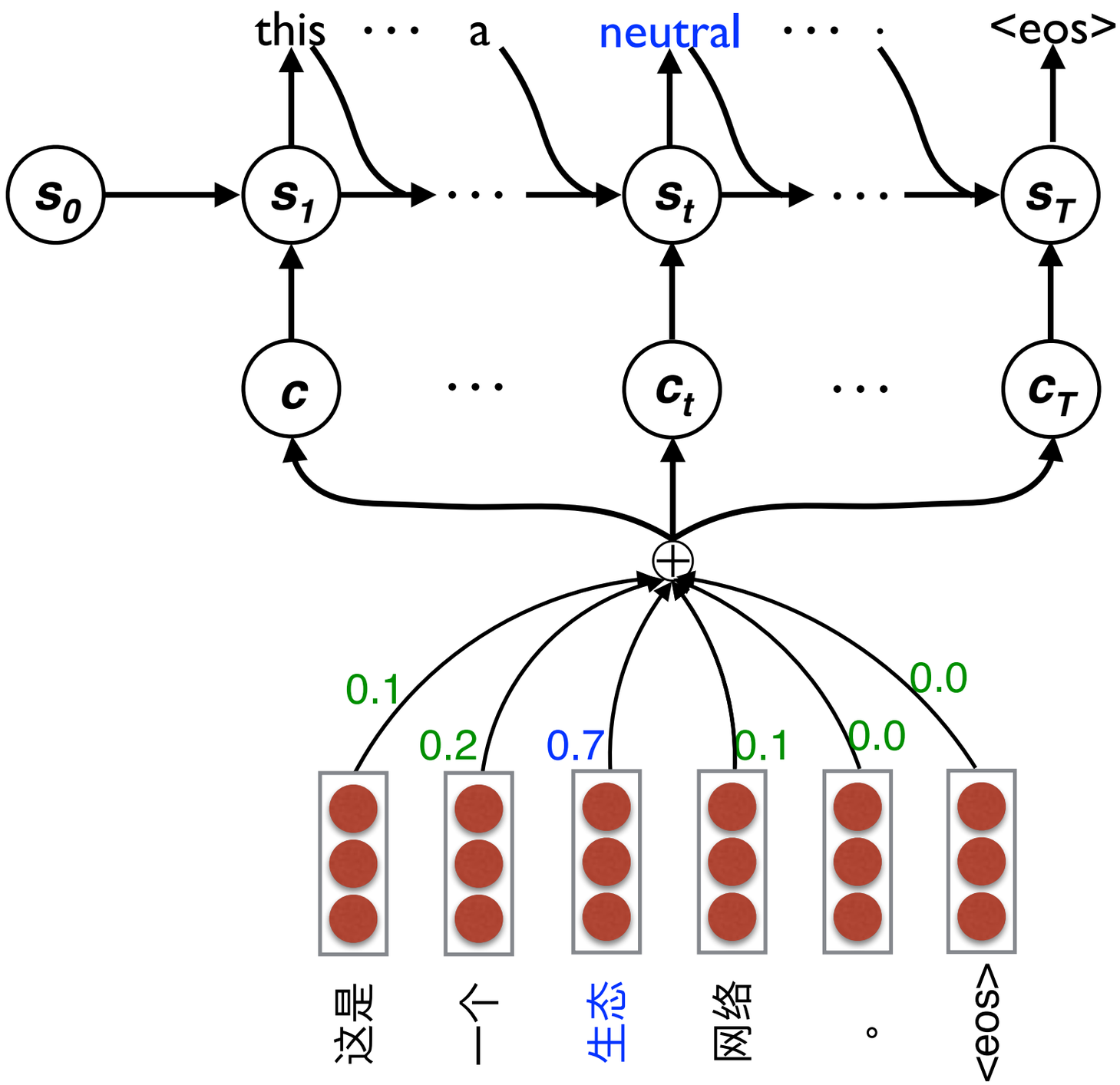
# *Exploiting Cross-Sentence Context for Neural Machine Translation*

Longyue Wang♥ **Zhaopeng Tu♠** Andy Way♥ Qun Liu♥

♥ ADAPT Centre, Dublin City University ♠ Tencent AI Lab

# Motivation

- The majority of NMT models is sentence-level



# Motivation

- The continuous vector representation of a symbol encodes ***multiple dimensions of similarity***.

| Word $x'$ | Axis | Nearest Neighbours |
|---|---|---|
| notebook | 1 | diary notebooks (notebook) sketchbook jottings |
| | 2 | palmtop notebooks (notebook) ipaq laptop |
| power | 1 | powers authority (power) powerbase sovereignty |
| | 2 | powers electrohydraulic microwatts hydel (power) |

(Choi et al., 2016)

# Motivation

- The continuous vector representation of a symbol encodes ***multiple dimensions of similarity***.
- ***Consistency*** is another critical issue in document-level translation.

| | |
|---|---|
| **Past** | 那么 在 这个 **问题** 上 ， 伊朗 的 …<br>well, on this **issue** , iran has a relatively …<br>在 任内 解决 伊朗 核 **问题** ， 不管是 用 和平 …<br>to resolve the iranian nuclear **issue** in his term , … |
| **Current** | 那 刚刚 提到 这个 … 谈判 的 **问题** 。<br>that just mentioned the **issue** of the talks … |

# Motivation

- The ***cross-sentence context*** has proven helpful for the aforementioned two problems in multiple sequential tasks (Sordoni et al., 2015; Vinyals and Le, 2015; Serban et al., 2016).

# Motivation

- The ***cross-sentence context*** has proven helpful for the aforementioned two problems in multiple sequential tasks (Sordoni et al., 2015; Vinyals and Le, 2015; Serban et al., 2016).
- However, it has ***received relatively little attention*** from the NMT research community.

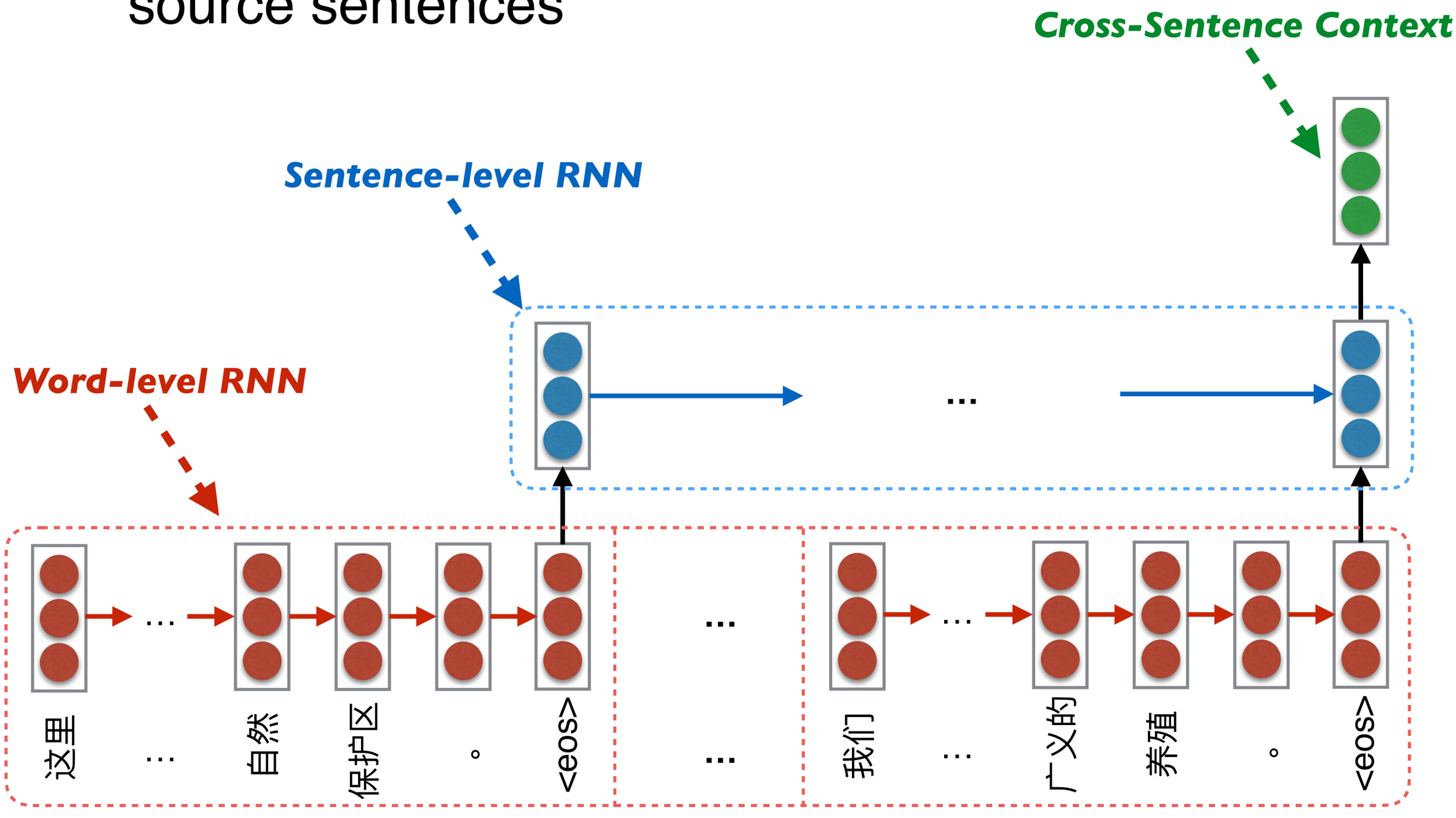
# Data and Setting

- Chinese-English translation task
- Training data: 1M sentence pairs from LDC corpora that contain document information
- Tuning: NIST MT05, Test: NIST MT06 and MT08

- Build the model on top of Nematus (*https://github.com/EdinburghNLP/nematus*)
- Vocabulary size: 35K for both languages
- Word embedding: 600; Hidden size: 1000

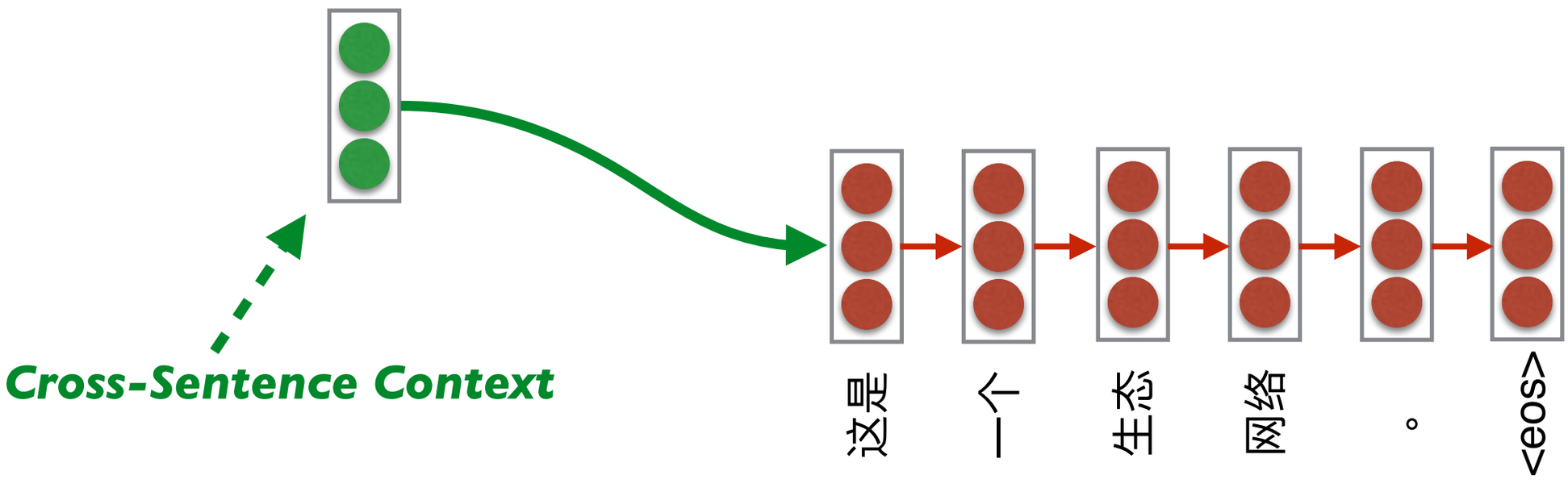
# Approach

- Use a ***Hierarchical RNN*** to summarize previous M source sentences



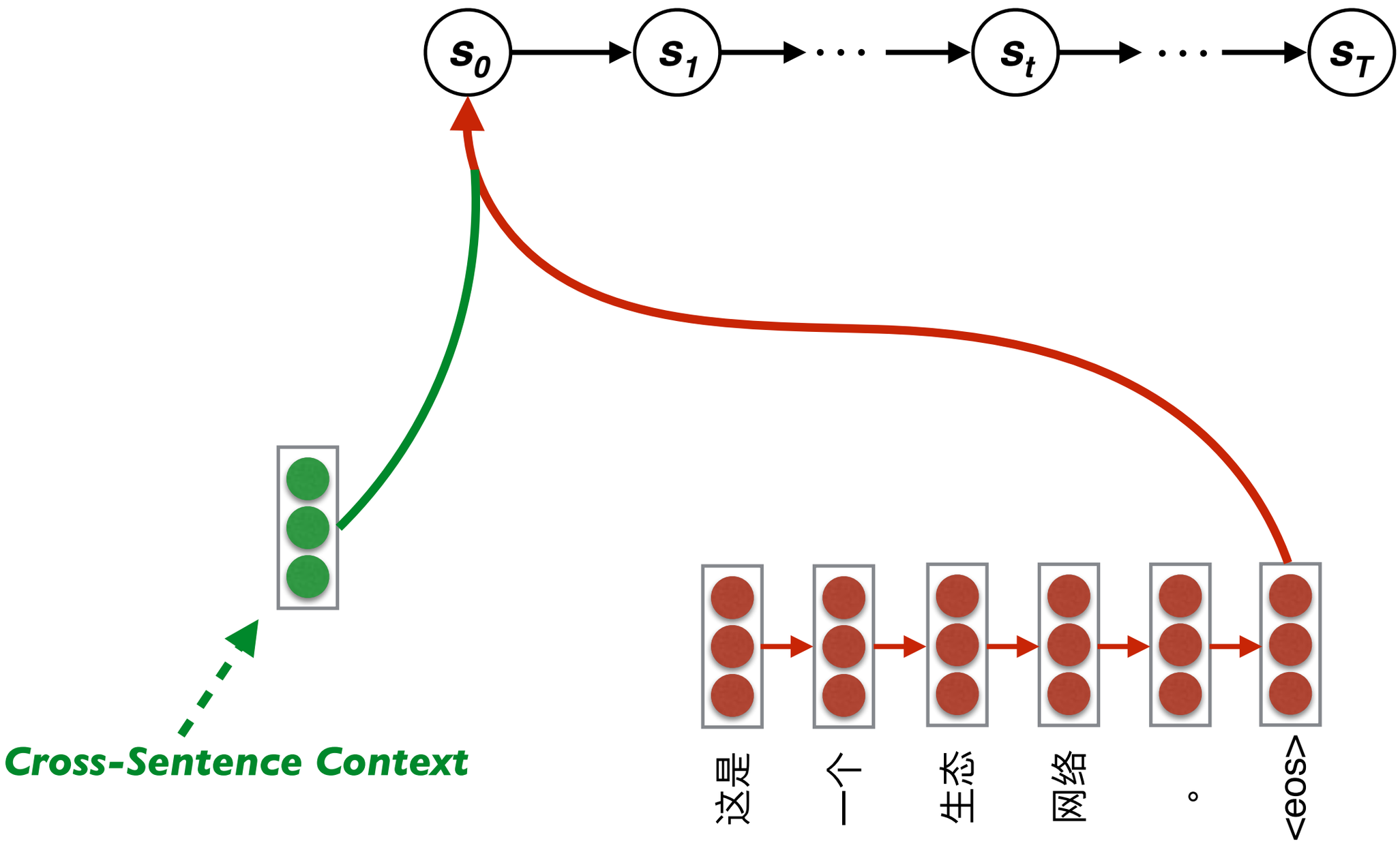
# Approach

- Strategy I: **Initialization** — *Encoder*



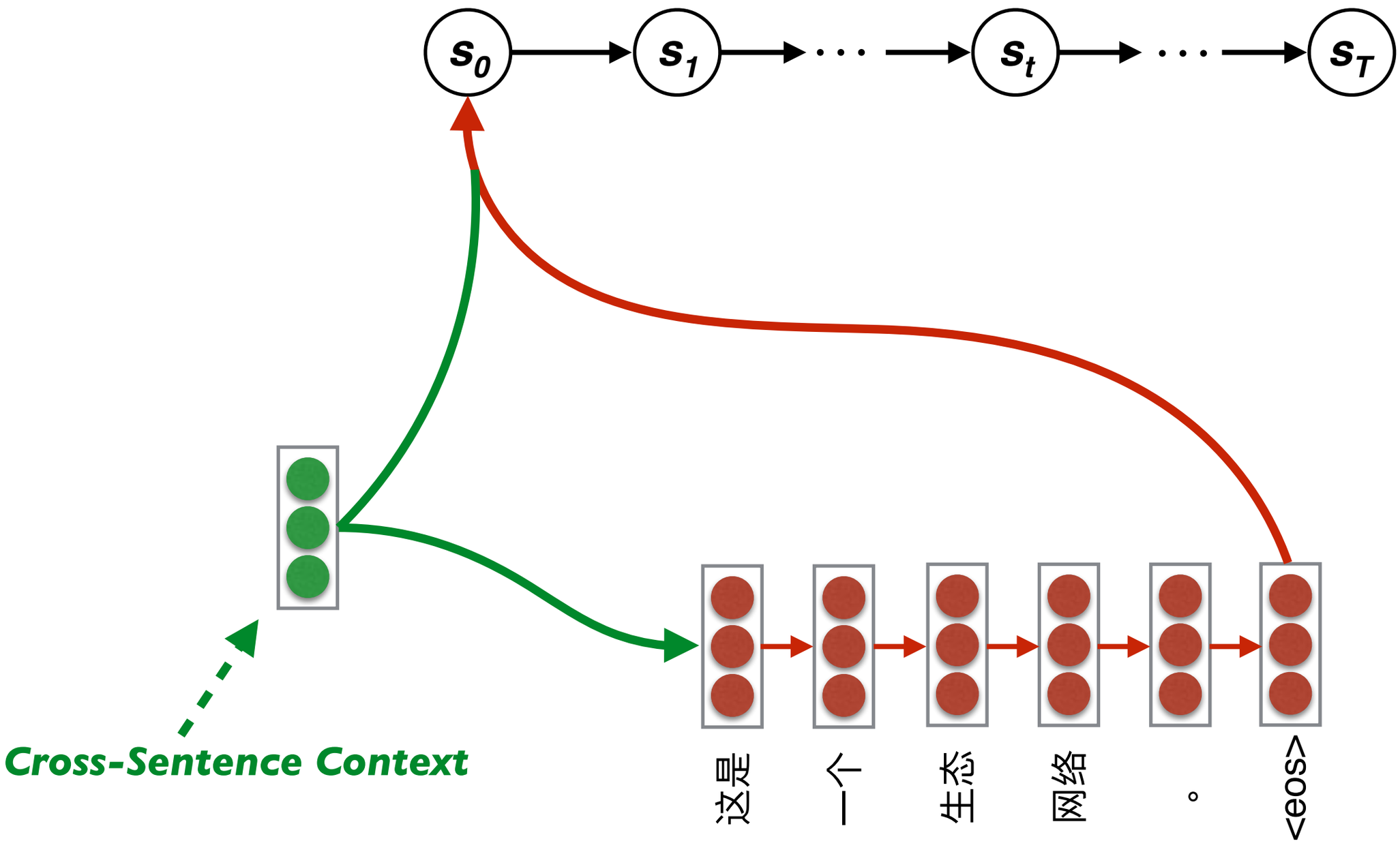
# Approach

- Strategy I: **Initialization** — *Decoder*



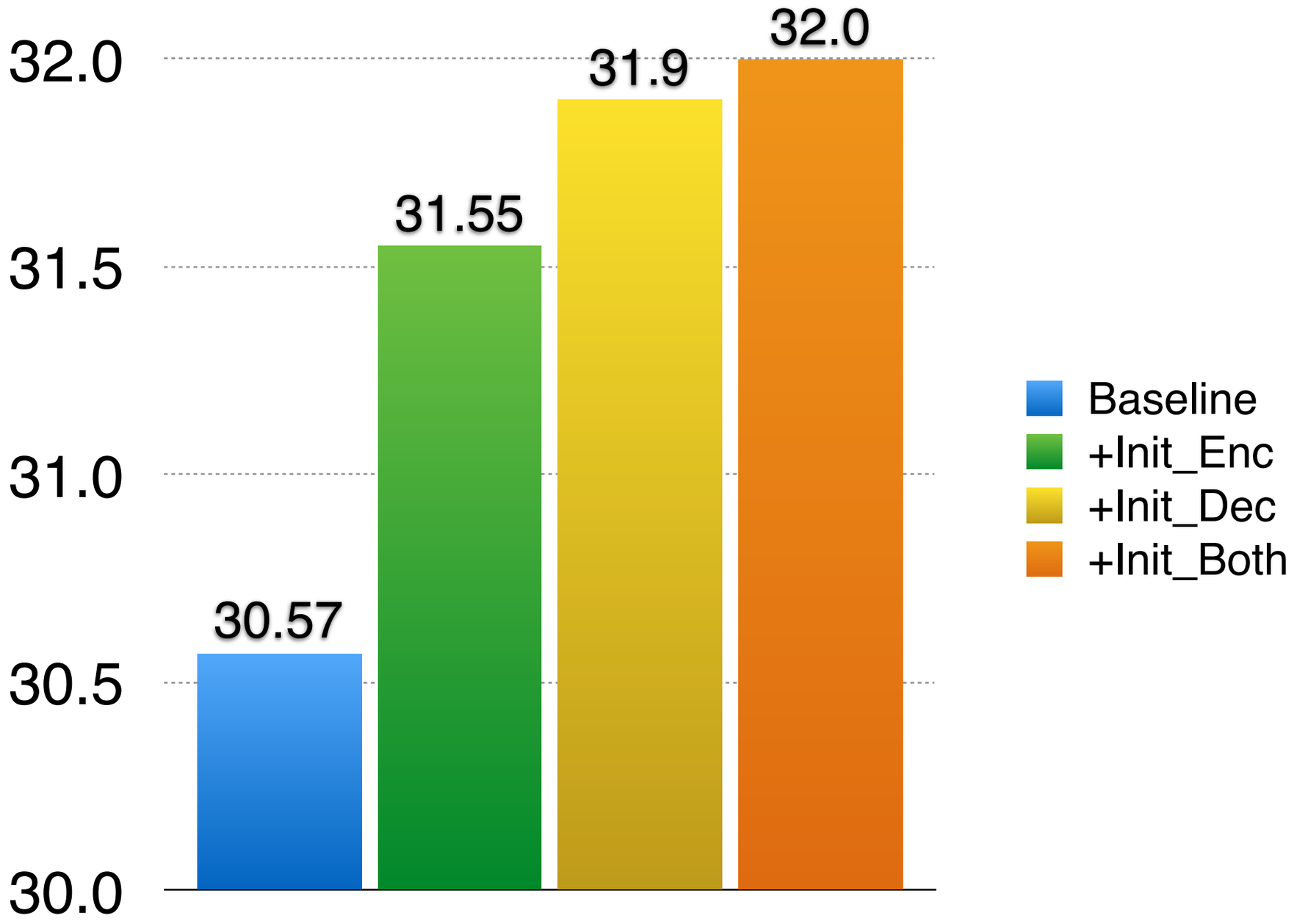
# Approach

- Strategy I: **Initialization** — *Both*



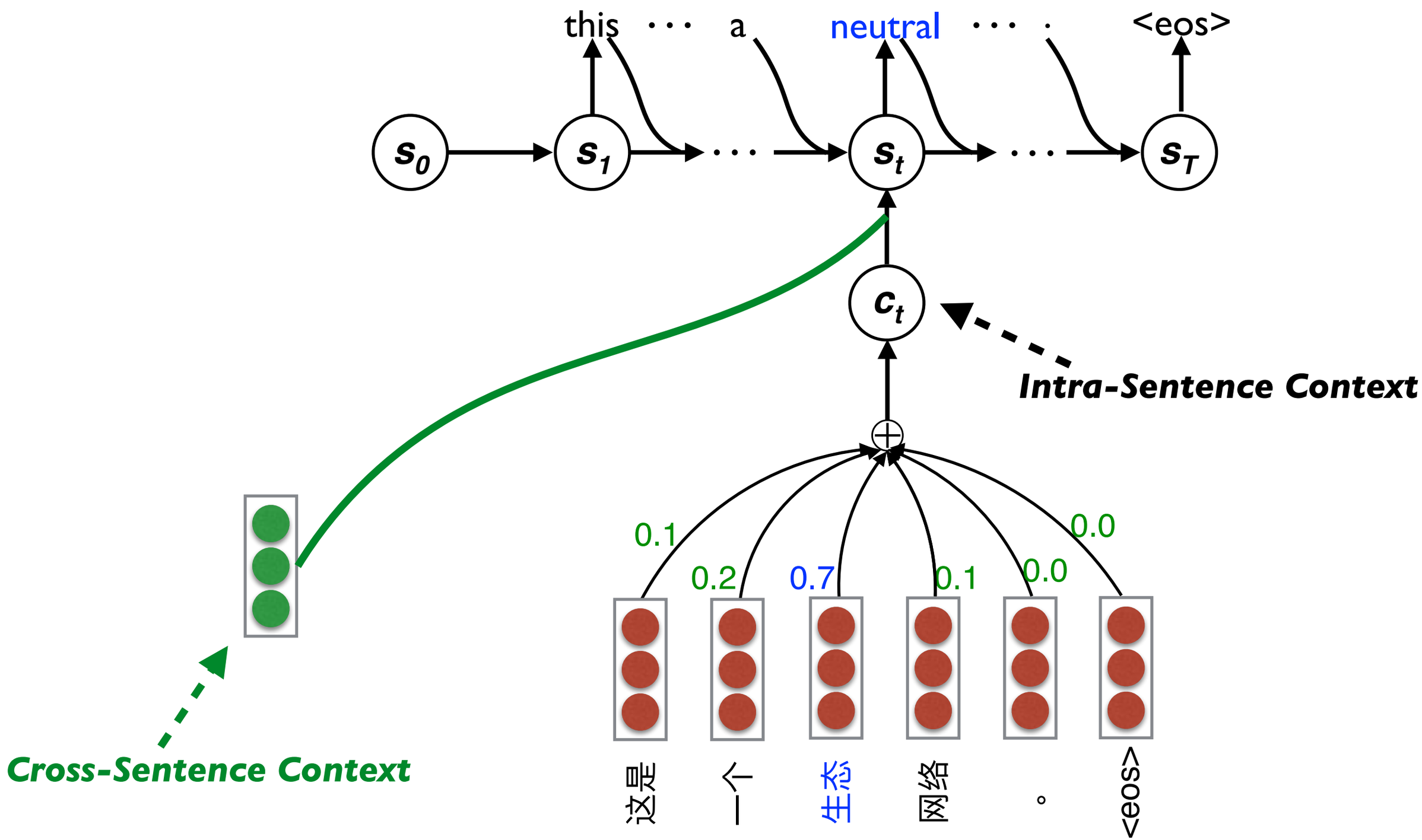
# Results

- Impact of components



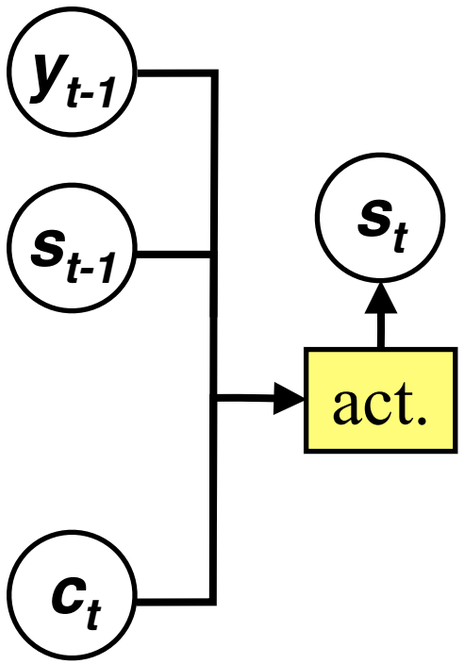
# Approach

- Strategy 2: Auxiliary Context

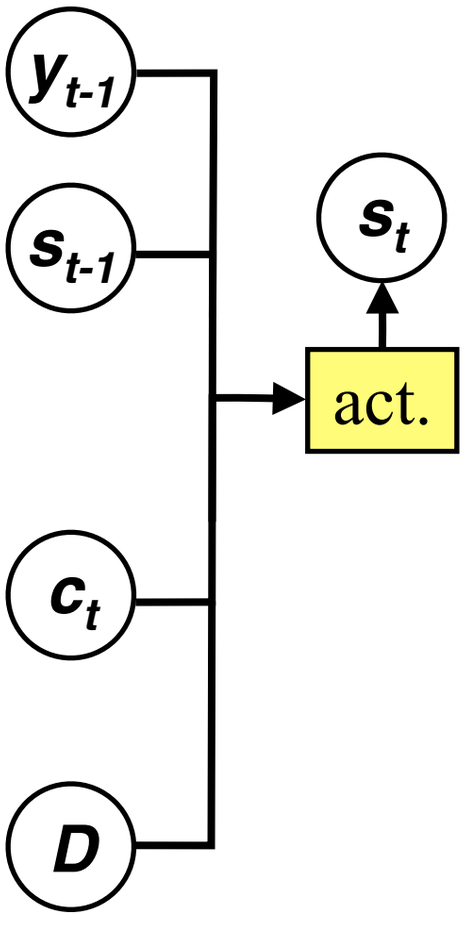


# Approach

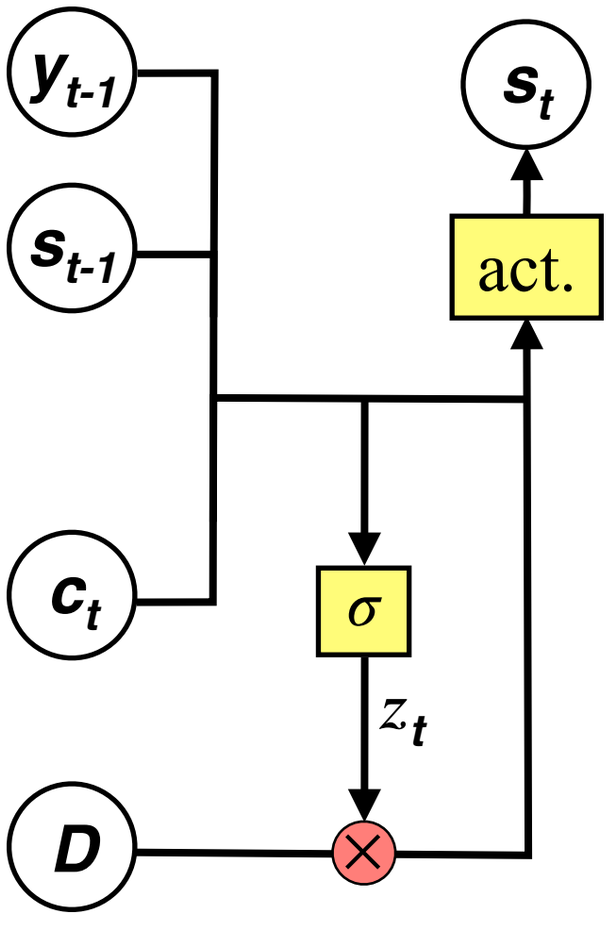
- Strategy 2: Auxiliary Context




(a) standard decoder



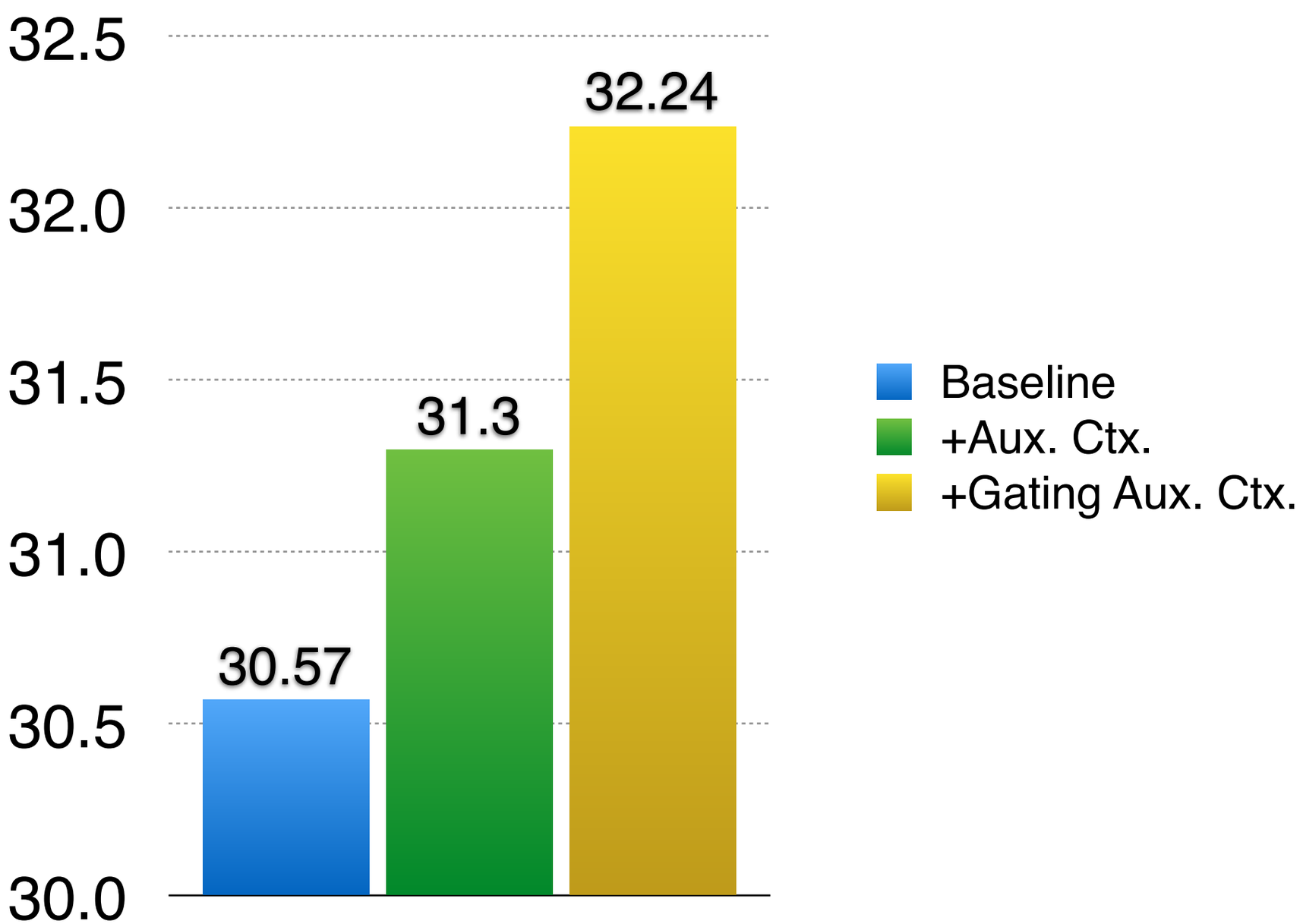

(b) decoder with auxiliary context




(c) decoder with gating auxiliary context

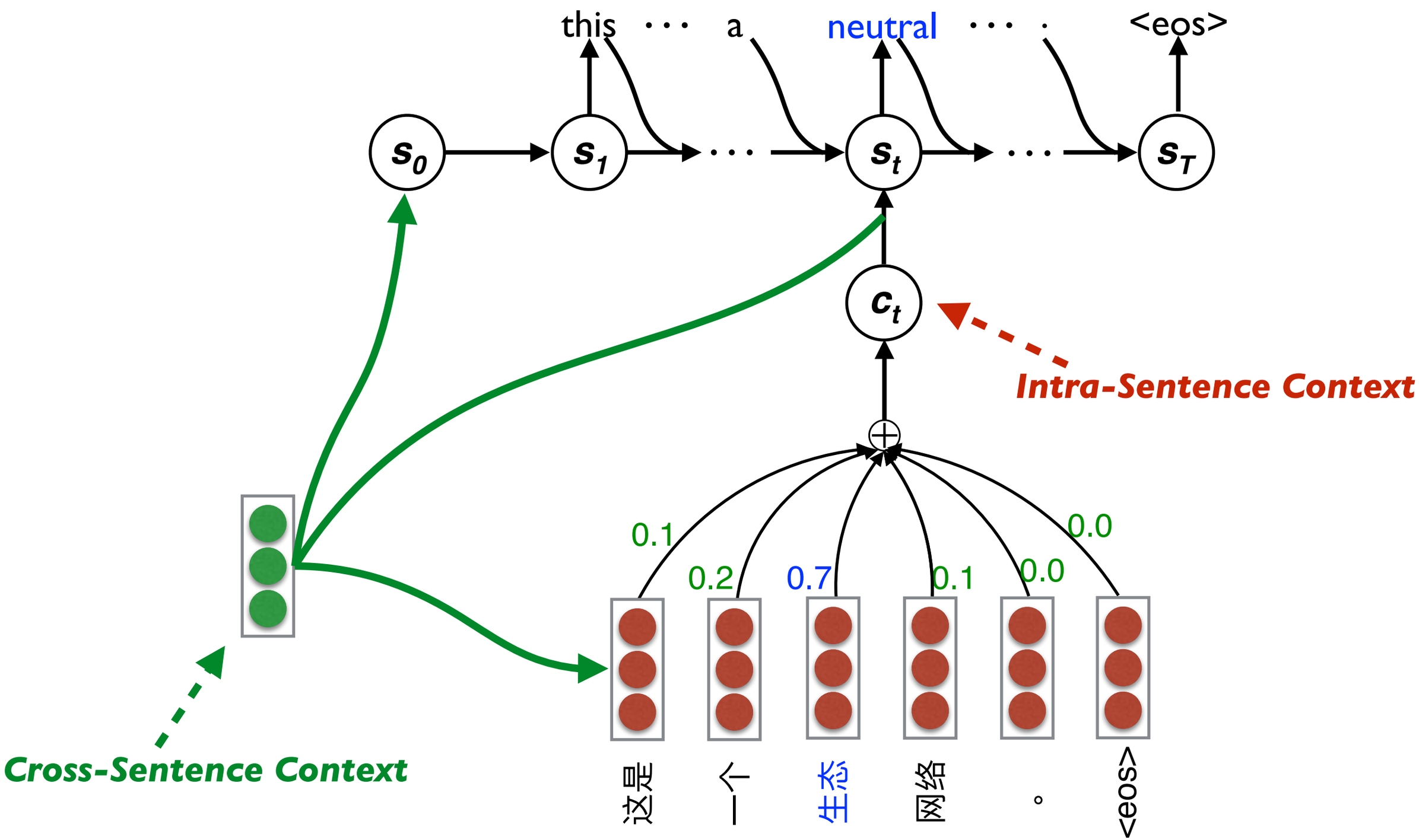
# Results

- Impact of components



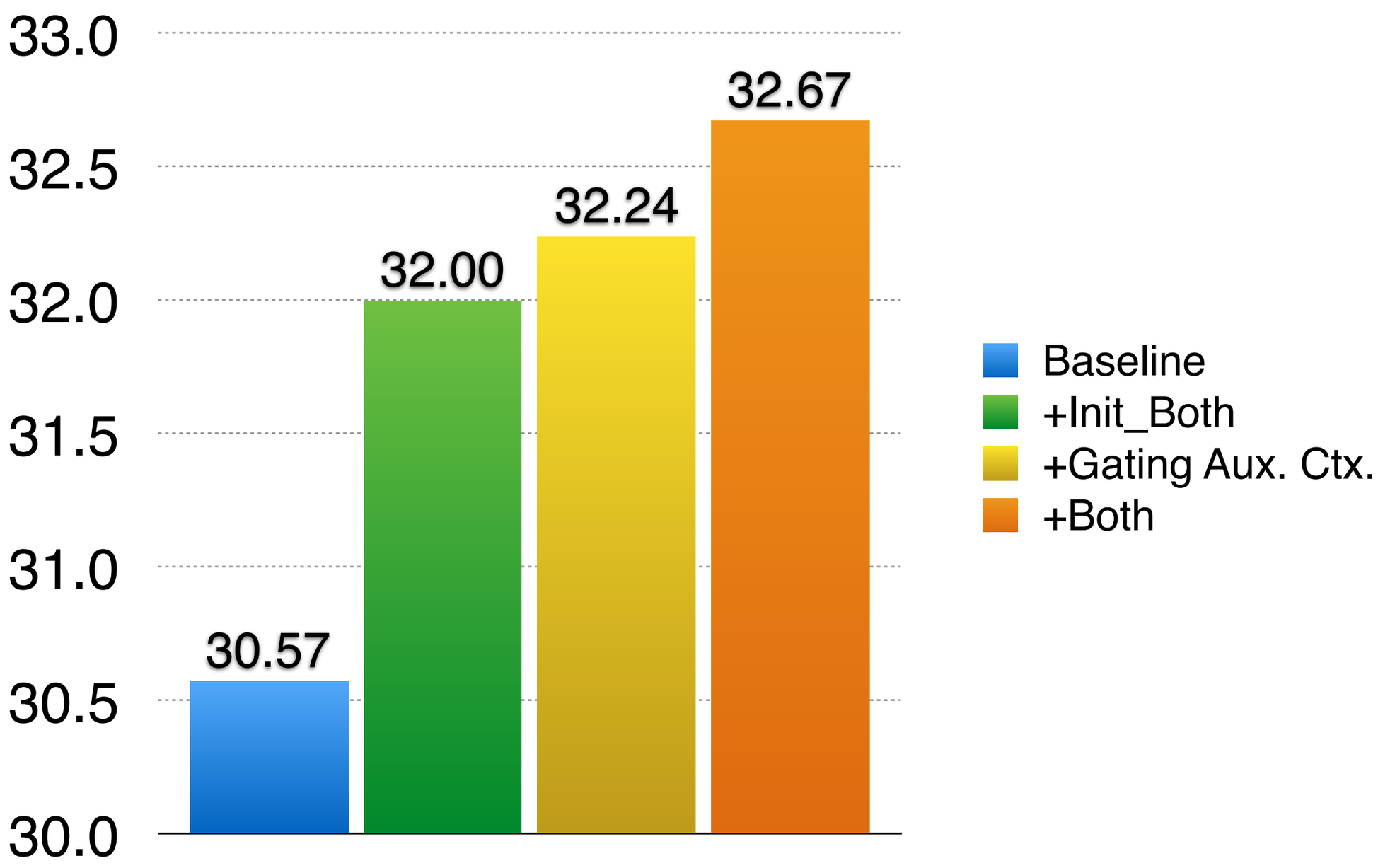
# Approach

- Initialization + Gating Auxiliary Context



# Results

- Impact of components



# Analysis

- Translation error statistics

| Errors | Ambiguity | Inconsistency | All |
|---|---|---|---|
| **Total** | 38 | 32 | 70 |
| **Fixed** | 29 | 24 | 53 |
| **New** | 7 | 8 | 15 |

# Analysis

- Case Study

| | |
|---|---|
| Hist. | 这 不 等于 明着 提前 告诉 贪官 们 赶紧 转移 罪证 吗 ? |
| Input | 能否 遏制 和 震慑 腐官 ? |
| Ref. | Can it inhibit and deter corrupt officials? |
| NMT | Can we contain and deter the *enemy*? |
| Our | Can it contain and deter the **corrupt officials**? |

# Summary

- We propose to use HRNN to summary previous source sentences, which aims at providing cross-sentence context for NMT

- Limitations
  - Computational expensive
  - Only exploit source sentences due to error propagation
  - Encoded into a single fixed-length vector, not flexible

# Publicly Available

- The source code is publicly available at *https://github.com/tuzhaopeng/LC-NMT*
- The trained models and translation results will be released

# Reference


1. Heeyoul Choi, Kyunghyun Cho, and Yoshua Bengio. *Context-dependent word representation for neural machine translation*. arXiv 2016.

2. Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen, and Jian- Yun Nie. A hierarchical recurrent encoder- decoder for generative context-aware query suggestion. CIKM 2015.

3. Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. AAAI 2016.

4. Oriol Vinyals and Quoc Le. A neural conversa- tional model. In Proceedings of the International Conference on Machine Learning, Deep Learning Workshop.

# Question & Answer