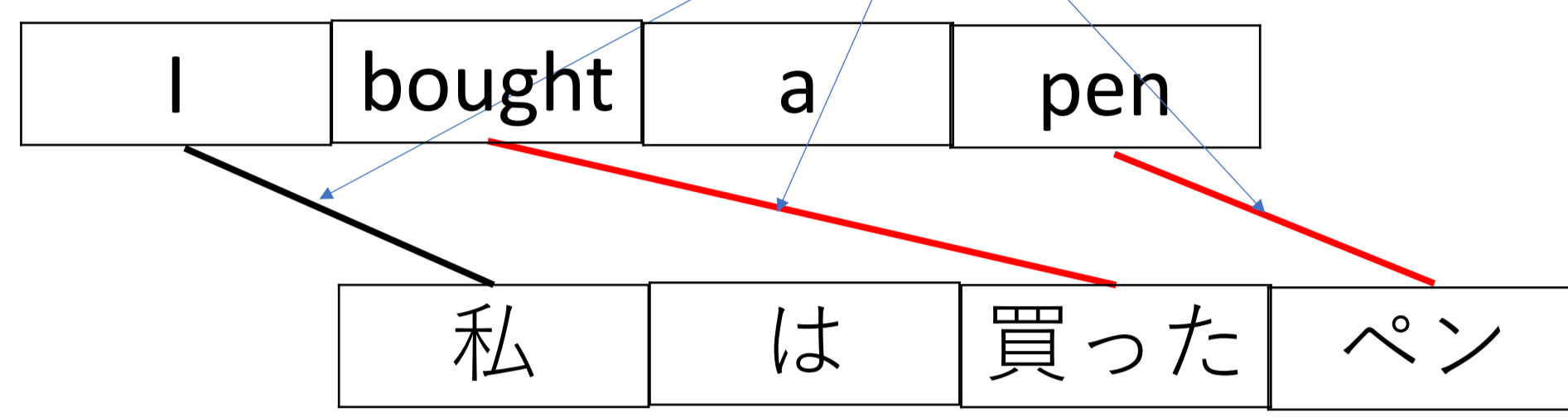




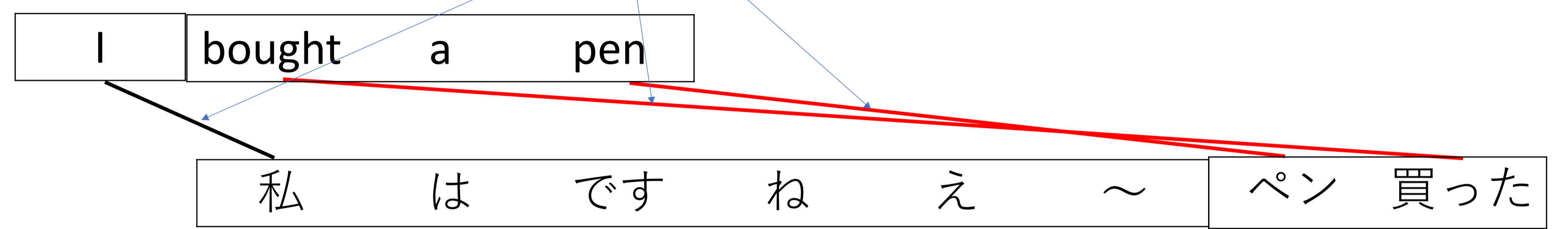
- **EVS (Ear-Voice-Span)**を測るには、音声認識、タイムスタンプ、アラインメントの取得が必要。
- **AL (Average Lagging)**という、よりシンプルな遅延評価尺度が自動同時通訳評価で使われてきた。
- **問題**は、ALは翻訳出力が長くなると遅延が小さく評価されること。
- **提案尺度ATD (Average Token Delay)**は、EVSにインスパイアされ、その問題を解決。
- **今後**は遅延の自動評価と人手評価の相関を取って評価精度を確かめることが重要。

EVS計算

Case 1: Avg. EVS = (1 + 2 + 1) / 3 = **1.3**



Case 2: Avg. EVS = (1 + 7 + 4) / 3 = **4.0**



出力:長→遅延:大

AL 定義

[Ma+, 2019]
$$AL = \frac{1}{\tau(|x|)} \sum_{t=1}^{\tau(|x|)} \left(g(t) - \frac{t-1}{\gamma} \right)$$

Average Lagging

Delay Catch-up Lagging

$x = x_1, x_2, \dots, x_{|x|}$: Input tokens

$y = y_1, y_2, \dots, y_{|y|}$: Output tokens

$$\gamma = \frac{|y|}{|x|}$$

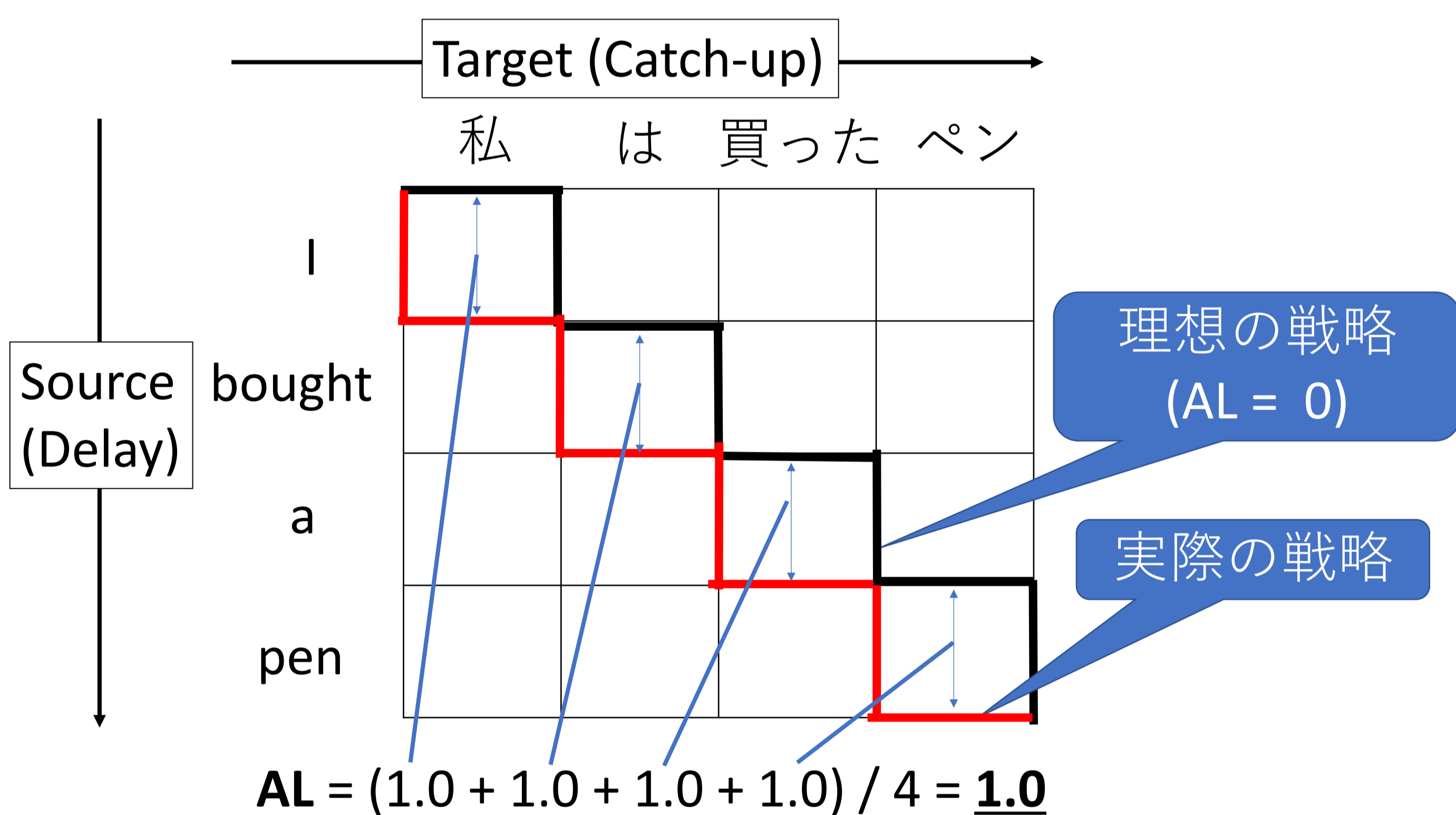
t : Target token index

$g(t)$: Number of input tokens read to output t th target token

$$\tau(|X|) = \min(t \mid g(t) = |x|)$$

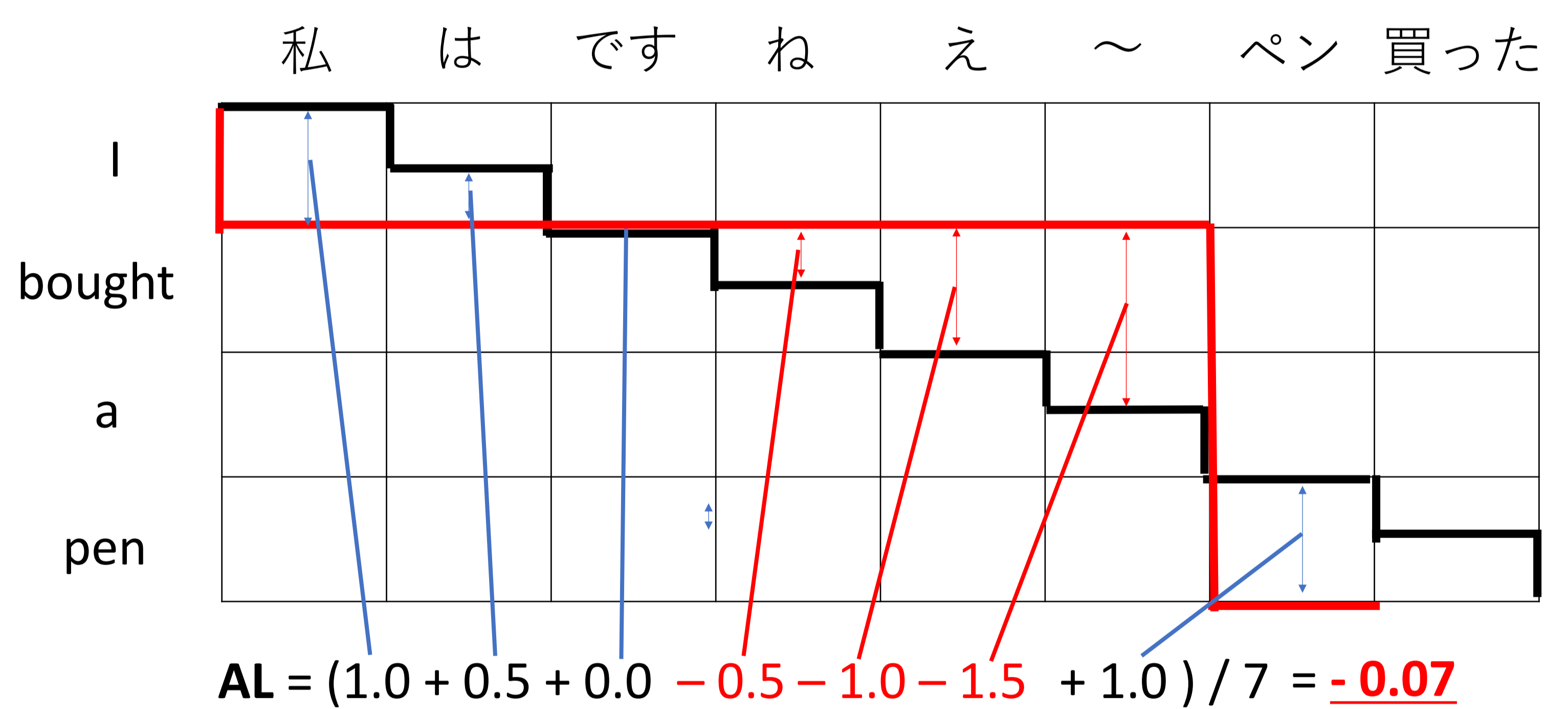
AL 計算

Case 1:



Case 3:

出力:長→遅延:小



ATD 定義

[Kano+, 2023]
$$ATD = \frac{1}{|y|} \sum_{t=1}^{|y|} (T(y_t) - T(x_{a(t)}))$$

Average Token Delay

Target token time Source token time Token Delay

$x = x_1, x_2, \dots, x_{|x|}$: Input tokens

$y = y_1, y_2, \dots, y_{|y|}$: Output tokens

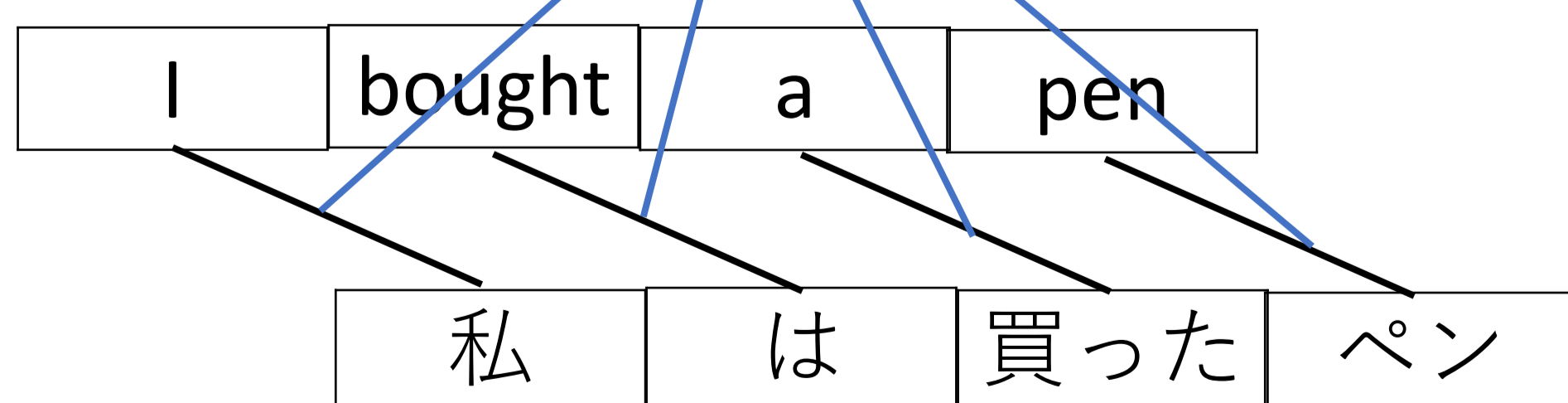
t : Target token index

$a(t)$: Source token index corresponding to target token index

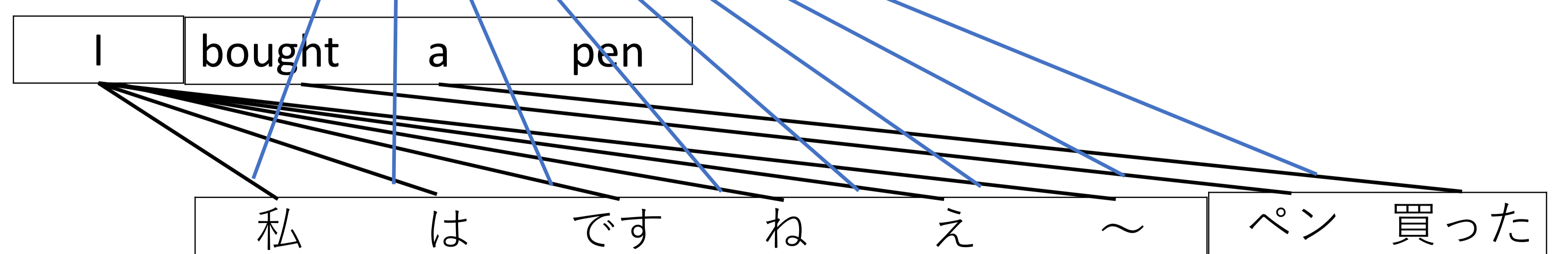
$T(*)$: Ending time of each token

ATD計算

Case 1: ATD = (1 + 1 + 1 + 1) / 4 = **1.0**



Case 2: ATD = (1 + 2 + 3 + 4 + 5 + 6 + 6 + 6) / 7 = **4.1**



出力:長→遅延:大

EVSは意味的対応をとるが、ATDは入出力タイミング情報のみから簡易的な対応をとる。

[Ma+, 2019] Mingbo Ma, Liang Huang, Hao Xiong, Renjie Zheng, Kaibo Liu, Baigong Zheng, Chuanqiang Zhang, Zhongjun He, Hairong Liu, Xing Li, Hua Wu, and Haifeng Wang. 2019. STACL: Simultaneous translation with implicit anticipation and controllable latency using prefix-to-prefix framework. In Proceedings of 57th ACL.

[Kano+, 2023] Yasumasa Kano, Katsuhito Sudoh, and Satoshi Nakamura.

2023. Average token delay: A latency metric for simultaneous translation. In Proceedings of Interspeech 2023.