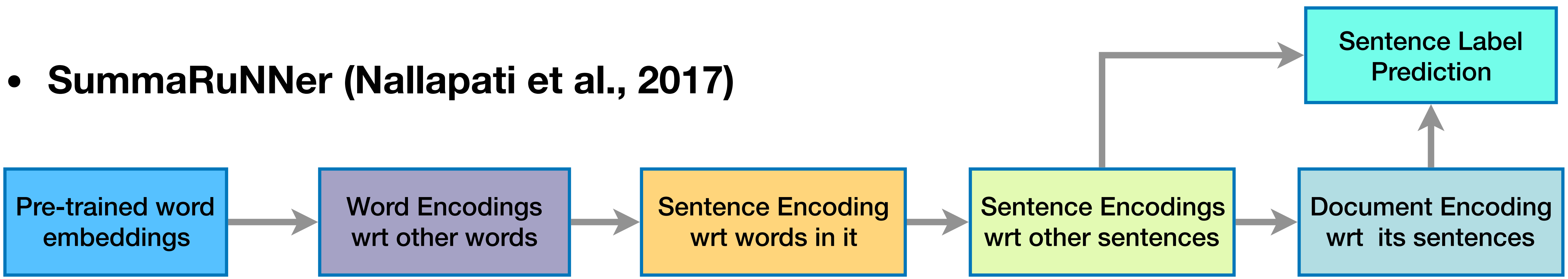
Extractive Summarization Methods

Recent extractive summarization methods

- **NN (Cheng and Lapata, 2016)**



- **SummaRuNNer (Nallapati et al., 2017)**



- Both assume **saliency of sentence s depends** on **salient sentences** appearing before **s**

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of docments. In Association for the Advancement of Artificial Intelligence, pages 3075–3081.

Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. 54th Annual Meeting of the Association for Computational Linguistics.

Intuition Behind Approach

***Question**: Which sentence should be considered salient (part of summary)?*

- Our hypothesis: **saliency of a sentence** depends on both **salient sentences and words appearing before that sentence** in the document

- Similar to graph based models by Wan et al. (2007)

- Along with labelling sentences we also **label words to determine their saliency**

- Moreover, **saliency of a word depends on previous salient words and sentences**

Xiaojun Wan, Jianwu Yang, and Jianguo Xiao. 2007. Towards an iterative reinforcement approach for simultaneous document summarization and keyword extraction. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pages 552–559.

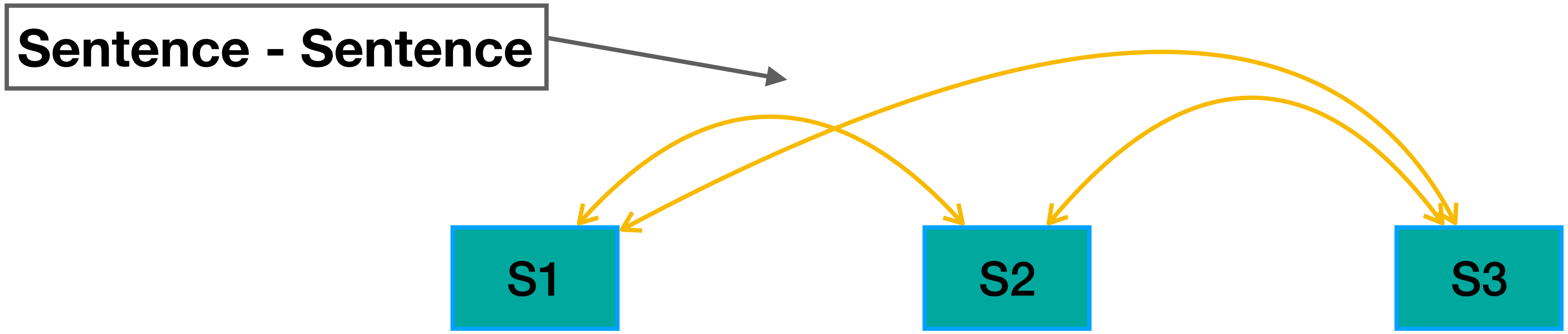
Intuition Behind Approach

Three types of Interactions:

- Sentence-Sentence Interaction
- Word-Word Interaction
- Sentence-Word Interaction

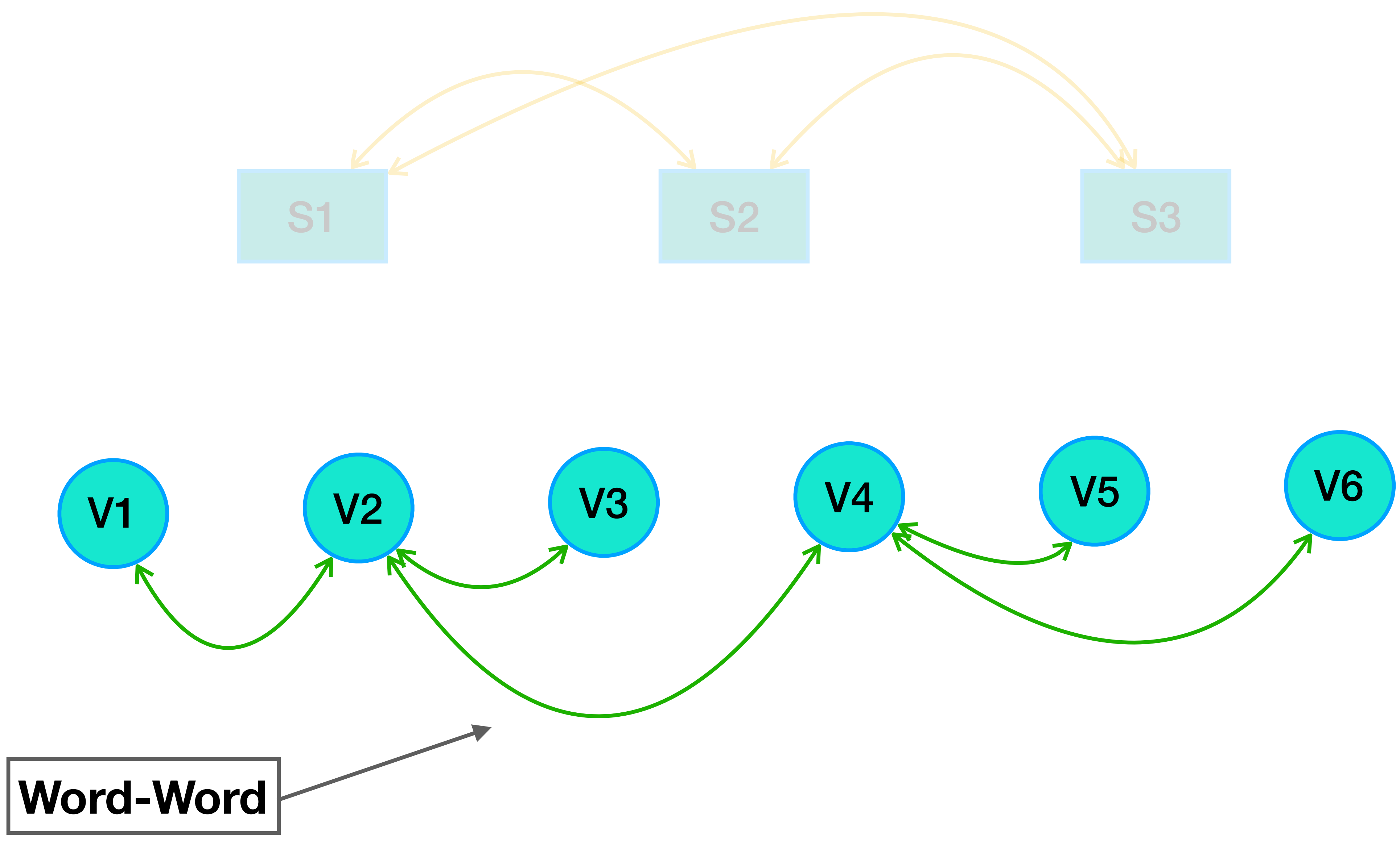
Intuition: Interaction Between Sentences

A sentence should be salient if it is heavily linked with other salient sentences



Intuition: Interaction Between Words

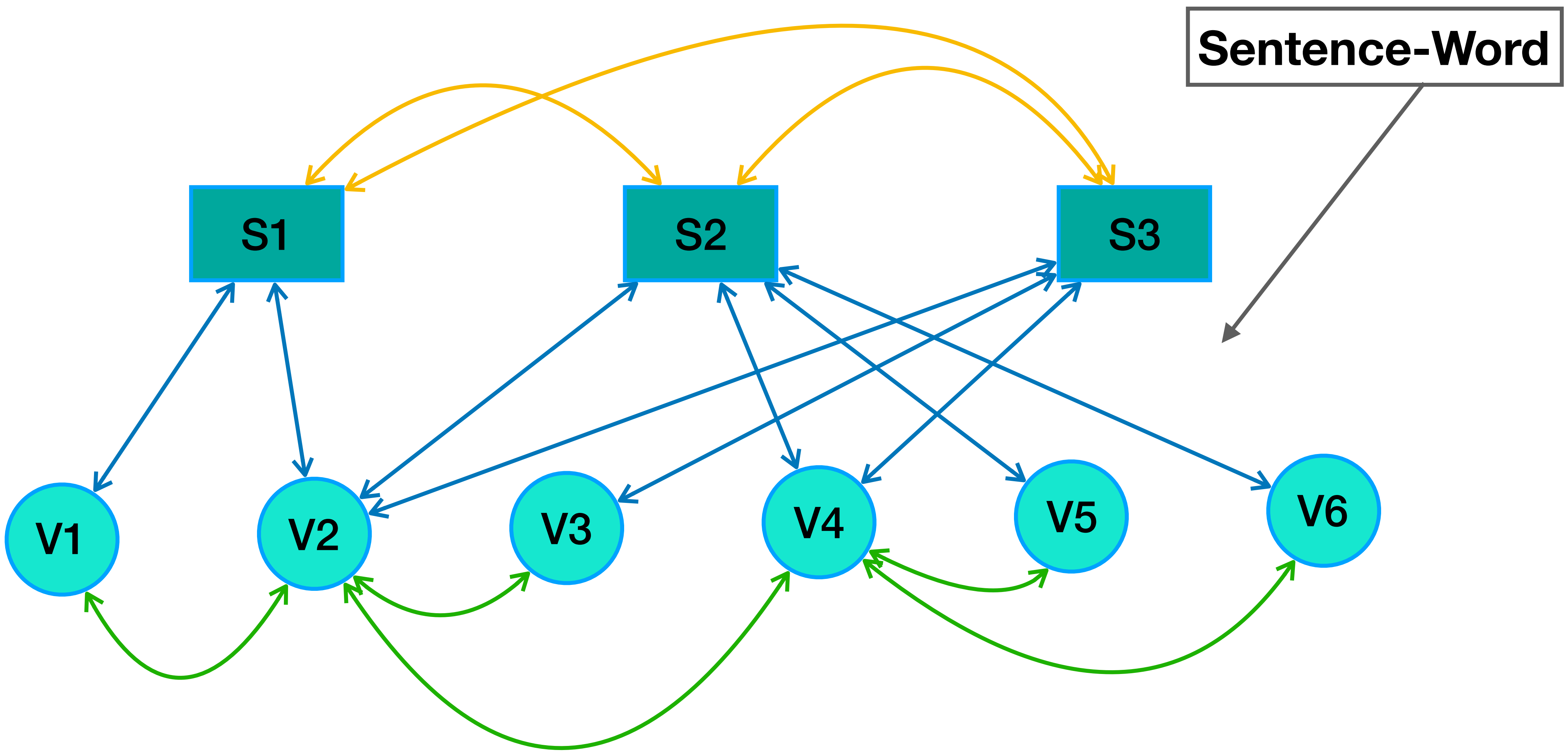
A word should be salient if it is heavily linked with other salient words



Intuition: Words and Sentences Interaction

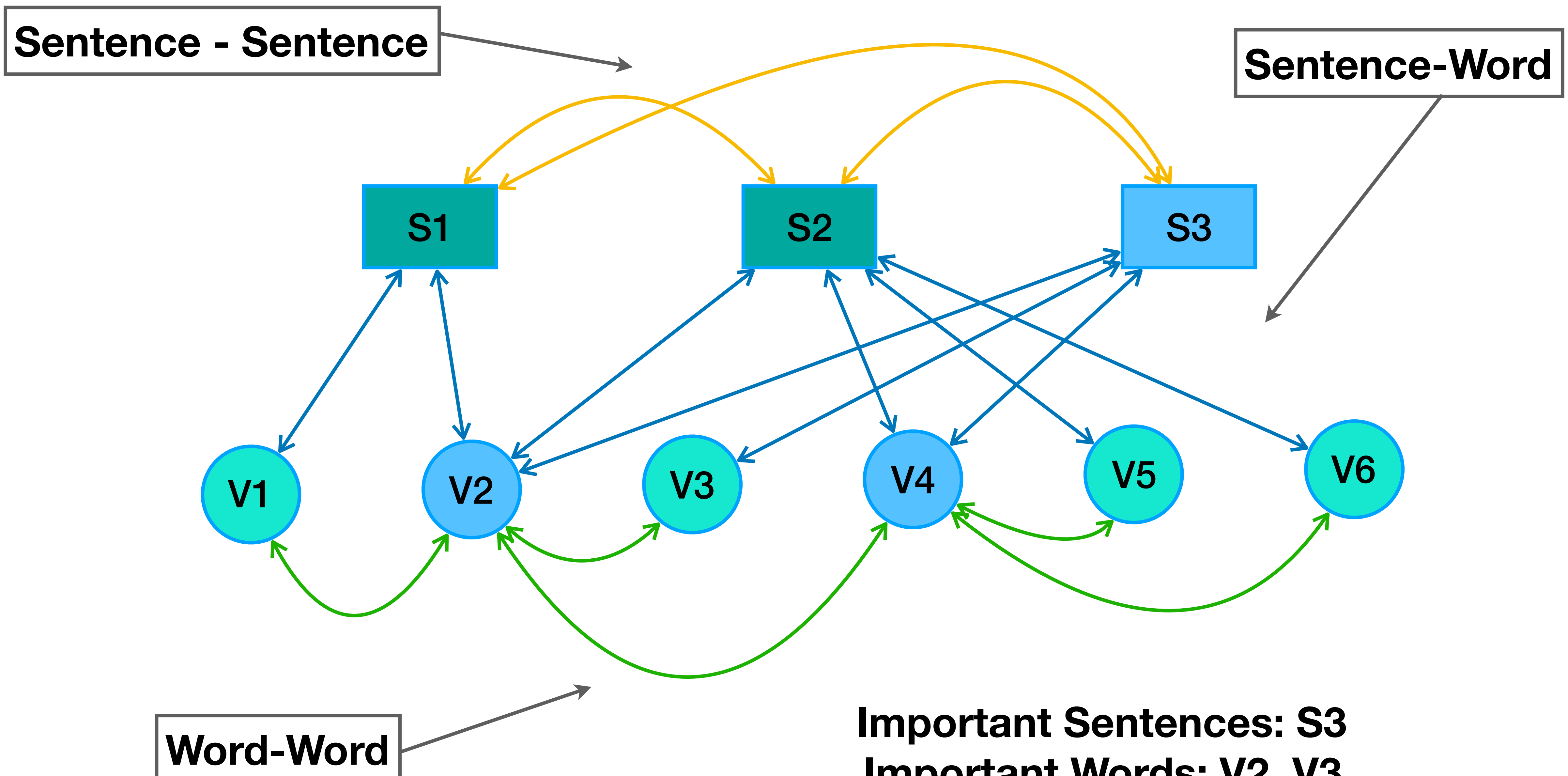
A sentence should be salient if it contains many salient words

A word should be salient if it appears in many salient sentences



Intuition: Words and Sentences Interaction

Generate **extractive summary** using both **important words and sentences**



Important Sentences: S3
Important Words: V2, V3

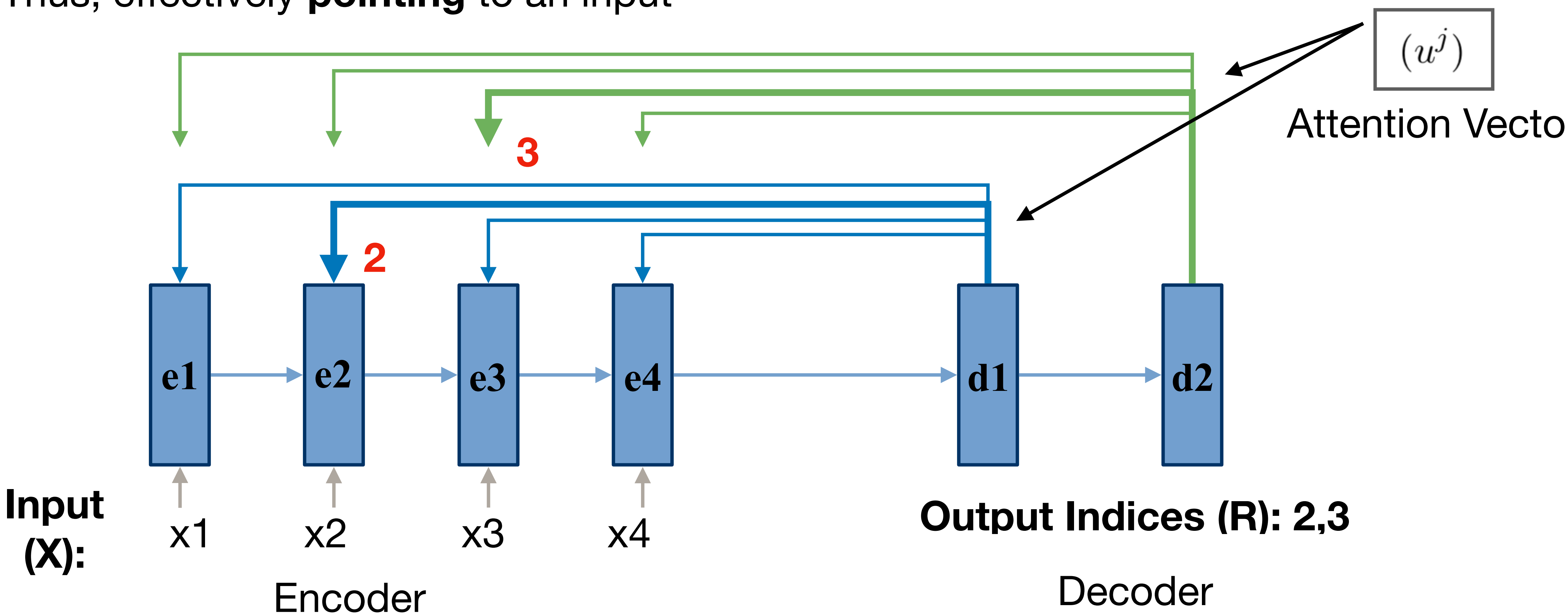
Keyword Extraction and Sentence Extraction

- Sentence to Sentence Interaction as **Sentence Extraction**
- Word to Word Interaction as **Word Extraction**
- For discrete sequences, pointer networks have been successfully used to learn how to **select positions from an input sequence**
- We use **two pointer networks** one at word-level and another at sentence-level

Pointer Network

Pointer network (Vinyals et al., 2015),

- Encoder-Decoder architecture with Attention
- Attention mechanism is used to **select one of the inputs** at each decoding step
- Thus, effectively **pointing** to an input

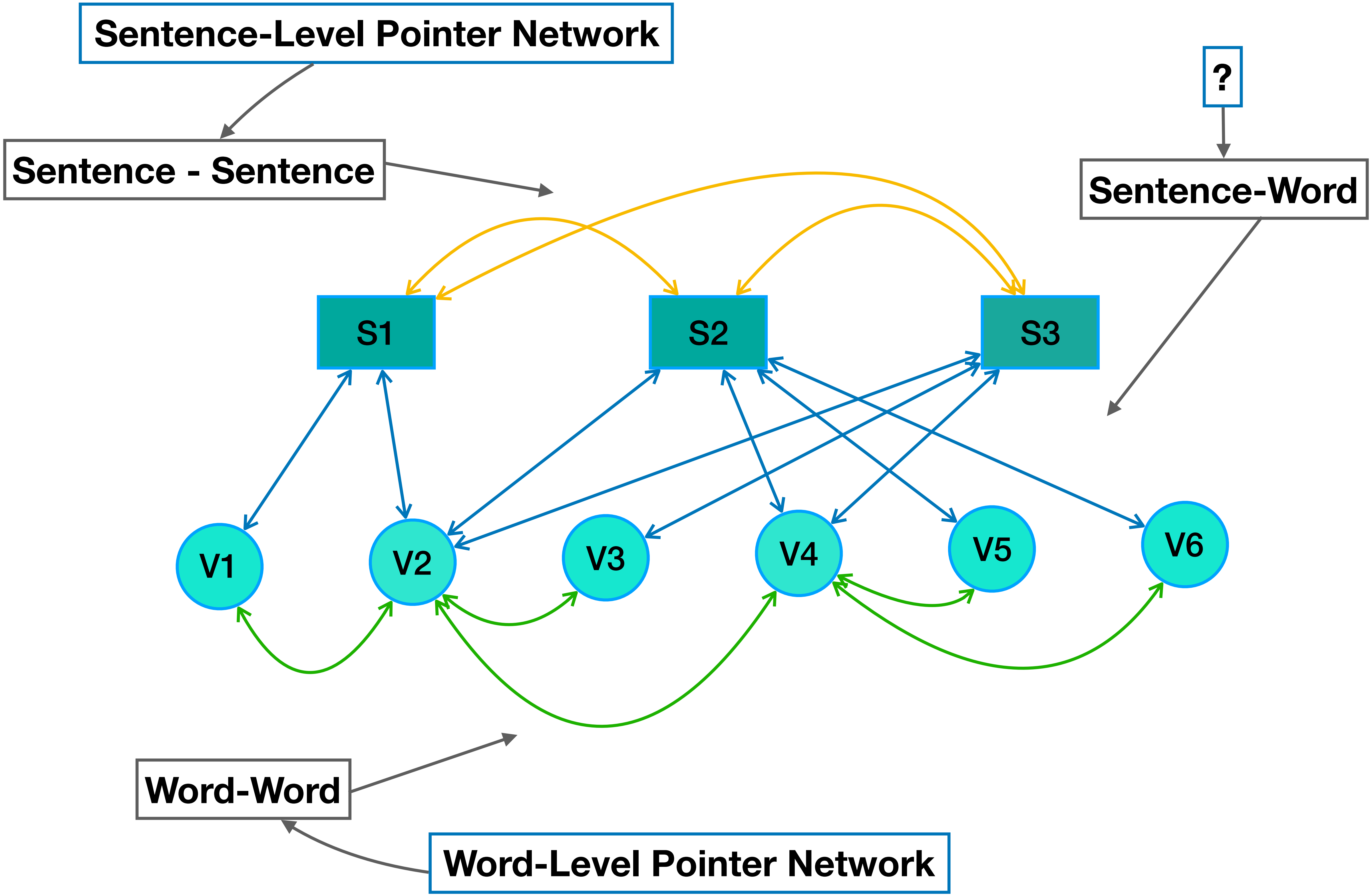


$$u_i^j = v^T \tanh(W_e e_i + W_d d_j), \ \ i \in (1, \ldots, n)$$

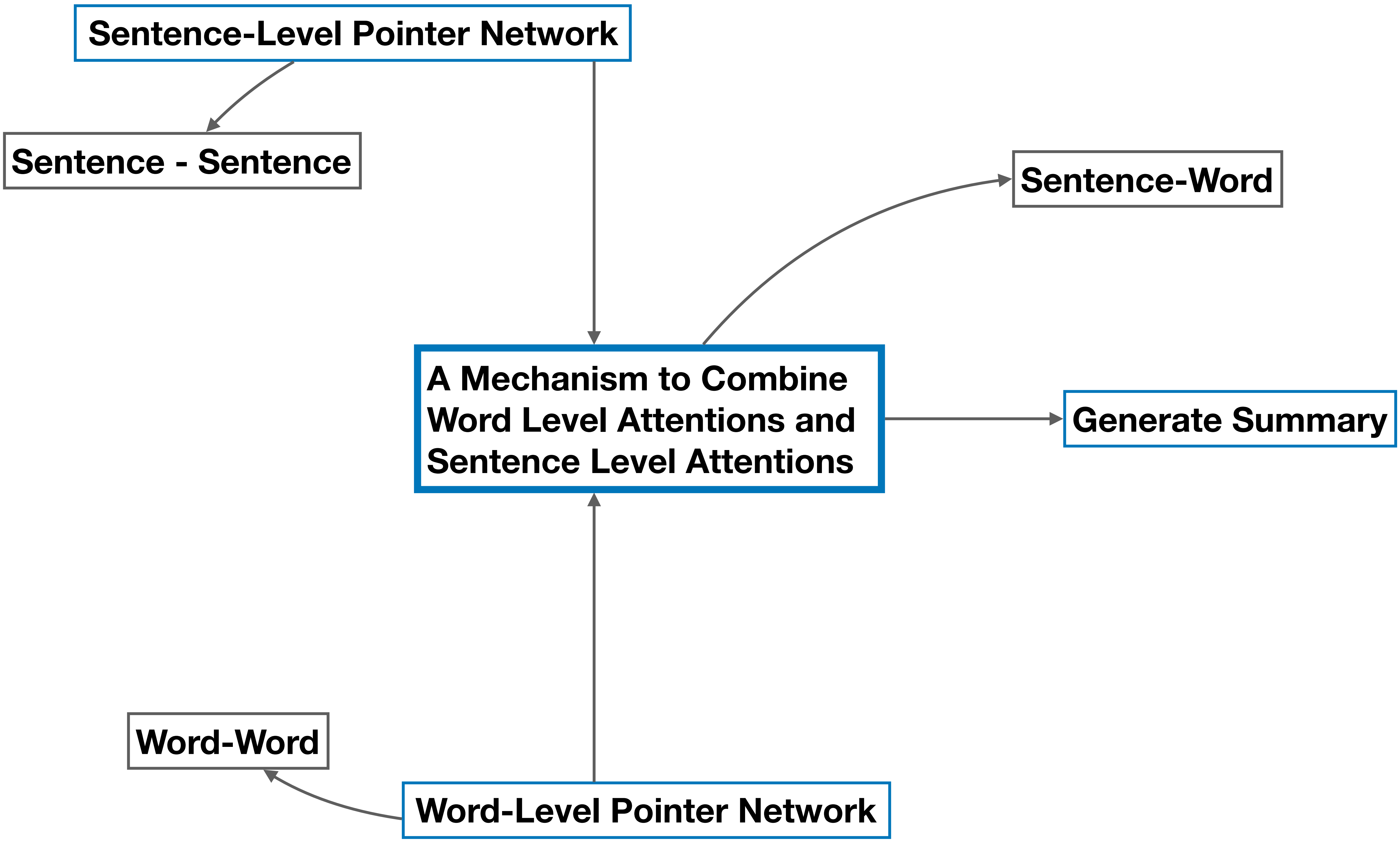
$$p(r_j | r_1, \ldots . r_{j-1}, X) = \text{softmax}(u^j)$$

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In *Advances in Neural Information Processing Systems*, pages 2692–2700.

Three Interactions



Three Interactions: SWAP-NET

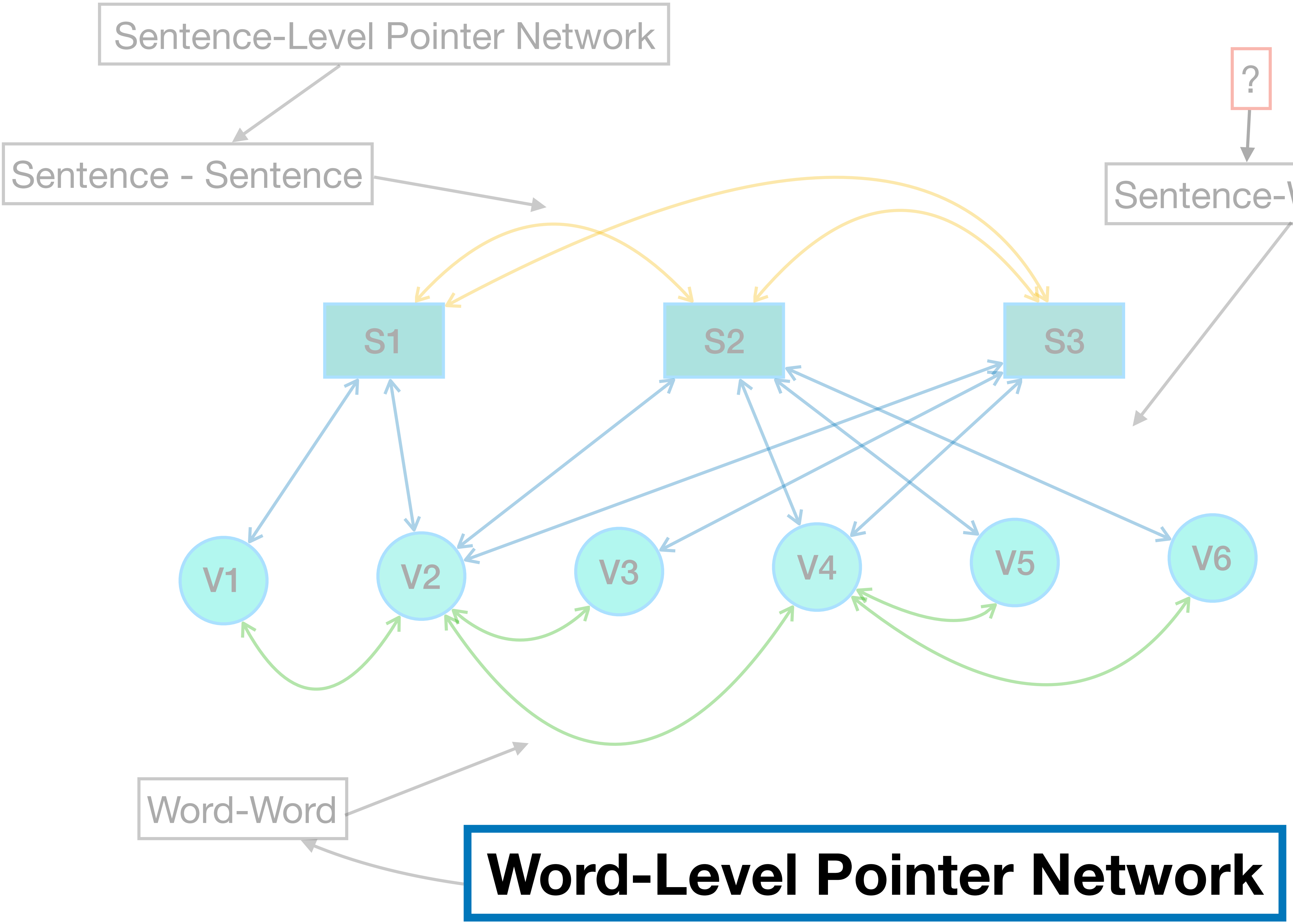


Questions

Q1 : ***How can the two attentions be combined?***

Q2 : ***How can the summaries be generated considering both the attentions?***

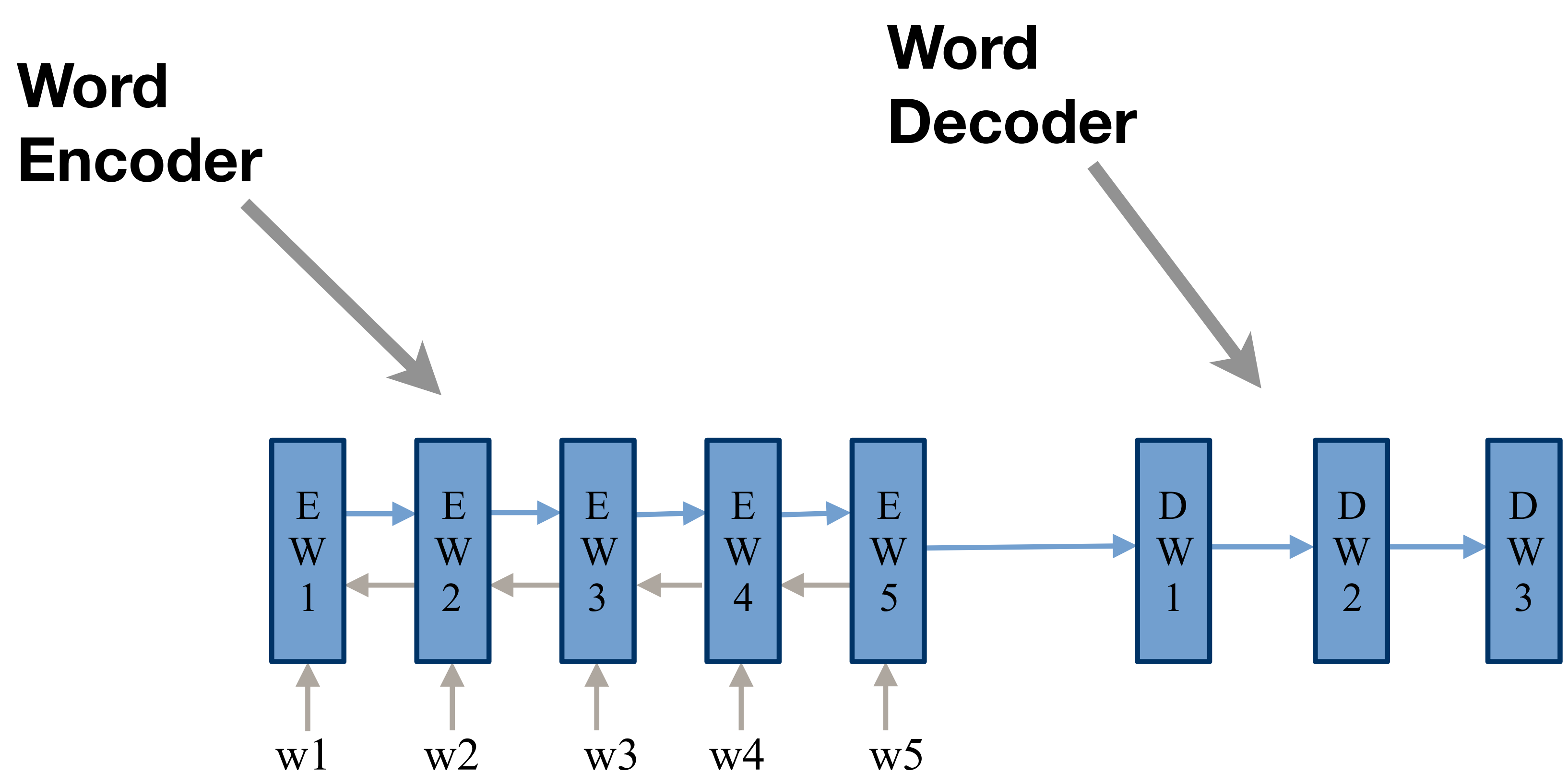
Three Interactions: SWAP-NET



SWAP-NET Architecture: Word-Level Pointer Network

Similar to Pointer Network,

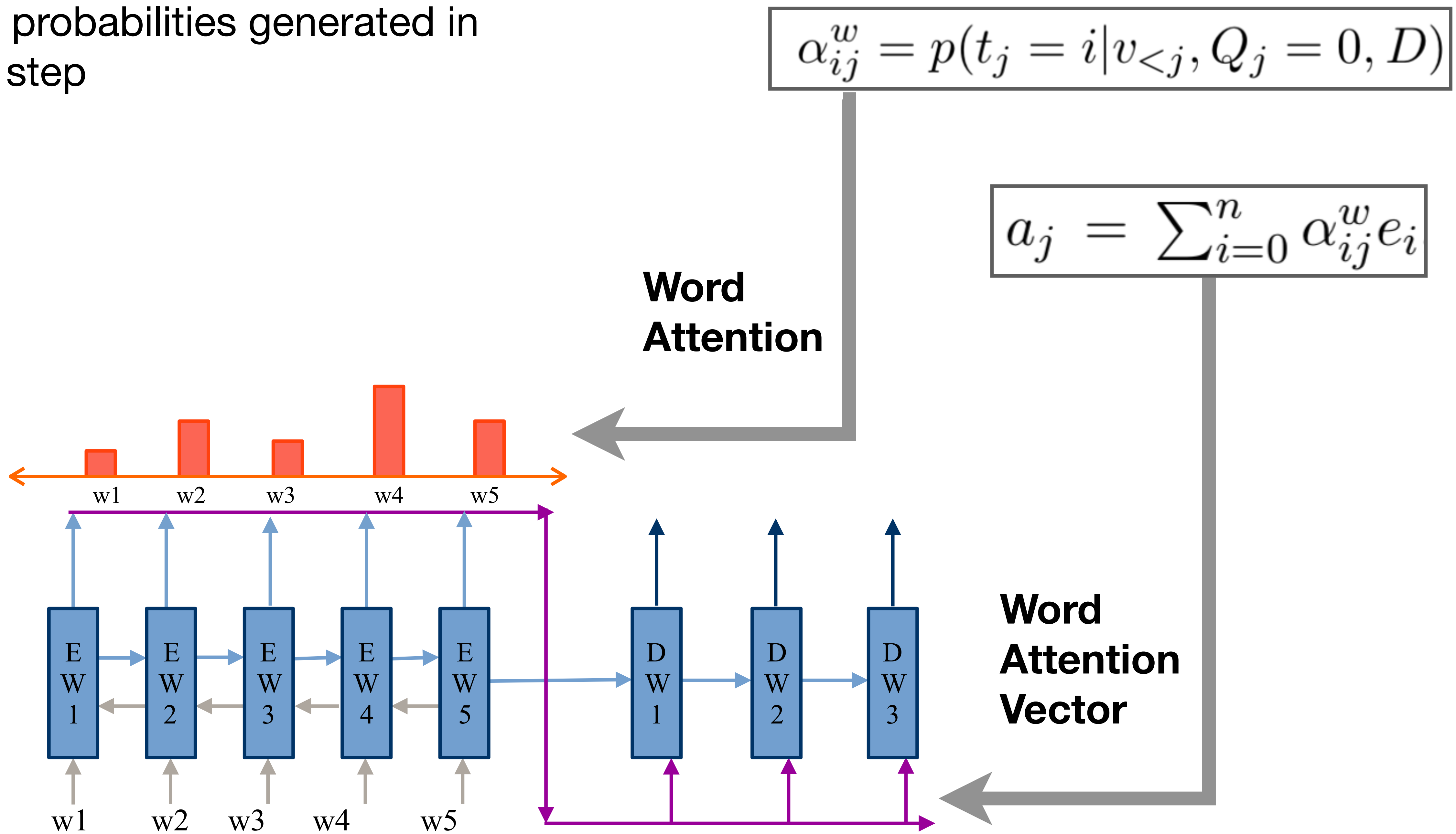
- The word encoder is bi-directional LSTM
- Word-level decoder learns to point to important words



SWAP-NET Architecture: Word-Level Pointer Network

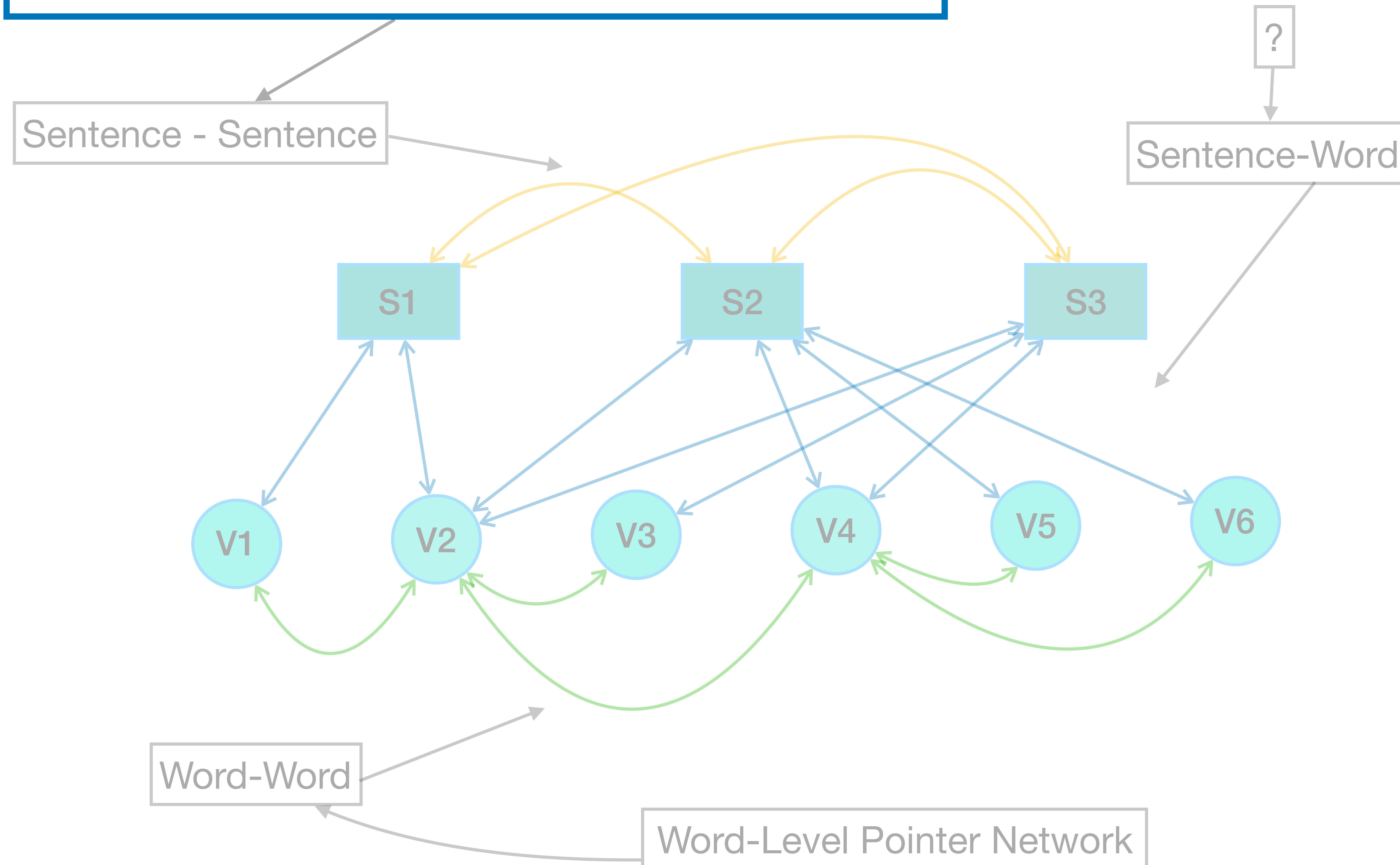
- Purple line: attention vector given as input to each decoding step
    - Sum of word encodings weighted by attention probabilities generated in previous step

Probability of word i, at decoding step j



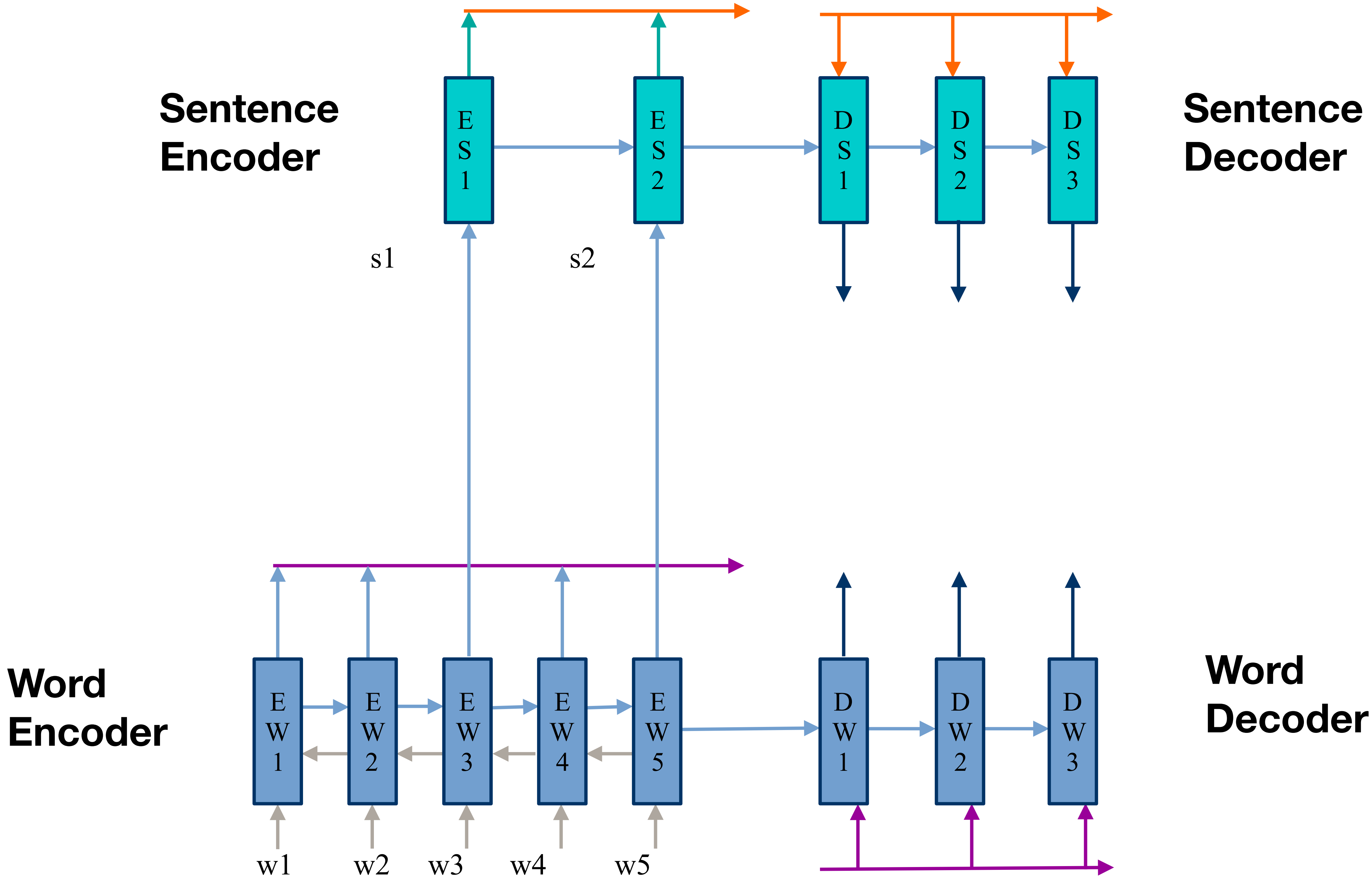
Three Interactions: SWAP-NET

Sentence-Level Pointer Network



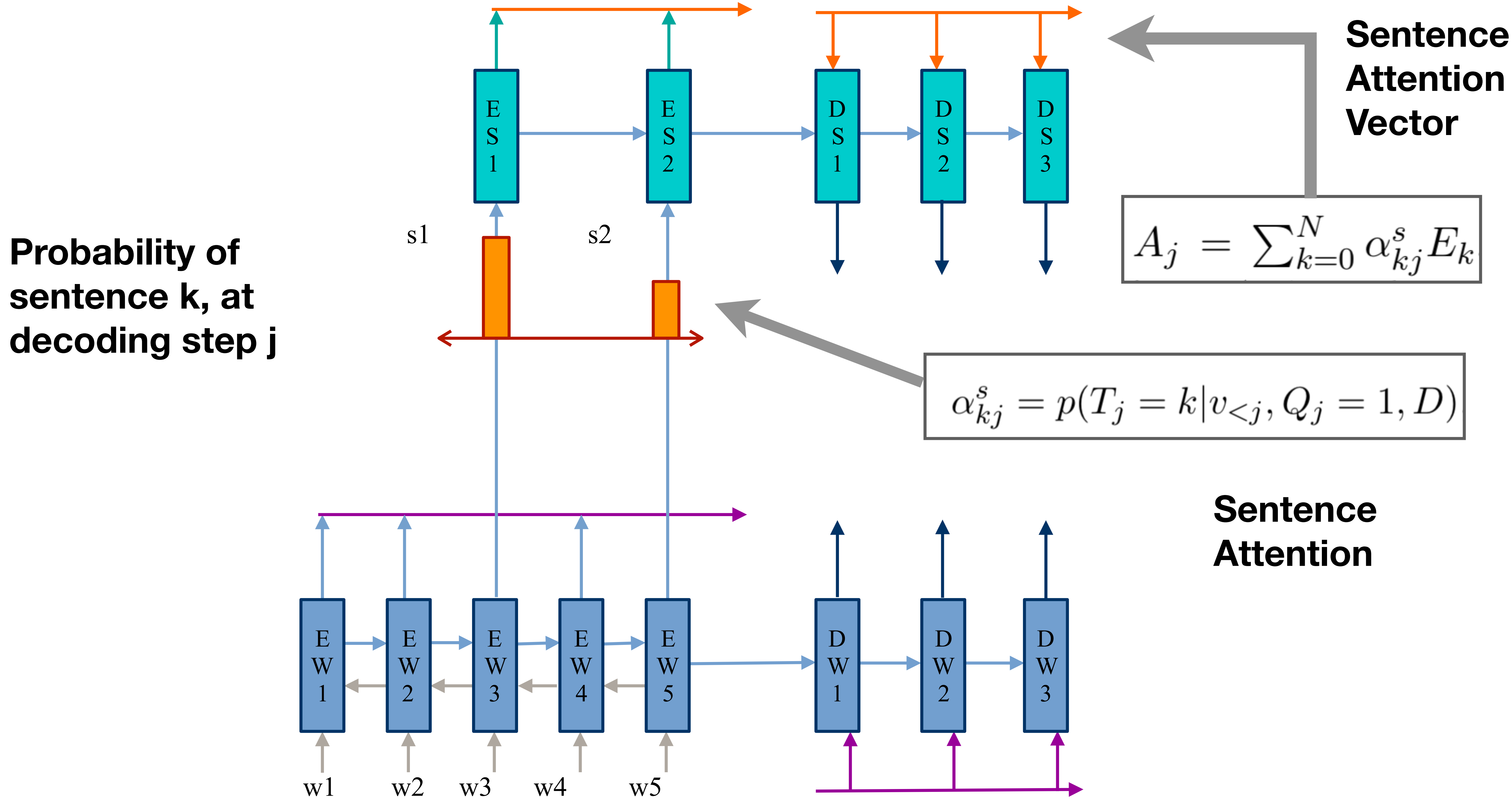
SWAP-NET Architecture:
Sentence-Level Hierarchical Pointer Network

Sentence is represented by encoding of last word of that sentence



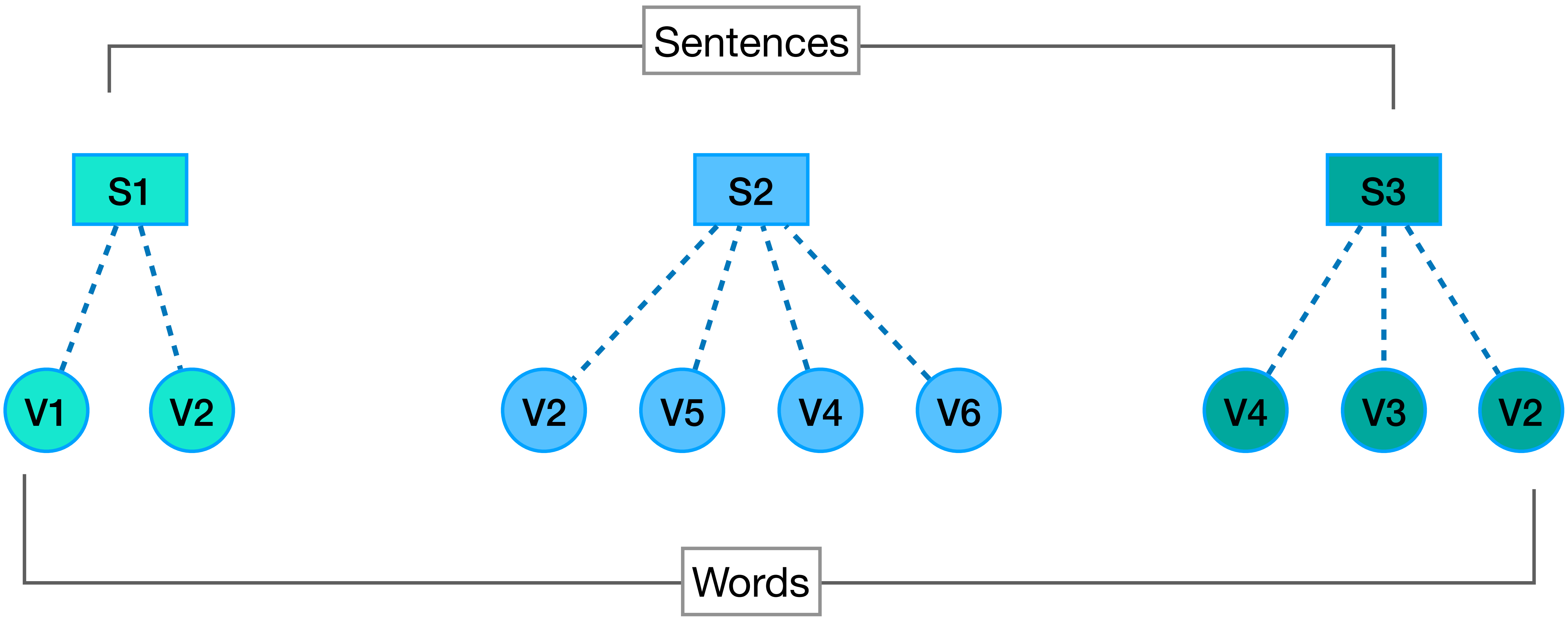
SWAP-NET Architecture:
Sentence-Level Hierarchical Pointer Network

Attention vectors are sum of sentence encodings weighted by attention probabilities by previous decoding step



Combining Sentence Attention and Word Attention

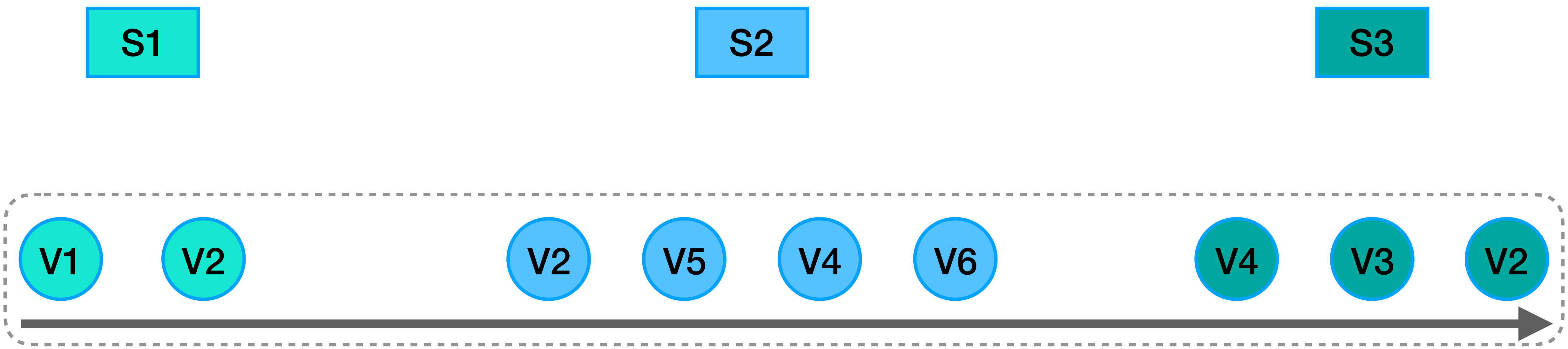
Q1 : *How can the two attentions be combined?*



A document with three sentences and corresponding words is shown

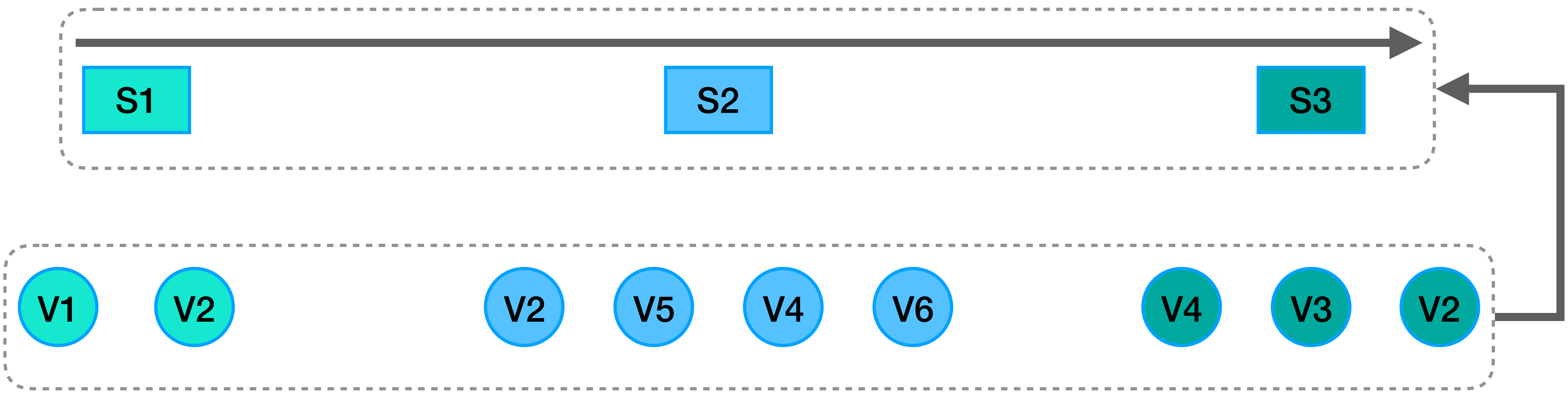
Sentence and Word Interactions

Possible Solution:
Step 1: Hold sentence processing. Then group all words and determine their saliency sequentially



Sentence and Word Interactions

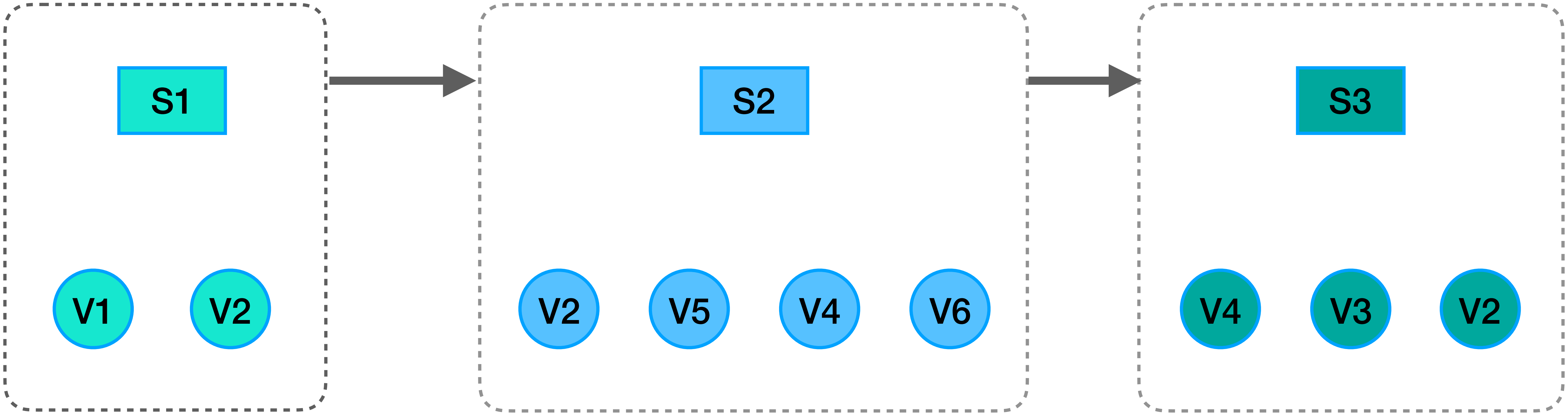
Possible Solution:
Step 2: Using output of step 1, i.e., using keywords, process sentences to determine salient sentences



INCOMPLETE SOLUTION : This methods processes sentence depending on words **but does not use sentences for processing words.**

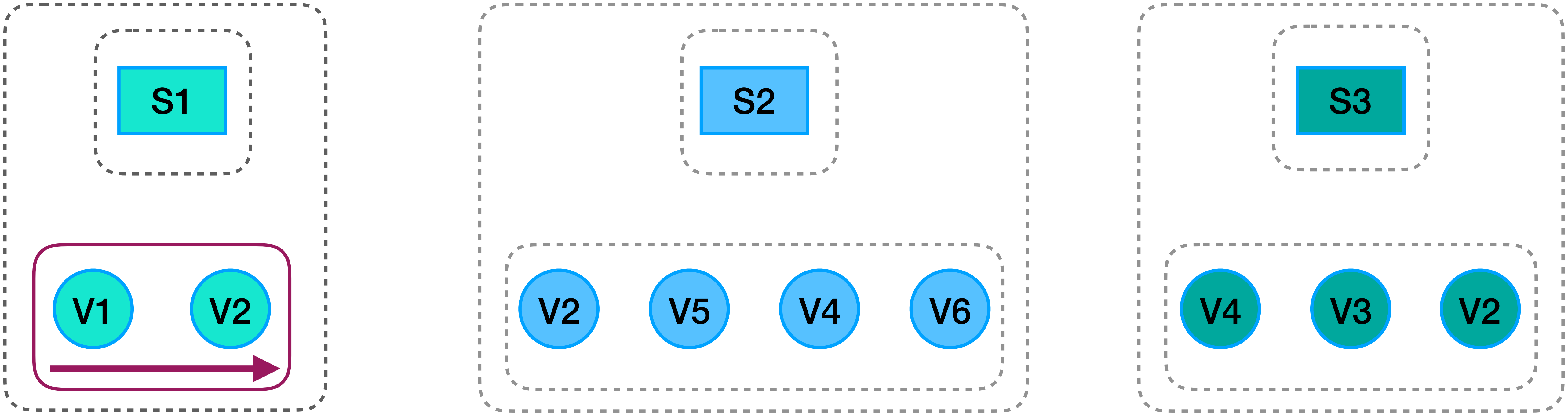
Sentence and Word Interactions

Solution:
Group each sentence and its words separately and process them sequentially



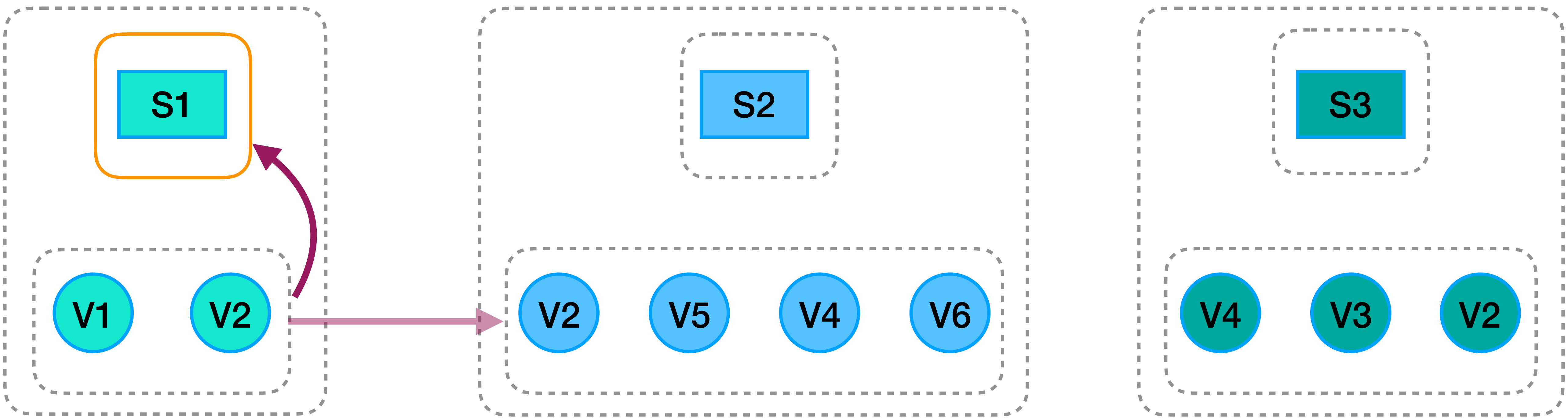
Sentence and Word Interactions

Step1:
Hold sentence processing. Determine saliency of words in S1



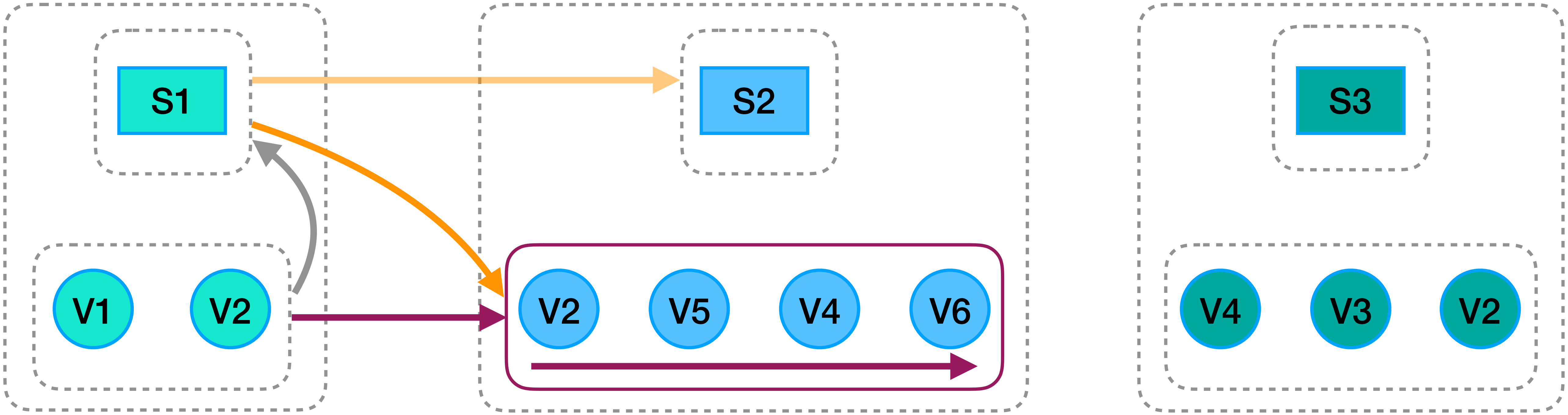
Sentence and Word Interactions

Step2:
Using information about saliency of words in S1
- Hold word processing and resume sentence processing.
- Determine saliency of S1



Sentence and Word Interactions

Step3:
Using information about saliency of both S1 and its words
- Hold sentence processing and resume word processing.
- Determine saliency of words in next sentence S2

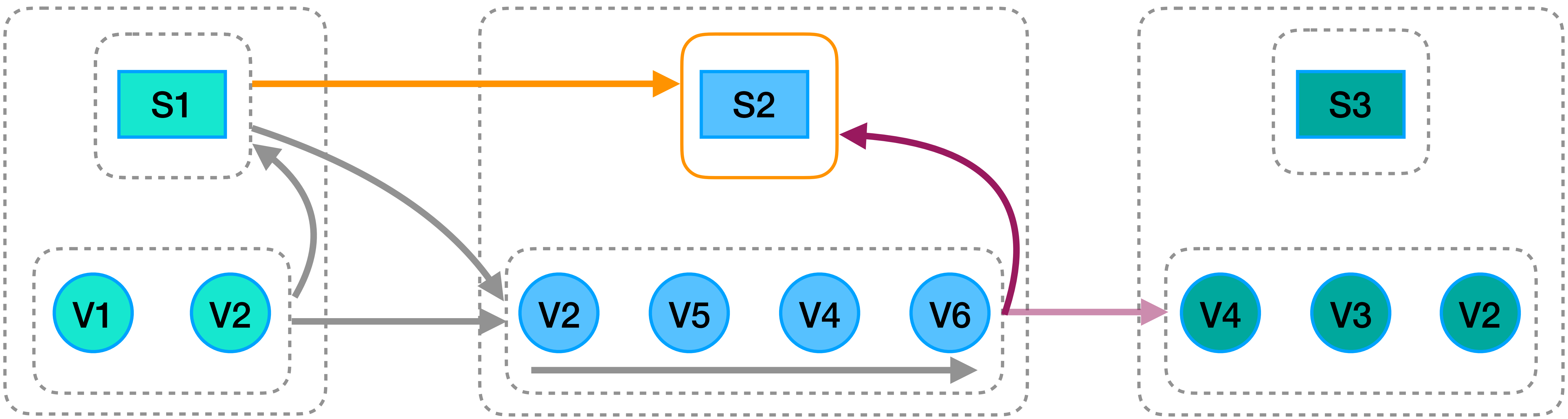


Sentence and Word Interactions

Step4:

Using information about saliency of words in S2 and saliency of previous sentence S1

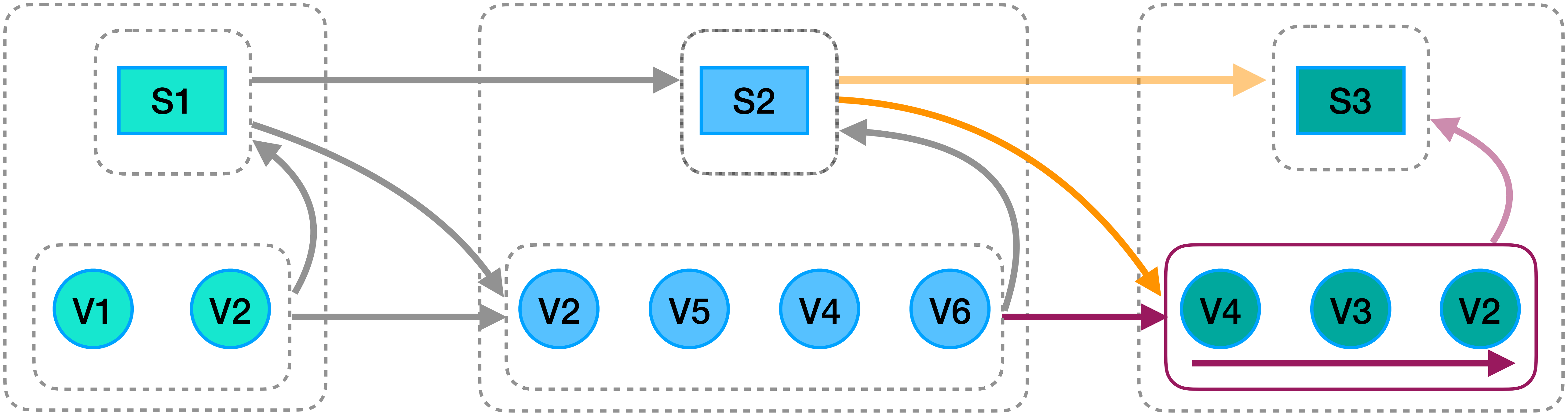
- Hold word processing and resume sentence processing.
- Determine saliency of sentence S2



Sentence and Word Interactions

Solution:
And so on.

This methods ensures that saliency of word and sentence is determined from previously predicted both salient sentences and words

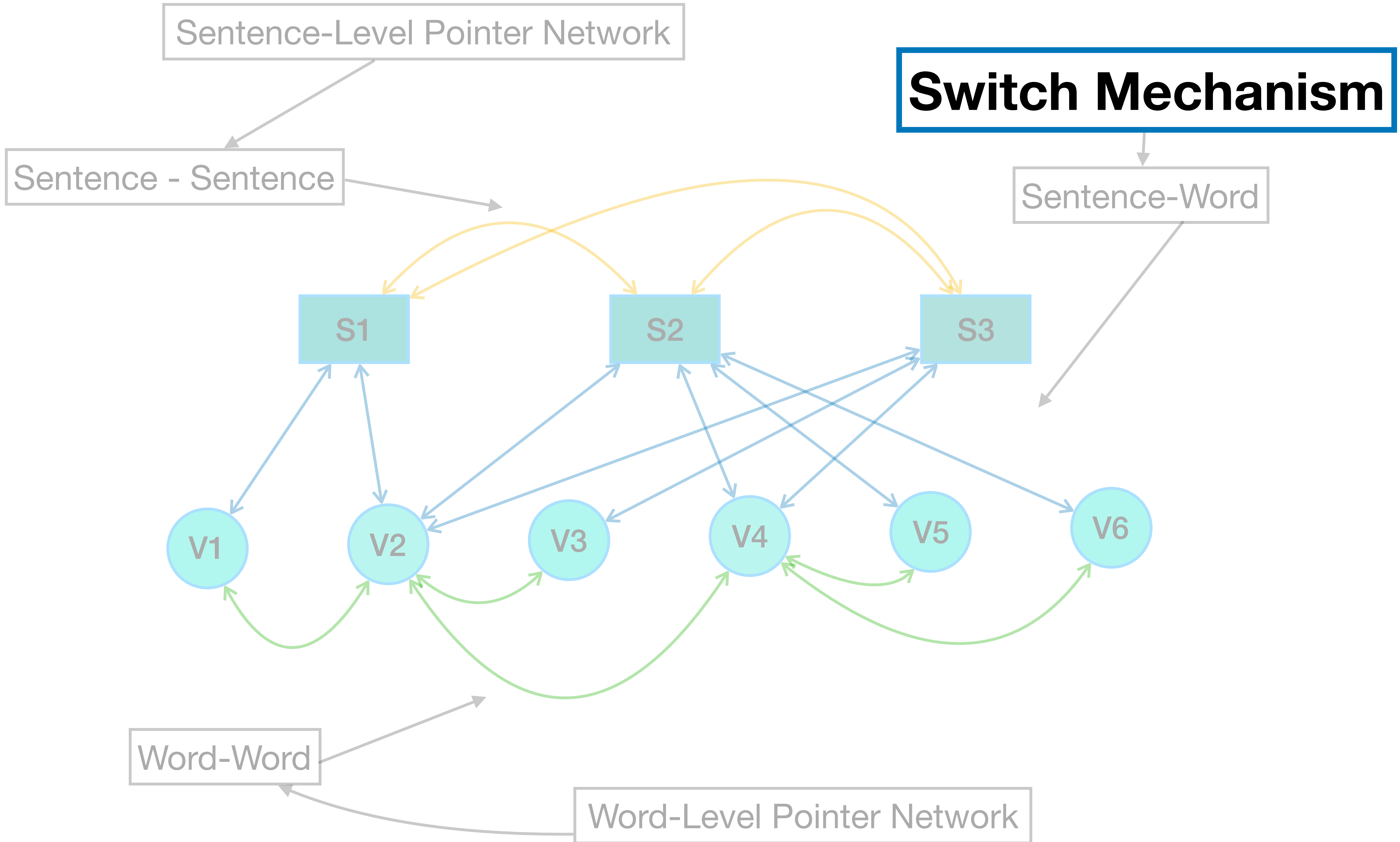


Sentence and Word Interactions

Using previously predicted salient word and sentences

- **Synchronising Decoding Steps:** Decide when to turn off and on word processing and sentence processing to synchronise word and sentence prediction
- **Sharing Attention Vectors:** Determine salient words and sentences

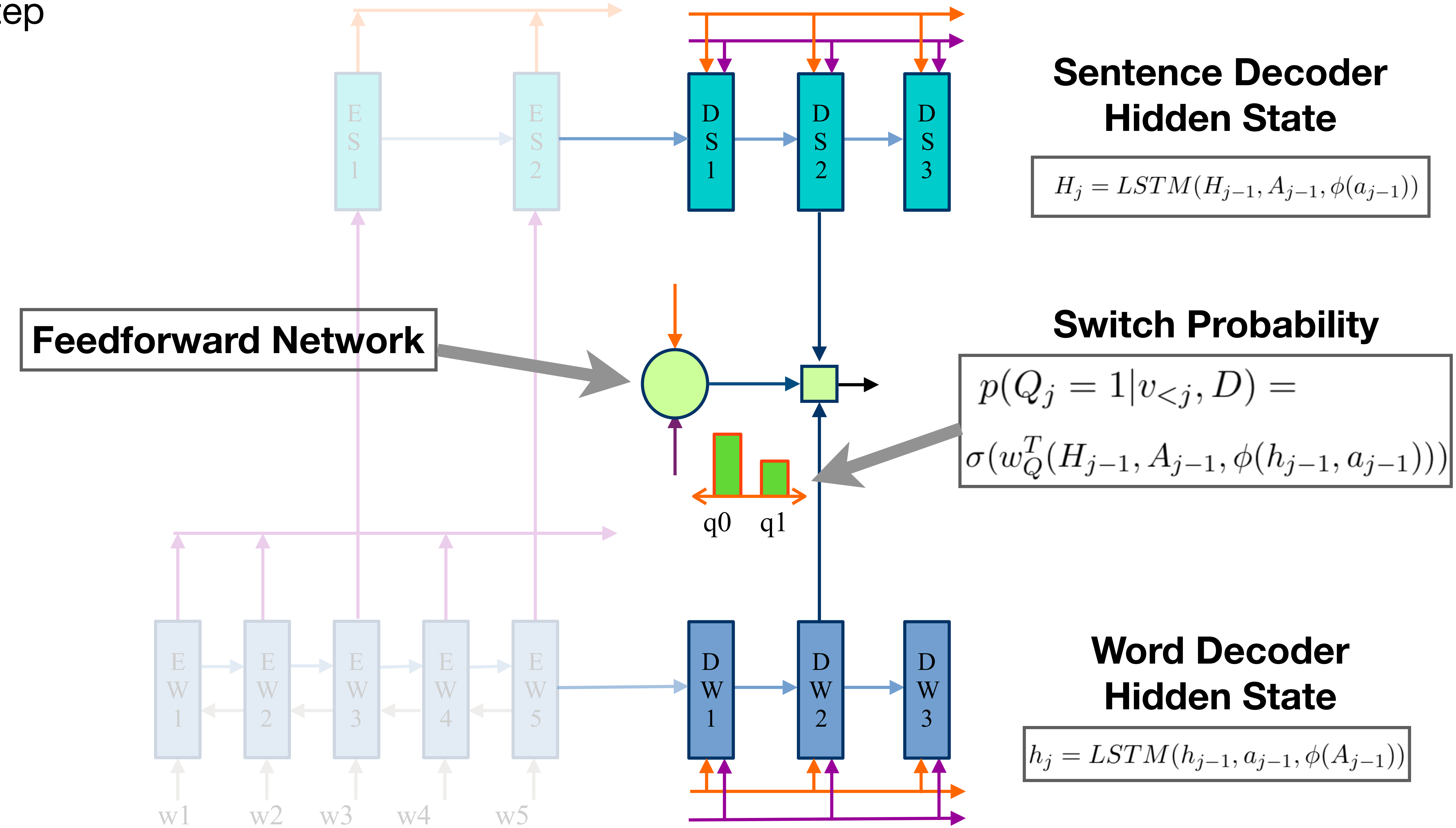
Three Interaction : SWAP-NET



SWAP-NET : Switch Mechanism

Sharing both attention vectors (purple and orange lines) between the two decoder

Synchronising decoding steps of the two decoders by allowing only one decoder output at a step

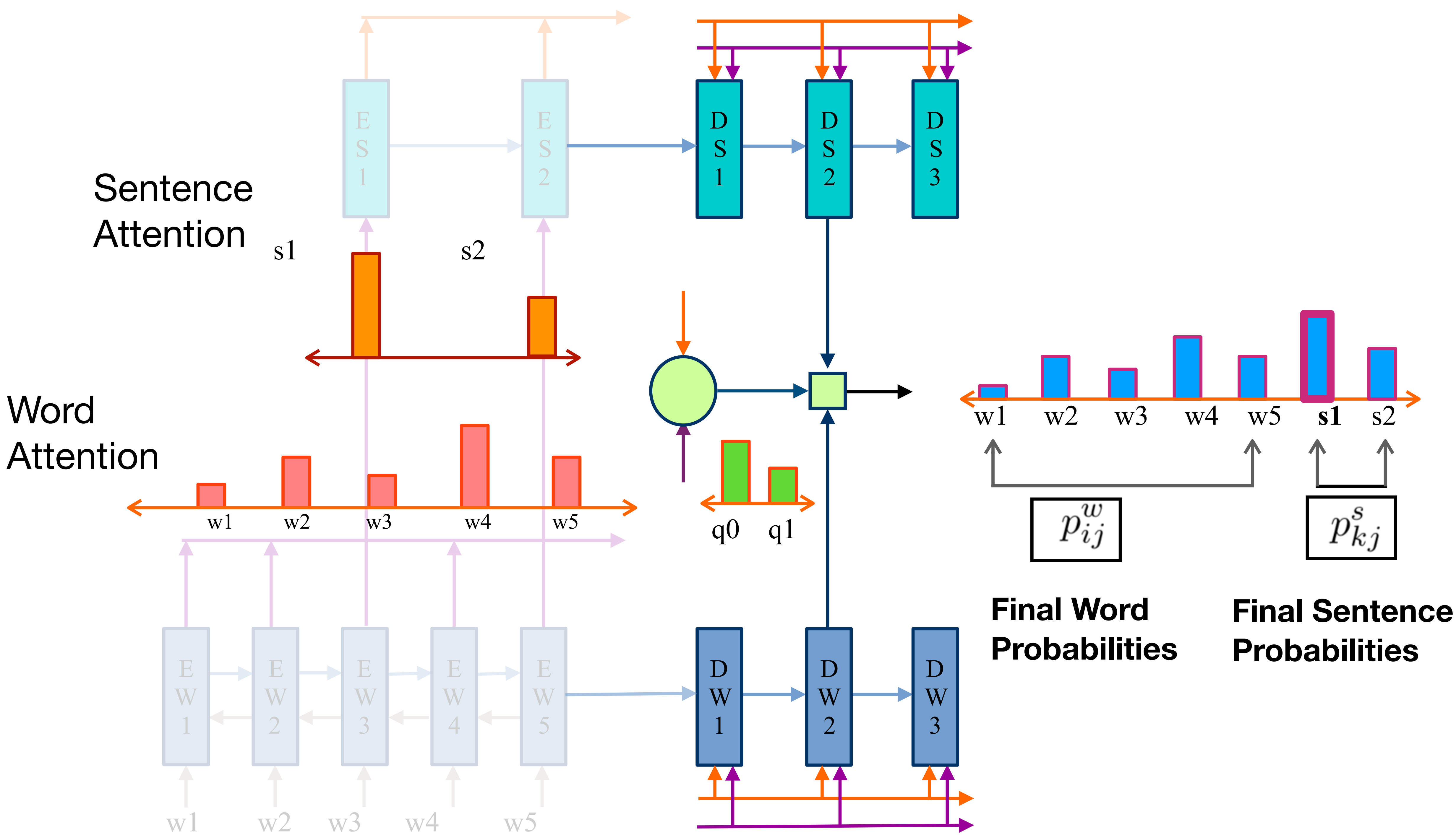


SWAP-NET : Switch Mechanism

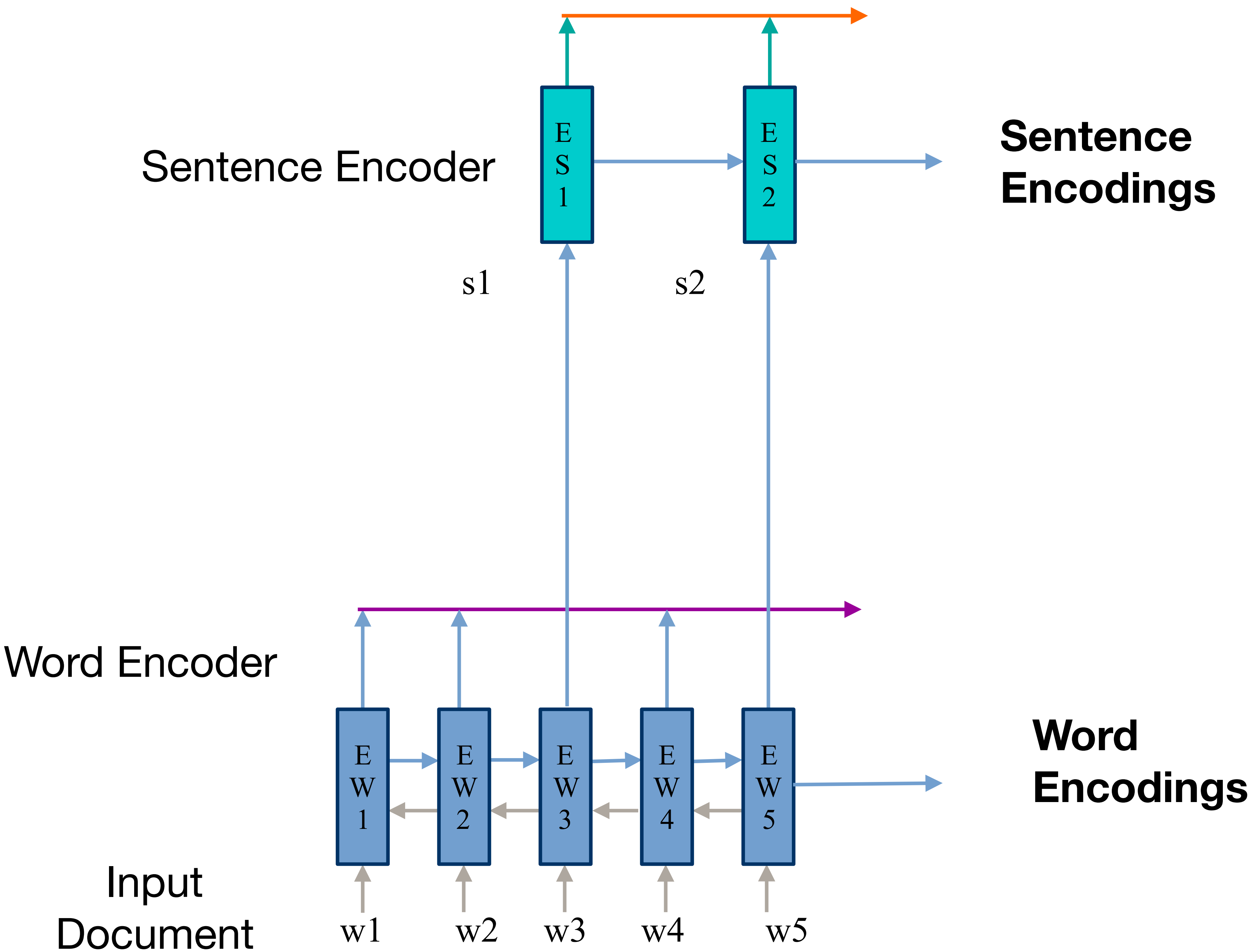
$$p^w_{ij} = \alpha^w_{ij} p(Q_j = 0 | v_{<j}, D)$$

$$p^s_{kj} = \alpha^s_{kj} p(Q_j = 1 | v_{<j}, D)$$

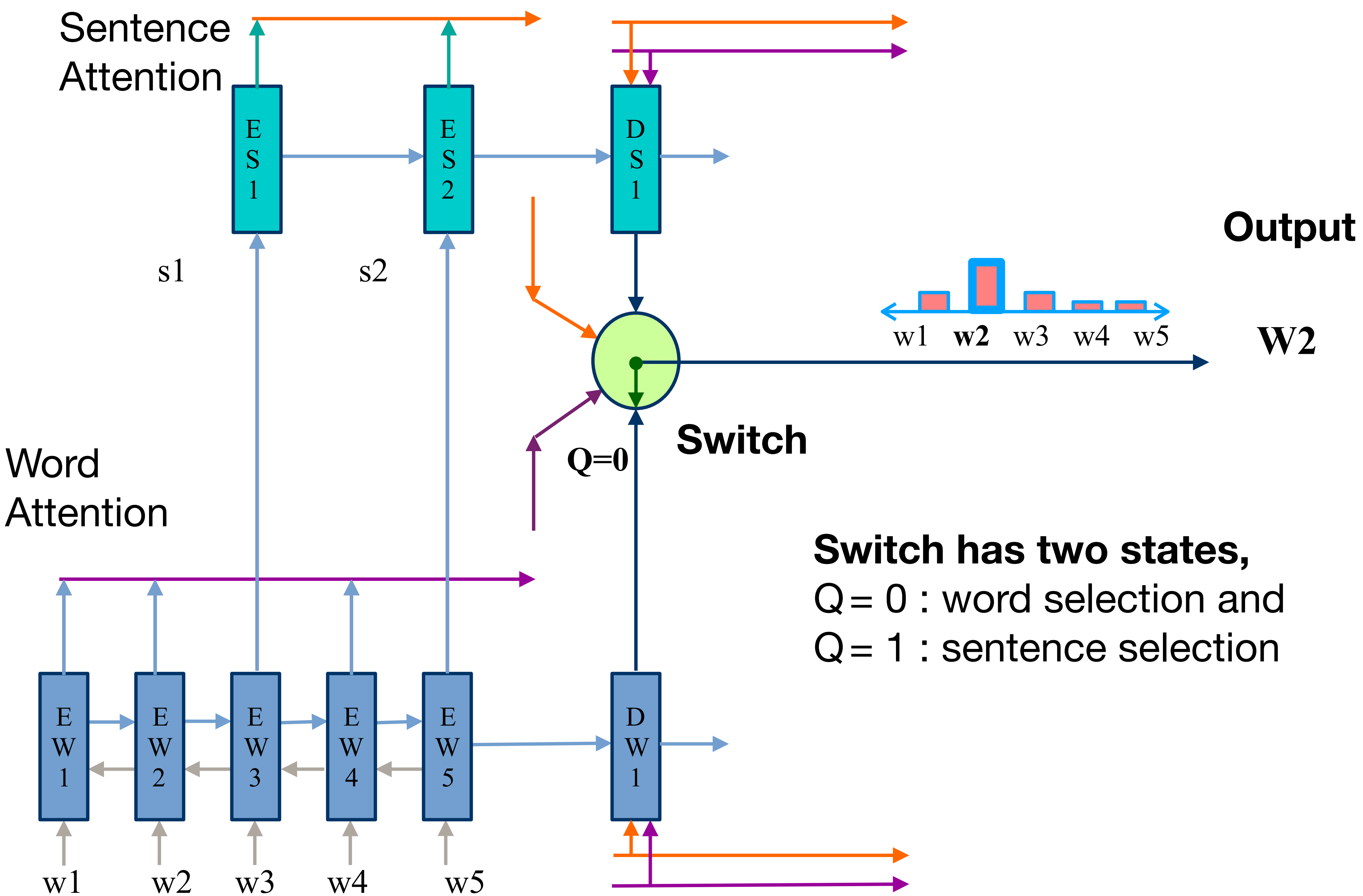
Output is selected with maximum of final word and sentence probabilities



Prediction with SWAP-NET: Encoding



Prediction with SWAP-NET: Decoding Step 1

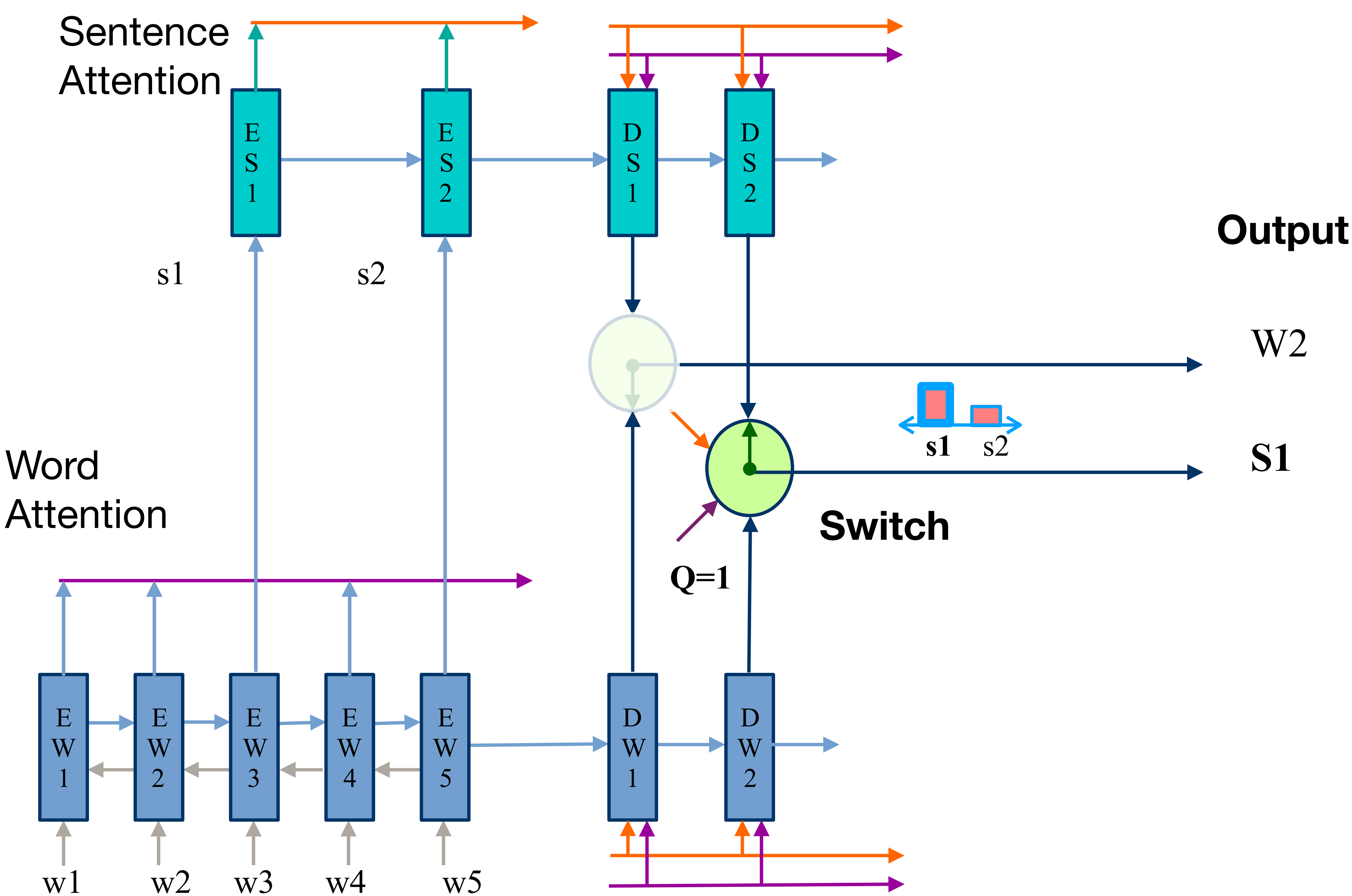


Switch has two states,

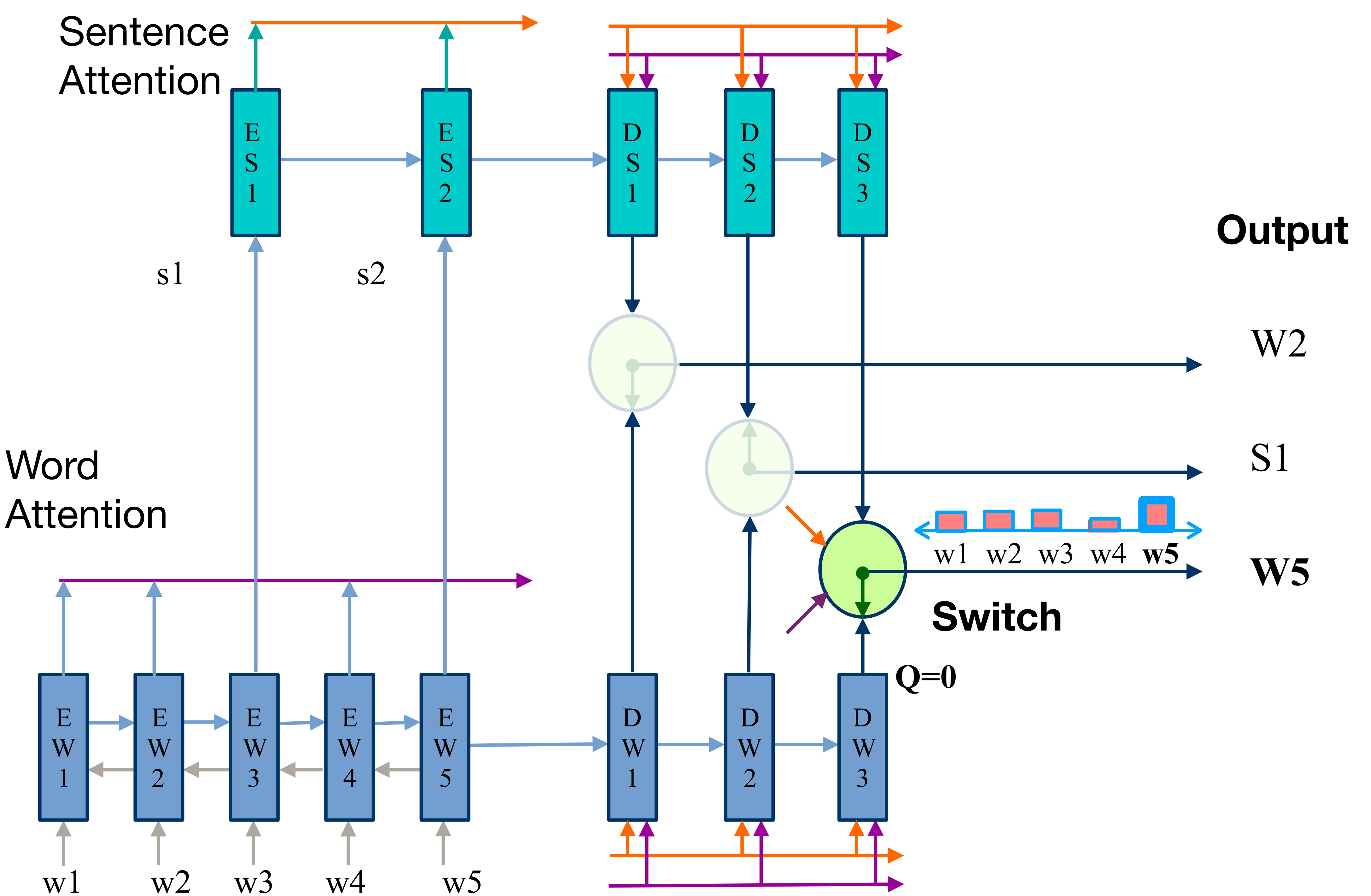
Q = 0 : word selection and

Q = 1 : sentence selection

Prediction with SWAP-NET: Decoding Step 2



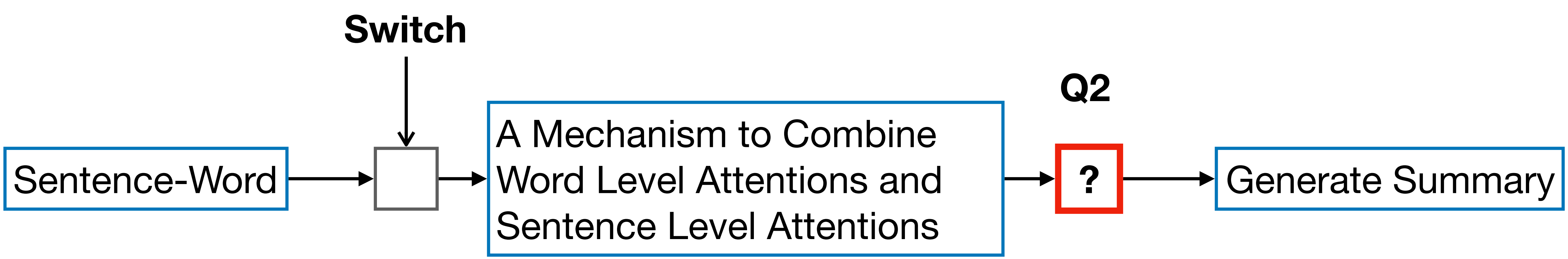
Prediction with SWAP-NET: Decoding Step 2



Questions

Q1 : *How can the two attentions be combined?*

Q2 : ***How can the summaries be generated considering both the attentions?***



Summary Generation

House prices across the UK will rise at a fraction of last year's frenetic pace, forecasts show.

KeyWord
Probability

Summary Generation

House **prices** across the UK will **rise** at a **fraction** of last year's **frenetic pace**, **forecasts show**

	prices	rise	fraction		frenetic	pace	forecasts	show
KeyWord Probability	P1	P2	P3		P4	P5	P6	P7

Probability of Sentence Ps

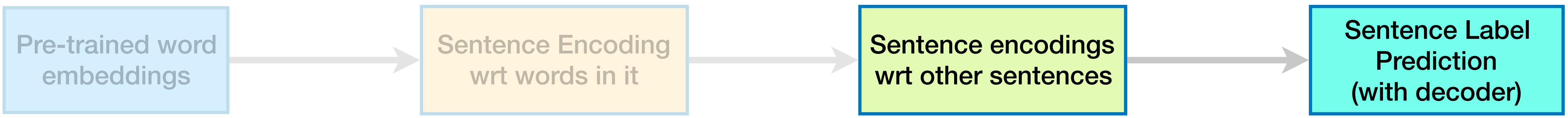
Score of Given Sentence = (Sentence Probability) + (Sum of its keyword Probabilities)

$$= P_s + \sum_{i=1}^{k} P_i$$ where k is number of keywords in sentence S

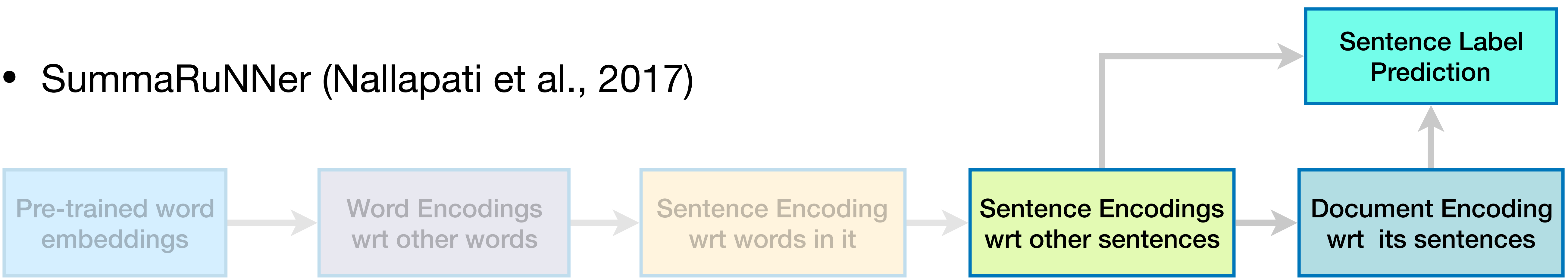
Top 3 sentences with maximum scores are chosen as summary

Extractive Summarization Methods

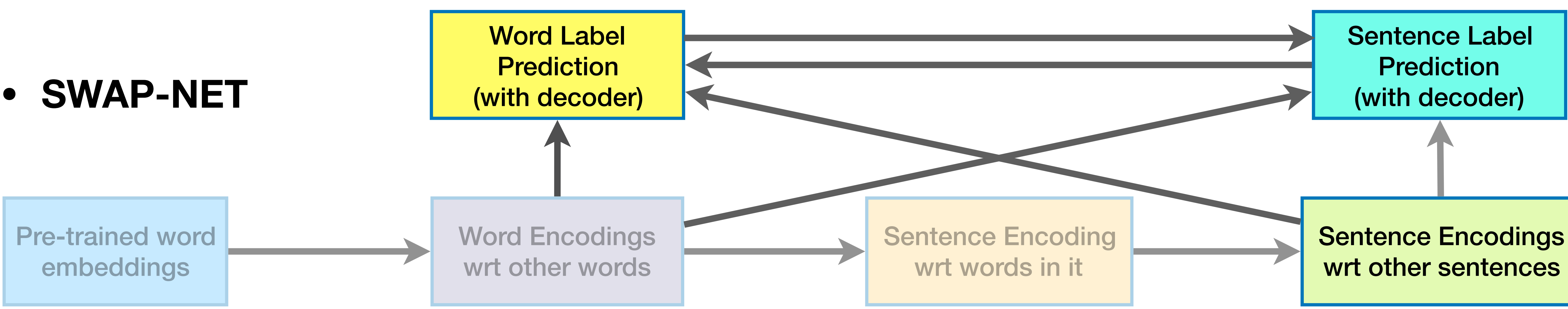
- NN (Cheng and Lapata, 2016)



- SummaRuNNer (Nallapati et al., 2017)



- **SWAP-NET**



Dataset and Evaluation

- **Large Benchmark Dataset CNN/DailyMail News Corpus**

News articles from CNN/DailyMail along with human generated summary (gold summary) for each article

- **GroundTruth Binary Labels For Training**

Sentences: Anonymised version of dataset given by (Cheng and Lapata, 2016)

Words: Extract keywords from each gold summary using RAKE

- **Number Labeled Documents**

Dataset	Training	Validation	Test
CNN	83568	1220	1093
Dailymail	193986	12147	10346

- **Standard Evaluation Metric: Three Variates of Rouge Score**

Comparing generated summaries and gold summaries for matching:

ROUGE-1 (R1): Unigrams

ROUGE-2 (R2): Bigrams

ROUGE-L (RL): Longest Common Subsequences

Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic key word extraction from individual documents. Text Mining: Applications and Theory.

Results

Performance on ***DailyMail Dataset*** using limited length **recall of Rouge**

	275 Bytes			**75 Bytes**		
Models	R1	R2	RL	R1	R2	RL
Lead-3	40.5	14.9	32.6	21.9	7.2	11.6
NN	42.2	17.3	34.8	22.7	8.5	12.5
SummaRuNNner-abs	40.4	15.5	32.0	23.8	9.6	13.3
SummaRuNNner	42.0	16.9	34.1	26.2	**10.8**	**14.4**
SWAP-NET	**43.6**	**17.7**	**35.5**	**26.4**	10.7	**14.4**

Results

Performance on ***CNN and Daily-Mail*** *test set* using the full length **Rouge F score**

Models	R1	R2	RL
Lead-3	39.2	15.7	35.5
ABS	35.4	13.3	32.6
SummaRuNNer-abs	37.5	14.5	33.4
SummaRuNNer	39.6	16.2	35.3
SWAP-NET	**41.6**	**18.3**	**37.7**

Example

Meet the four immigrant students each accepted to ALL EIGHT Ivy League schools who want to pay back their parents who moved to the U.S. to give them a better

PUBLISHED: 19:56 BST, 9

Their parents came to the U.S. for opportunities and now these four teens have them in abundance .
The high-achieving high schoolers have each been accepted to all eight Ivy League schools : Brown University , Columbia University , Cornell University , Dartmouth College , Harvard University , University of Pennsylvania , Princeton University and Yale University .
And as well as the Ivy League colleges , each of them has also been accepted to other top schools .
While they all grew up in different cities , the students are the offspring of immigrant parents who moved to America - from Bulgaria , Somalia or Nigeria .
And all four - Munira Khalif from Minnesota , Stefan Stoykov from Indiana , Victor Agbafe from North Carolina , and Harold Ekeh from New York - say they have their parents ' hard work to thank .
Now they hope to use the opportunities for good - whether its effecting positive social change , improving education across the world or becoming a neurosurgeon .
The teens have one more thing in common : they do n't know which school they 're going to pick yet .
The daughter of Somali immigrants who has already received a U.N. award and wants to improve education across the world
Star pupil : Munira Khalif , from St. Paul , Minnesota , says she has always been driven by the thought that her parents , who left Somalia during the civil war , fled to the U.S. so she would have better opportunities
Munira Khalif , who attends Mounds Park Academy in St. Paul , Minnesota , was shocked when she was accepted by eight Ivy Schools and three others - but her teachers were not .
\` She is composed and she is just articulate all the time , ' Randy Comfort , an upper school director at the private school , told KMSP . \` She 's pretty remarkable . '
The 18-year-old student , who was born and raised in Minnesota after her parents fled Somalia during the civil war , she said she was inspired to work hard because of the opportunities her family and the U.S. had given her .
\` The thing is , when you come here as an immigrant , you 're hoping to have opportunities not only for yourself , but for your kids , ' she told the channel . \` And that 's always been at the back of my mind . '
As well as achieving top grades , Khalif has immersed herself in other activities both in and out of school - particularly those aimed at doing good .
She was one of nine youngsters in the world to receive the UN Special Envoy for Global Education 's Youth Courage Award for her education activism , which she started when she was just 13 .

Summary Generated by SWAP-NET

Gold Summary

Munira_Khalif from Minnesota , Stefan_Stoykov from Indiana , Victor_Agbafe from North_Carolina , and Harold_Ekeh from New_York got multiple offers
All have immigrant parents - from Somalia , Bulgaria or Nigeria - and say they have their parents ' hard work to thank for their successes
They hope to use the opportunities for good , from improving education across the world to becoming neurosurgeons

SWAP-NET Predicted Keywords

Summary Generated by SWAP-NET

While they all grew up in different cities , the **students** are the offspring of **immigrant parents** who moved to America - from **Bulgaria , Somalia** or **Nigeria** .
And all four - **Munira_Khalif** from **Minnesota , Stefan_Stoykov** from Indiana , **Victor_Agbafe** from **North_Carolina** , and **Harold_Ekeh** from **New_York** - say they have their **parents** ' hard work to thank .
Now they hope to use the **opportunities** for **good** - whether its effecting **positive social** change , improving education across the **world** or becoming a **neurosurgeon**

SWAP-NET predictions highlighted in green

Keywords: Ground truth vs. SWAP-NET predictions

SWAP-NET key words (green) and Ground truth (blue)

While they all grew up in different cities , the **students** are the offspring of **immigrant parents** who moved to America - from **Bulgaria , Somalia** or **Nigeria** . And all four - **Munira_Khalif** from **Minnesota , Stefan_Stoykov** from Indiana , **Victor_Agbafe** from **North_Carolina** , and **Harold_Ekeh** from **New_York** - say they have their **parents** ' hard work to thank . **Now** they hope to use the **opportunities** for **good** - whether its effecting **positive social** change , improving education across the **world** or becoming a **neurosurgeon**
While they all grew up in different cities , the students are the offspring of **immigrant parents** who moved to America - from **Bulgaria** , **Somalia** or **Nigeria** . And all four - **Munira_Khalif** from **Minnesota** , **Stefan_Stoykov** from Indiana , **Victor_Agbafe** from **North_Carolina** , and **Harold_Ekeh** from **New_York** - say they have their **parents** ' **hard work** to thank . Now they **hope** to use the **opportunities** for **good** - whether its effecting positive social change , **improving education** across the **world** or becoming a **neurosurgeon**

Gold Summary

Munira_Khalif from **Minnesota** , **Stefan_Stoykov** from **Indiana** , **Victor_Agbafe** from **North_Carolina** , and **Harold_Ekeh** from **New_York** got multiple offers All have **immigrant parents** - from **Somalia** , **Bulgaria** or **Nigeria** - and say they have their **parents** ' **hard work** to thank for their successes They **hope** to use the **opportunities** for **good** , from **improving education** across the **world** to becoming **neurosurgeons**

Observations

- **Almost no keyword is repeated** across different sentence in the summary
- **Presence of key words** in all the overlapping segments of text with the gold summary
- Most of the **predicted keywords are actual keywords**
- Most of the extracted **summary sentences contain keywords**
- **Large proportion of key words** from the gold summary present in the generated summary

Summary Generated by SWAP-NET:

While they all grew up in different cities , the **students** are the offspring of **immigrant parents** who moved to America - from **Bulgaria , Somalia** or **Nigeria** .
And all four - **Munira_Khalif** from **Minnesota ,** **Stefan_Stoykov** from Indiana , **Victor_Agbafe** from **North_Carolina** , and **Harold_Ekeh** from **New_York** - say they have their **parents** ' hard work to thank .
Now they hope to use the **opportunities** for **good** - whether its effecting **positive social** change , improving education across the **world** or becoming a **neurosurgeon**

While they all grew up in different cities , the students are the offspring of **immigrant parents** who moved to America - from **Bulgaria** , **Somalia** or **Nigeria** .
And all four - **Munira_Khalif** from **Minnesota** , **Stefan_Stoykov** from Indiana , **Victor_Agbafe** from **North_Carolina** , and **Harold_Ekeh** from **New_York** - say they have their **parents** ' **hard work** to thank .
Now they **hope** to use the **opportunities** for **good** - whether its effecting positive social change , **improving education** across the **world** or becoming a **neurosurgeon**

Gold Summary:

Munira_Khalif from **Minnesota** , **Stefan_Stoykov** from **Indiana** , **Victor_Agbafe** from **North_Carolina** , and **Harold_Ekeh** from **New_York** got multiple offers
All have **immigrant parents** - from **Somalia** , **Bulgaria** or **Nigeria** - and say they have their **parents** ' **hard work** to thank for their successes
They **hope** to use the **opportunities** for **good** , from **improving education** across the **world** to becoming **neurosurgeons**

Experiments

- **Key word coverage** measures the proportion of key words from those in the gold summary present in the generated summary
- **Sentences with key words** measures the proportion of sentences containing at least one key word

Statistics	Lead-3	SWAP-NET
KW coverage	61.6%	73.8%
Sentences with KW	92.2%	98%

- **Average pairwise cosine distance** between paragraph vector representations of sentences in summaries to **measure semantic redundancy** in summaries

Gold summary	Lead-3	SWAP-NET
0.81	0.553	0.8

SWAP-NET summaries are similar in redundancy to the Gold summary

Highlights the importance of key words in finding salient sentences for extractive summaries

Conclusion

- We develop SWAP-NET, a neural sequence-to- sequence model for extractive summarization
- By **effective modelling of interactions between sentences and key words**, SWAP- NET **outperforms state-of-the-art** extractive single-document summarizers
- SWAP-NET models these interactions using a new two-level pointer network based architecture with **a switching mechanism**
- Experiments suggest that modelling sentence-keyword interaction has the desirable property of **less semantic redundancy in summaries** generated by SWAP-NET

An implementation of SWAP-NET and generated summaries from the test sets are available online:
https://github.com/aishj10/swap-net