

# Языковое моделирование

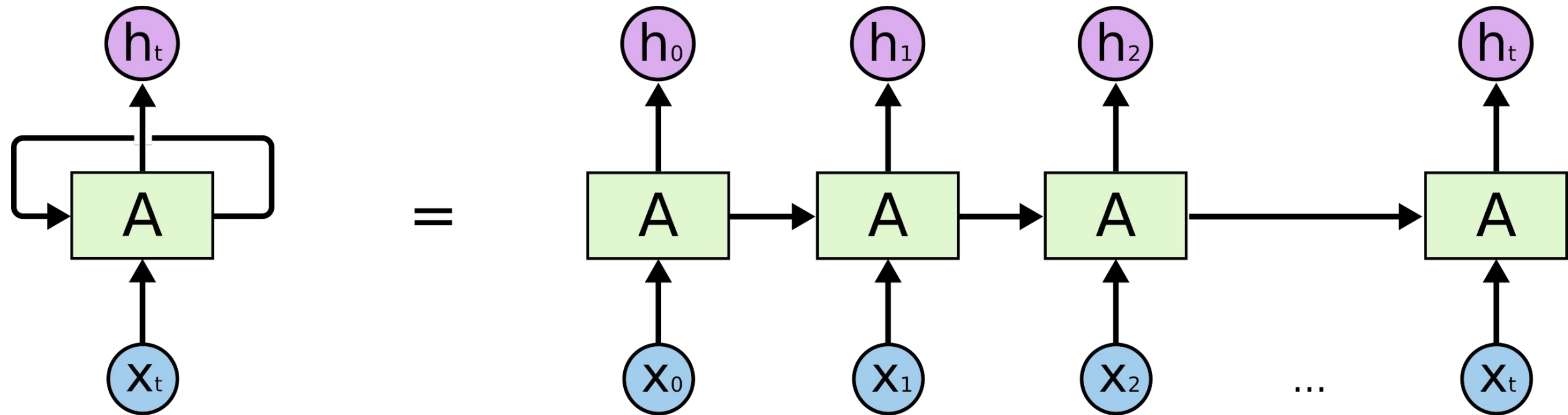
## Лекция 3

Решение задачи языкового моделирования с помощью RNN

# План занятия

1. Напоминание работы RNN
2. Алгоритм обучения RNN для решения задачи языкового моделирования
3. Плюсы и минусы RNN

# RNN



# Как обучить языковую модель с помощью RNN

1. Взять большой корпус текста. Разбить его на отдельные последовательности слов  $x_0, \dots, x_k$

# Как обучить языковую модель с помощью RNN

1. Взять большой корпус текста. Разбить его на отдельные последовательности слов  $x_0, \dots, x_k$
2. Подать эти последовательности на вход в RNN. Рассчитать вероятностное распределение  $\hat{y}_0, \dots, \hat{y}_k$

# Как обучить языковую модель с помощью RNN

1. Взять большой корпус текста. Разбить его на отдельные последовательности слов  $x_0, \dots, x_k$
2. Подать эти последовательности на вход в RNN. Рассчитать вероятностное распределение  $\hat{y}_0, \dots, \hat{y}_k$
3. Посчитать лосс функцию — кросс-энтропию между предсказанием и истинным распределением  $y_0, \dots, y_k$ . Каждый  $y_i$  является one-hot вектором для  $x_{i+1}$

# Как обучить языковую модель с помощью RNN

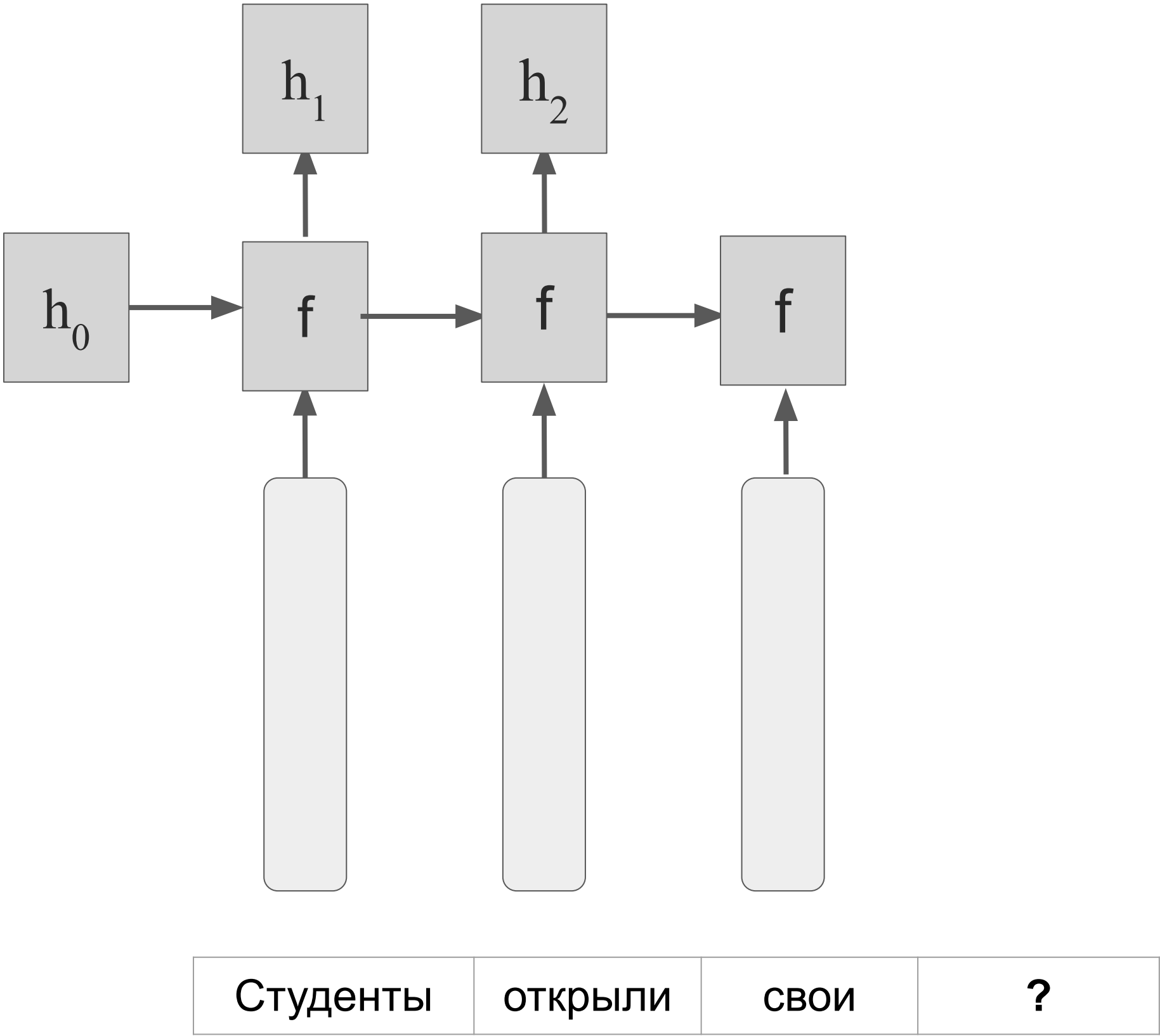
1. Взять большой корпус текста. Разбить его на отдельные последовательности слов  $x_0, \dots, x_k$
2. Подать эти последовательности на вход в RNN. Рассчитать вероятностное распределение  $\hat{y}_0, \dots, \hat{y}_k$
3. Посчитать лосс функцию — кросс-энтропию между предсказанием и истинным распределением  $y_0, \dots, y_k$ . Каждый  $y_i$  является one-hot вектором для  $x_{i+1}$
4. Оптимизировать градиентным спуском

# Пример

Скрытые состояния

Эмбеддинги

Слова



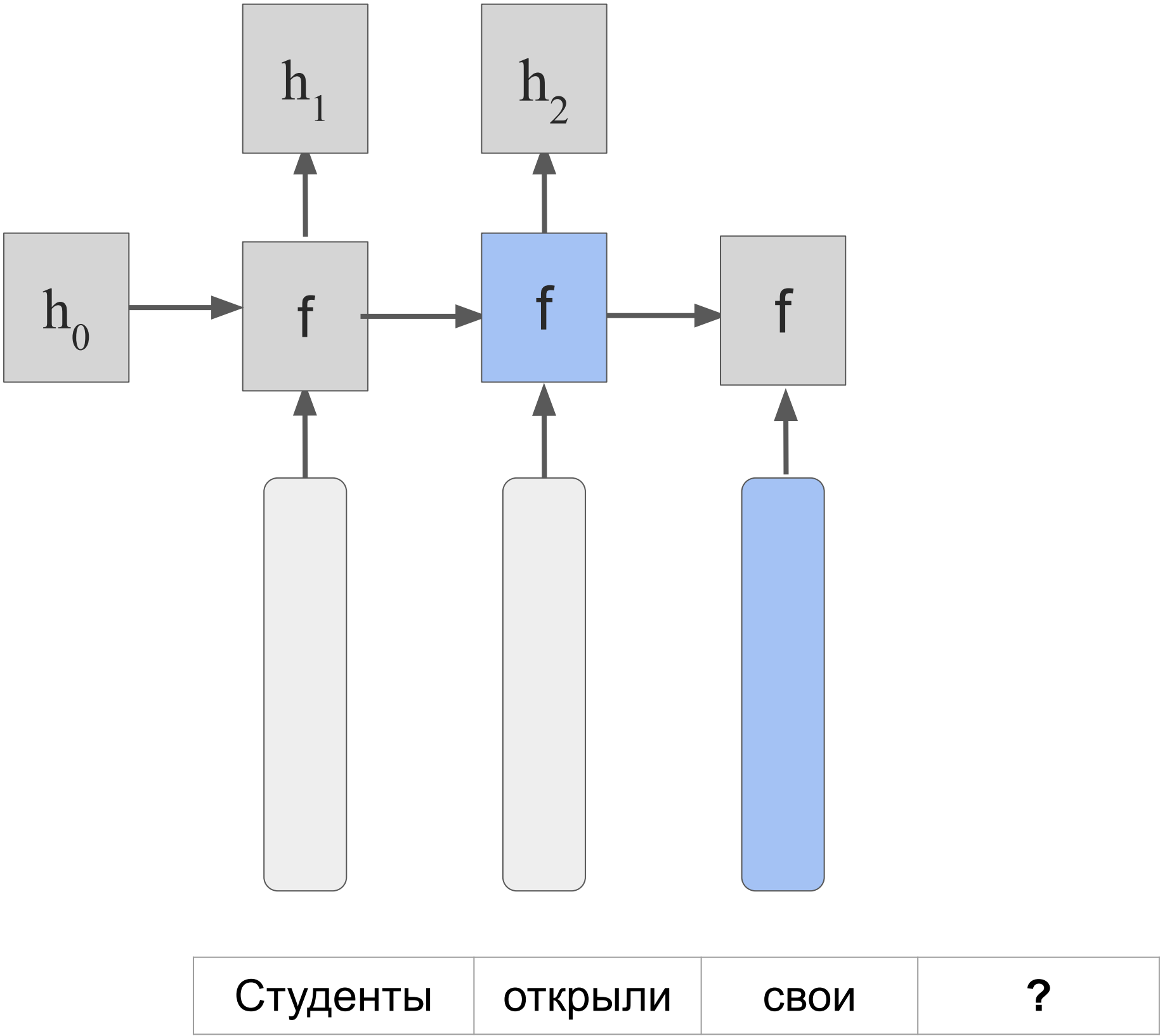


# Пример

Скрытые состояния

Эмбеддинги

Слова

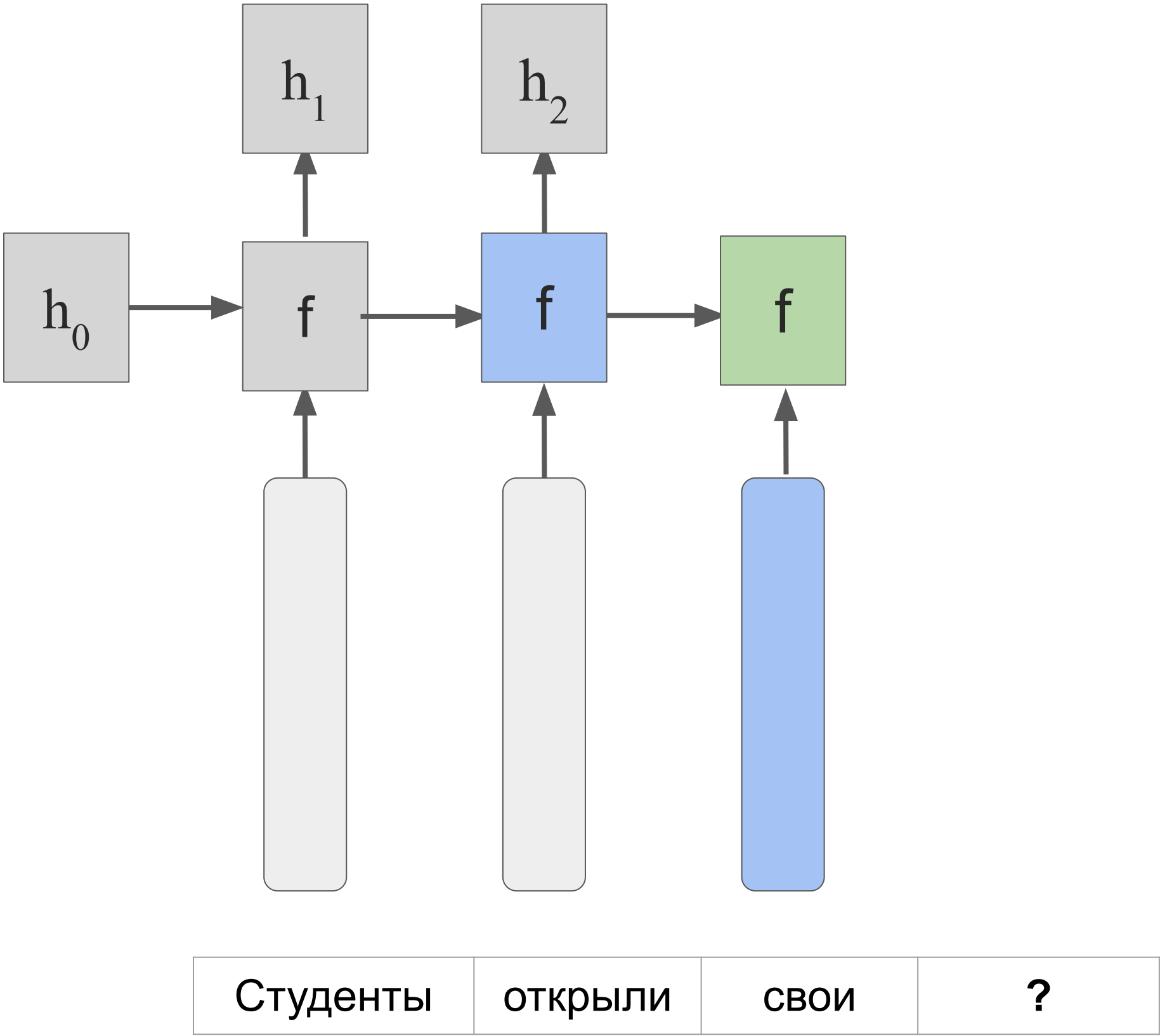


# Пример

Скрытые состояния

Эмбеддинги

Слова

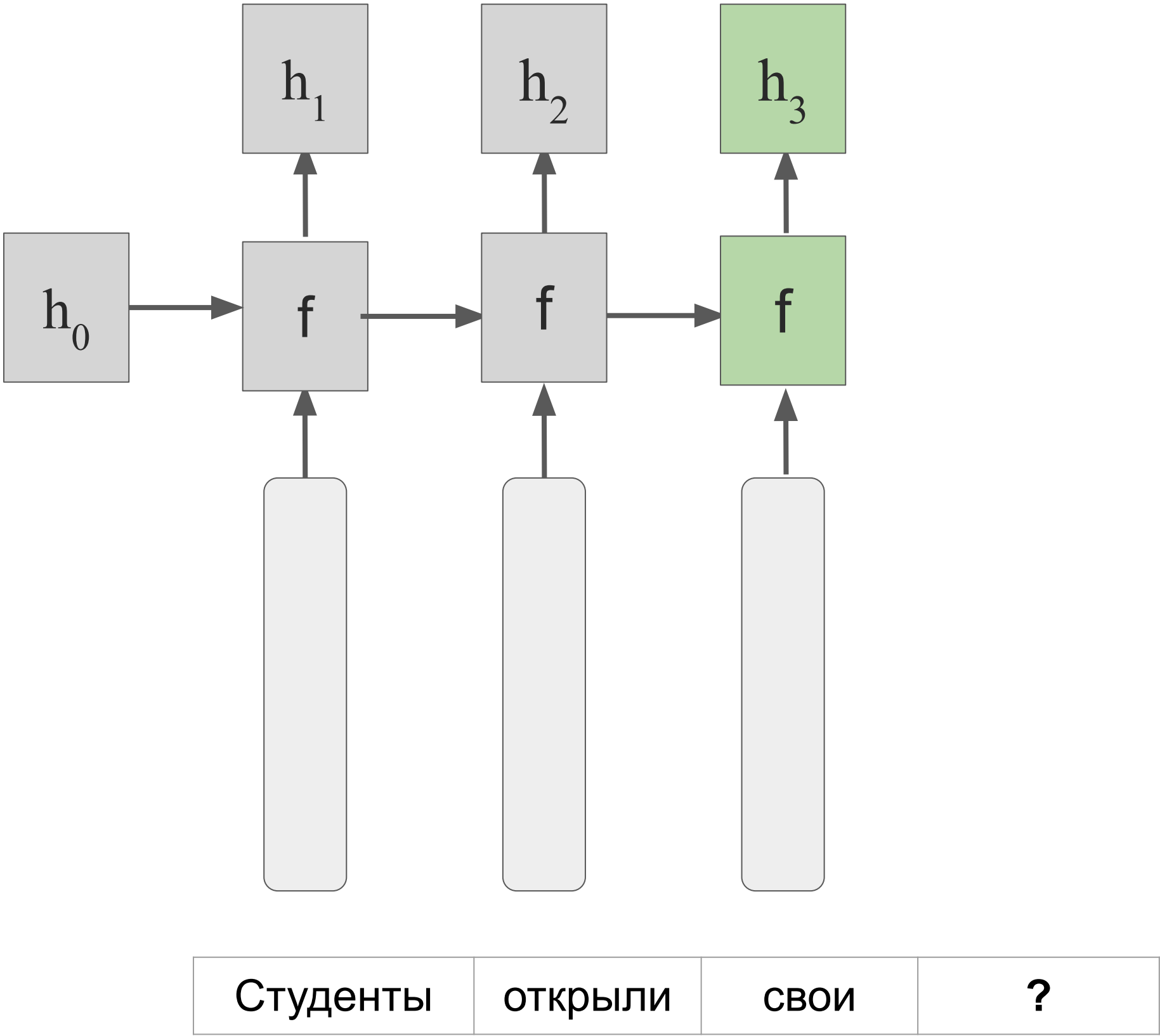


# Пример

Скрытые состояния

Эмбеддинги

Слова

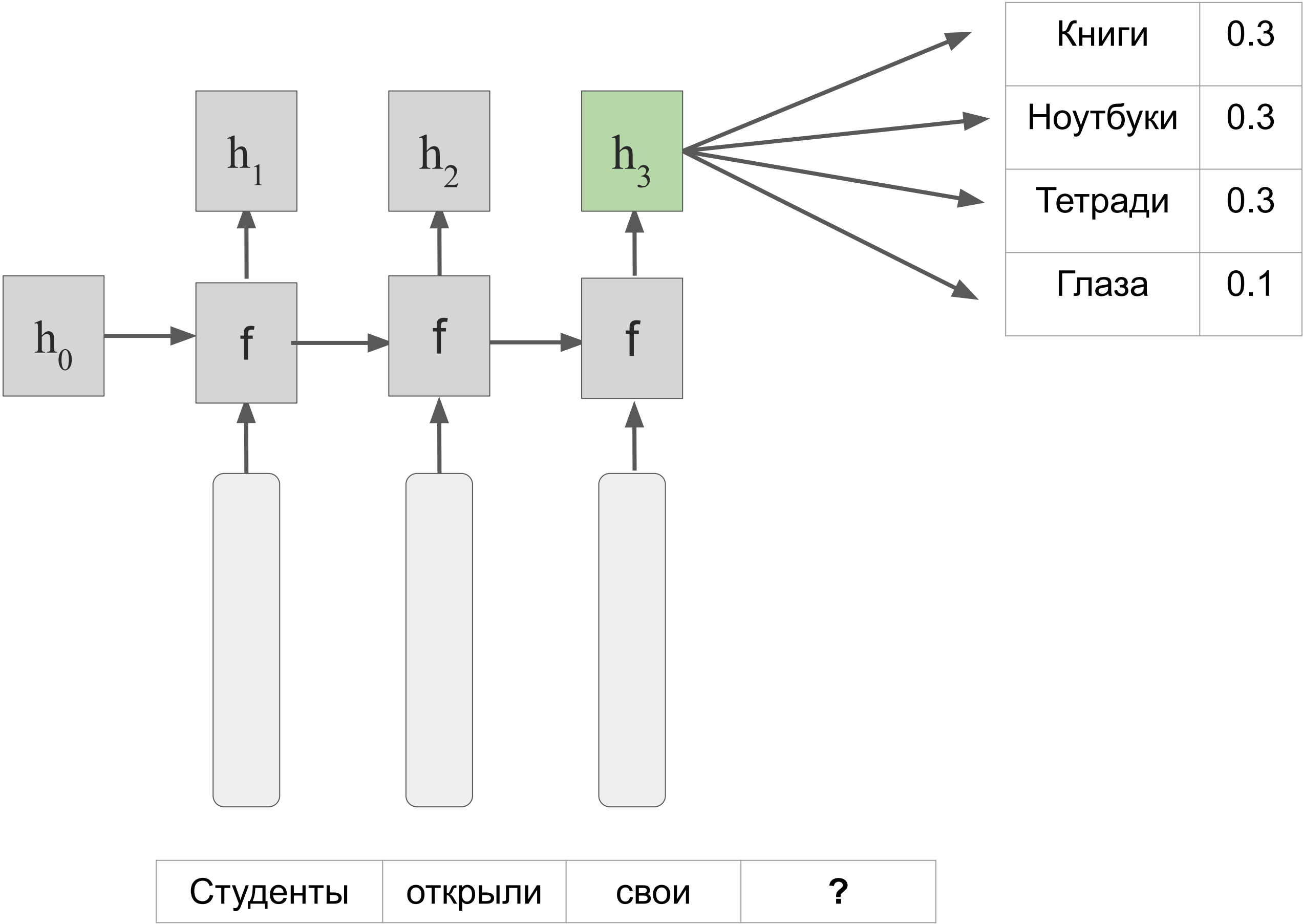


# Пример

Скрытые состояния

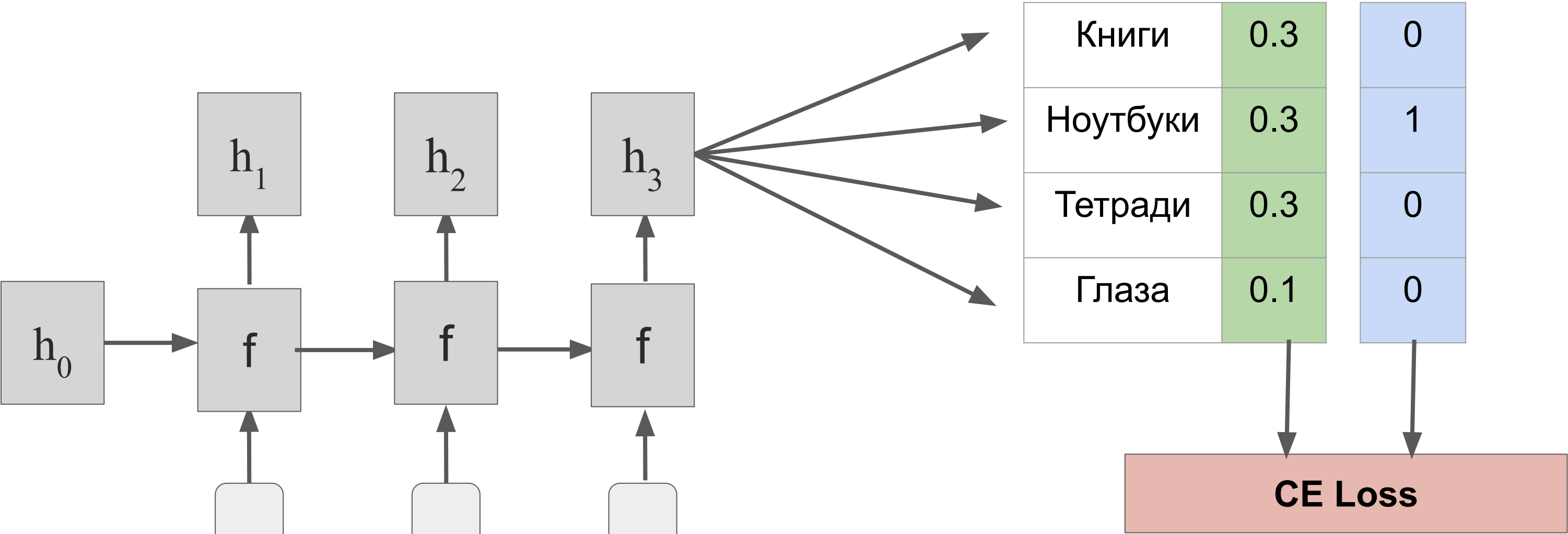
Эмбеддинги

Слова



# Пример

Скрытые состояния



Эмбеддинги

Слова

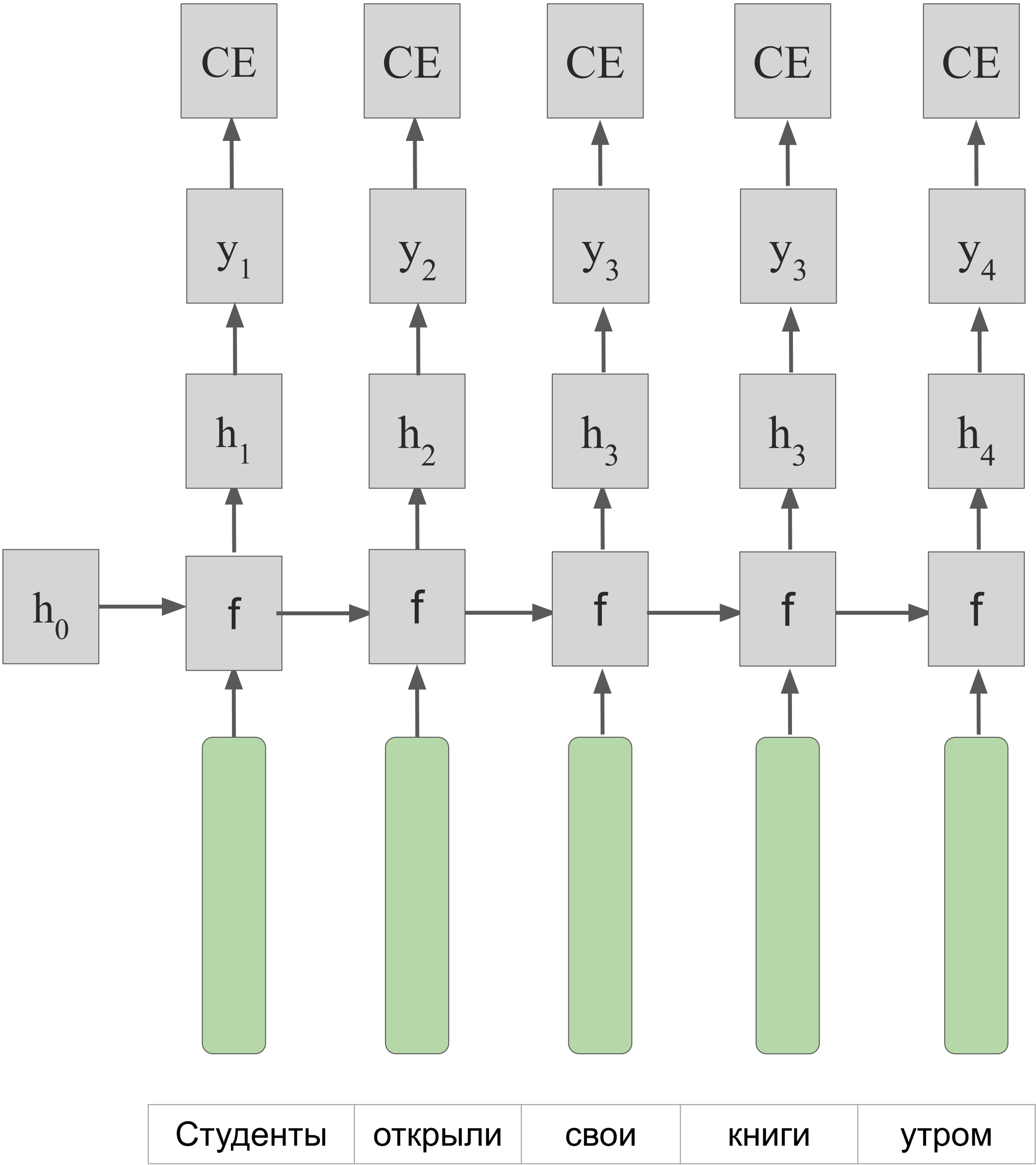
Студенты	открыли	свои	?
----------	---------	------	---

# Пример

Скрытые состояния

Эмбеддинги

Слова

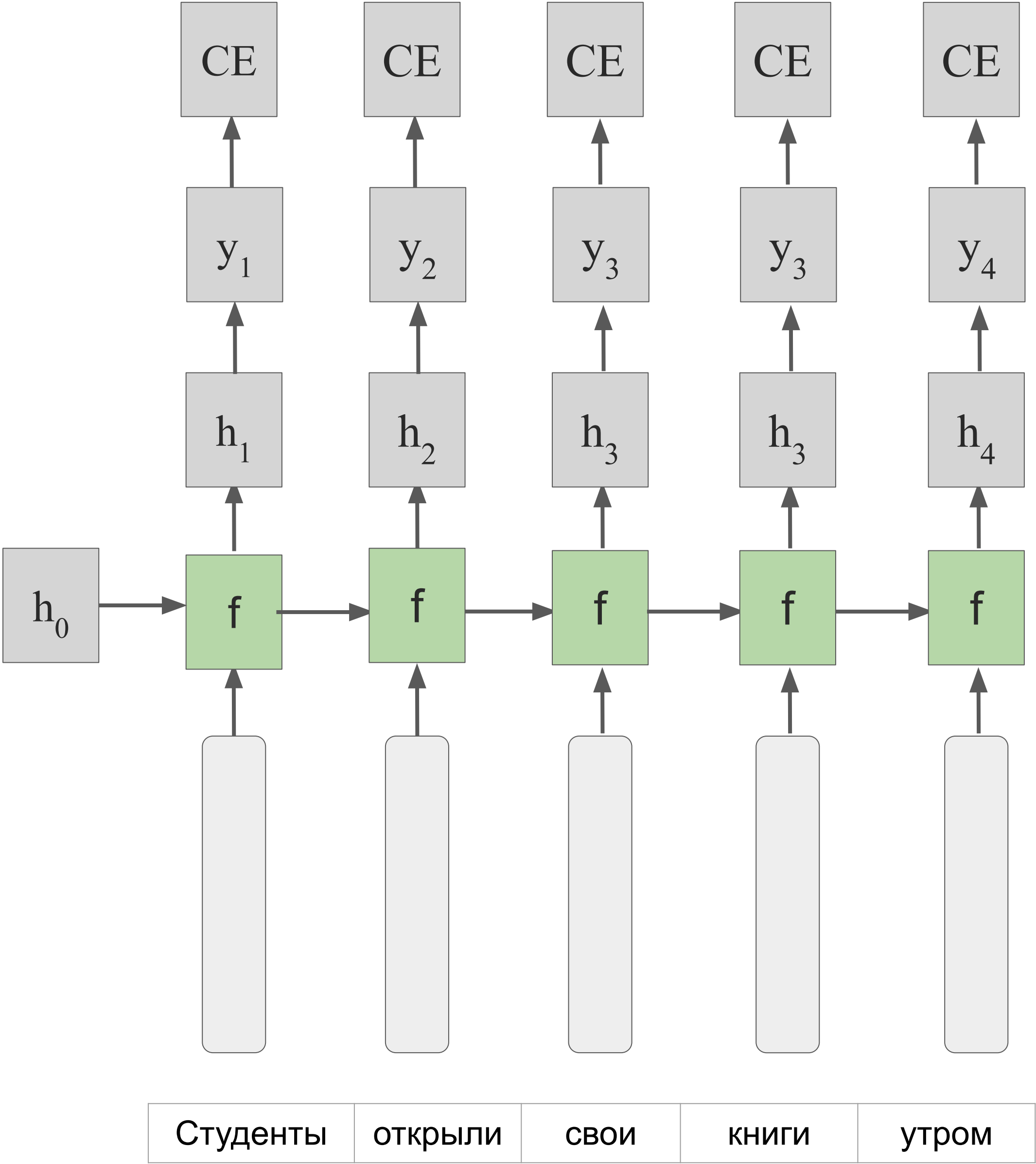


# Пример

Скрытые состояния

Эмбеддинги

Слова

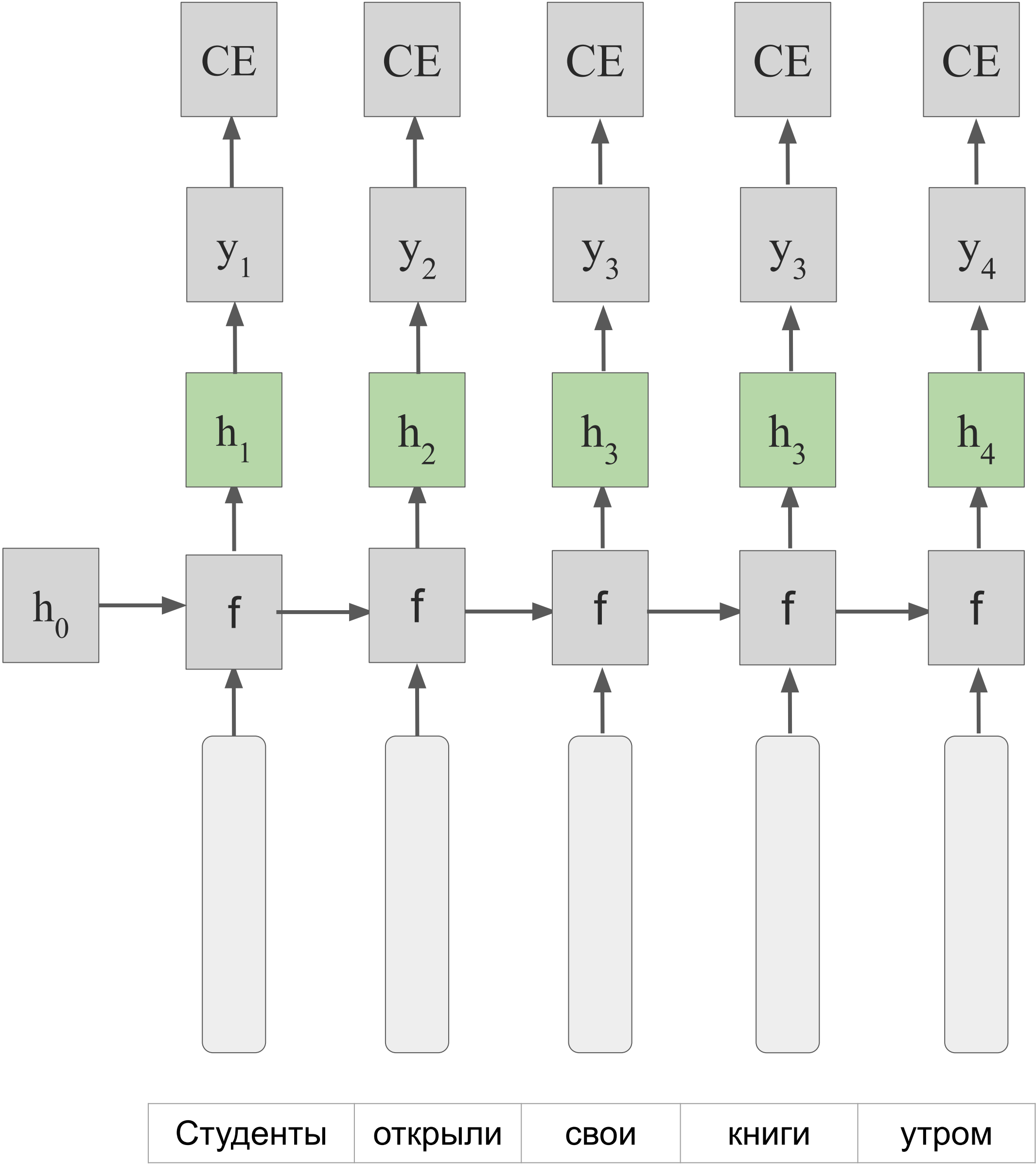


# Пример

Скрытые состояния

Эмбеддинги

Слова



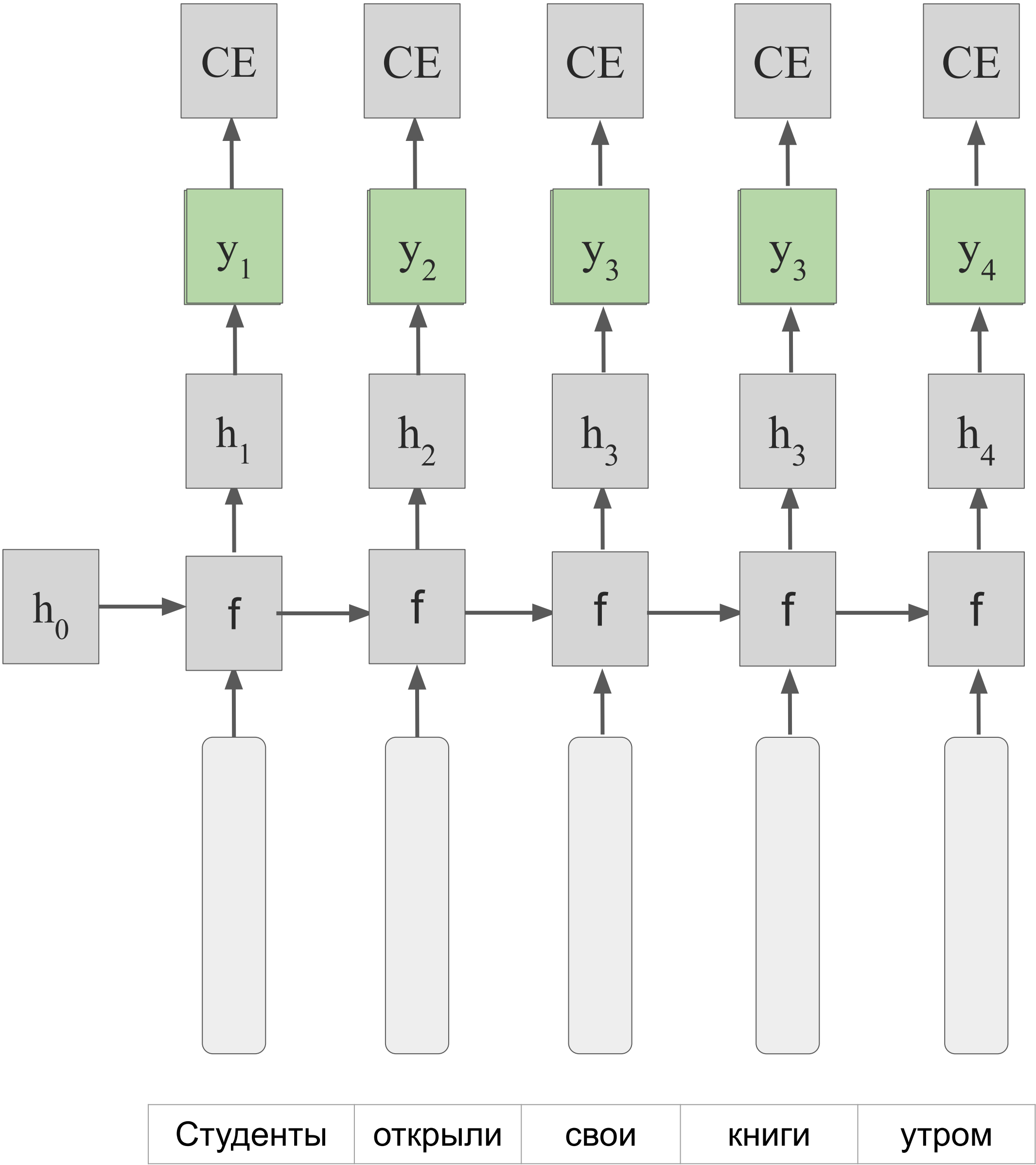


# Пример

Скрытые состояния

Эмбеддинги

Слова

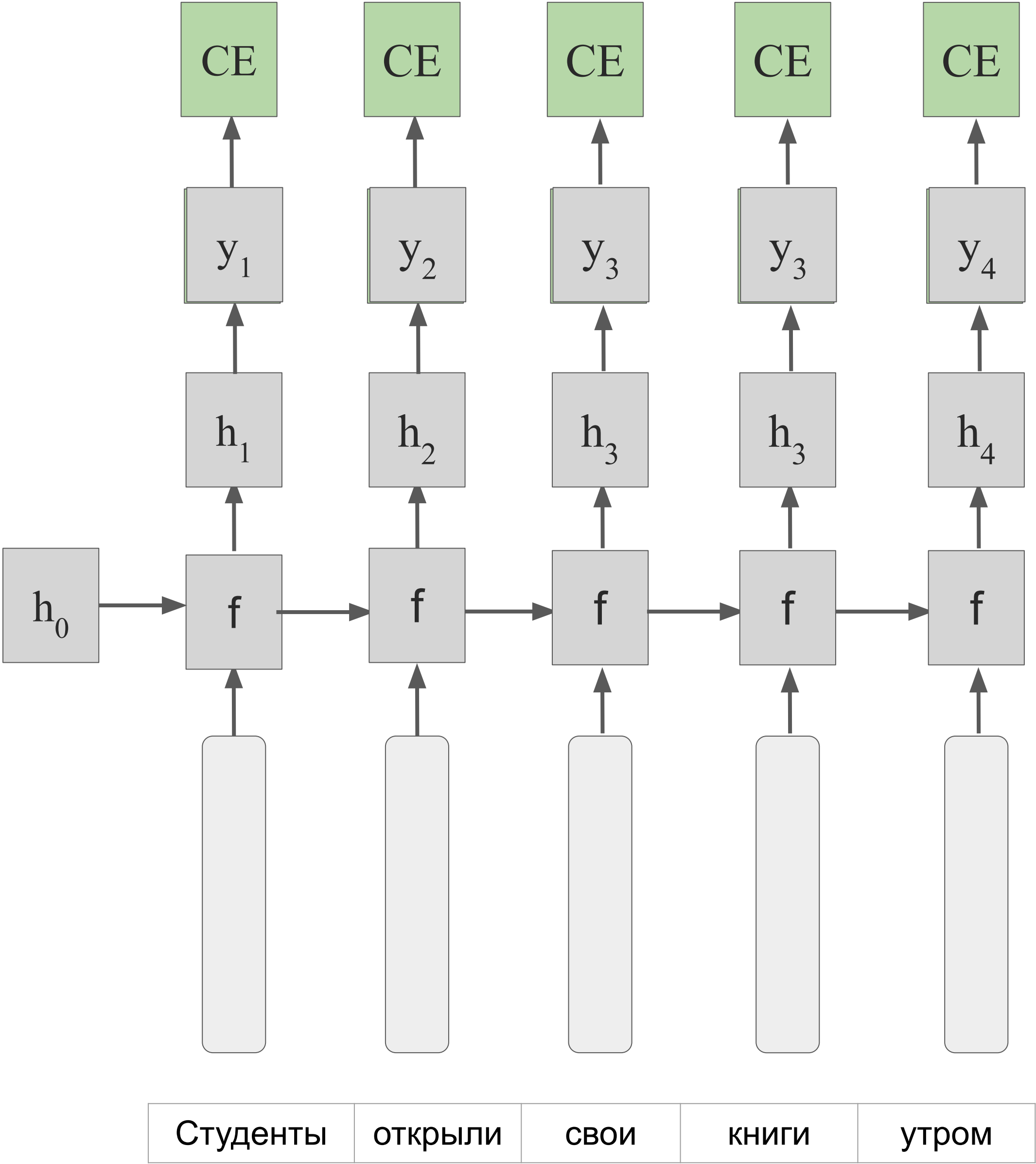


# Пример

Скрытые состояния

Эмбеддинги

Слова



# Расчет лосса

$$J(\theta) = \frac{1}{T} ( \begin{array}{c} \boxed{\text{CE}} \\ \uparrow \\ \boxed{y_1} \end{array} + \begin{array}{c} \boxed{\text{CE}} \\ \uparrow \\ \boxed{y_2} \end{array} + \begin{array}{c} \boxed{\text{CE}} \\ \uparrow \\ \boxed{y_3} \end{array} + \begin{array}{c} \boxed{\text{CE}} \\ \uparrow \\ \boxed{y_3} \end{array} + \begin{array}{c} \boxed{\text{CE}} \\ \uparrow \\ \boxed{y_4} \end{array} ) = \frac{1}{T} \sum_{t=1}^T J^{(t)}(\theta) = \frac{1}{T} \sum_{t=1}^T -\log \hat{\mathbf{y}}_{\mathbf{x}_{t+1}}^{(t)}$$

# Плюсы и минусы RNN для решения задачи языкового моделирования

## Плюсы:

1. Обработывают гораздо больший контекст по сравнению с N-граммными моделями
2. Модель не требует больше памяти для улавливания более длинного контекста

## Минусы:

1. Медленно вычисляют предсказание из-за рекуррентной природы
2. Взрывающиеся и затухающие градиенты требуют архитектурных улучшений в виде LSTM и GRU

# Итоги занятия

1. Вспомнили идею работы рекуррентных сетей
2. Узнали, как обучить языковую модель на основе RNN
3. Поняли, что RNN способна учитывать более длинный контекст с меньшими затратами по памяти в сравнении с N-граммными моделями
4. Поняли, что RNN работают медленнее N-граммных моделей и требуют применения архитектурных трюков для обучения