Aprendizaje profundo

Mecanismos de atención

Gibran Fuentes-Pineda

Octubre 2023

Modelos secuencia a secuencia (seq2seq)

- Necesitan codificar todo el contexto de la entrada en un sólo vector

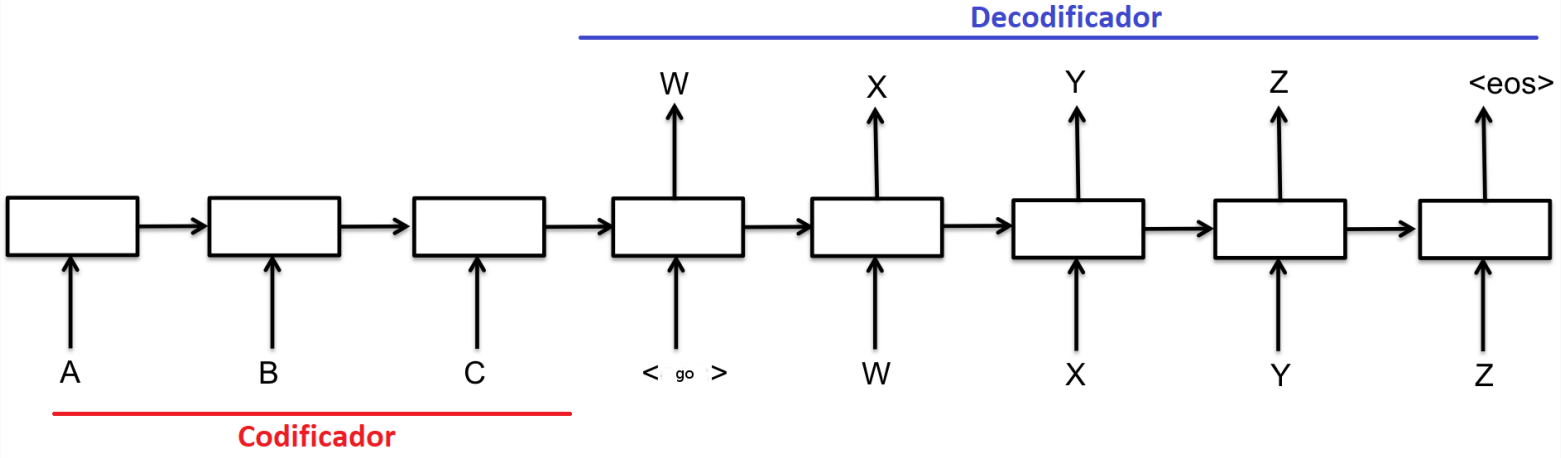

Imagen derivada de `https://www.tensorflow.org/tutorials/seq2seq`

Ejemplo de seq2seq para traducción

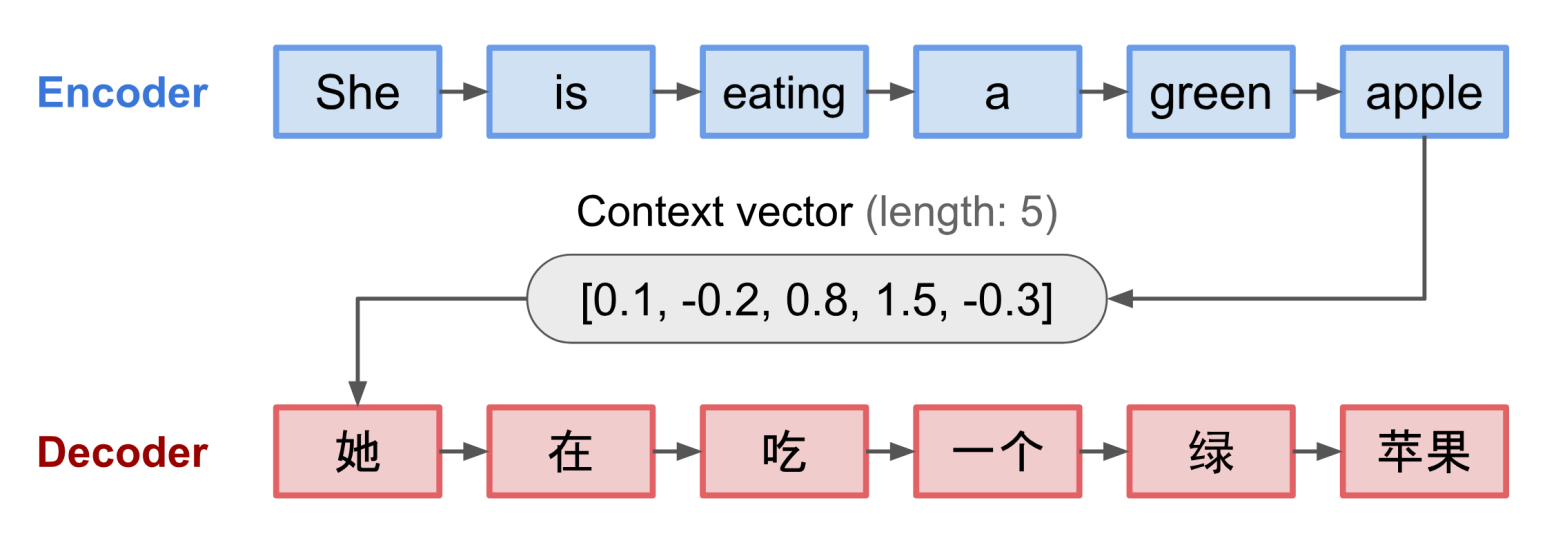

Imagen tomada de `https://lilianweng.github.io/lil-log/2018/06/24/attention-attention.html`

Mecanismos de atención

- Información relevante de la entrada se codifica en un solo vector de contexto (último estado de una red recurrente)
  - Difícil en secuencias largas

- Mecanismos de atención: se calcula un vector de contexto distinto por cada paso del decodificador a partir de:
  1. Estados del decodificador
  2. Estados del codificador
  3. Función de alineación

Esquema general de los mecanismos de atención

- Dado un paso t del decodificador, calcular un vector de contexto $\mathbf{c}^{[t]}$ a partir de todos los estados del codificador

$$\mathbf{c}^{[t]} = \sum_{i=1}^{T} \alpha_{t,i} \cdot \hat{\mathbf{h}}^{[i]}$$

donde $\hat{\mathbf{h}}^{[i]}$ es el estado del codificador en el paso i, T es el número total de estados del codificador y $\alpha_{t,i}$ es un valor de atención para $\mathbf{h}^{[i]}$ calculada de la siguiente manera:

$$\begin{aligned} \alpha_{t,i} &= \text{alineación}(\mathbf{h}^{[t]}, \hat{\mathbf{h}}^{[i]}) \\ &= \mathit{softmax}(\mathit{puntaje}(\mathbf{h}^{[t]}, \hat{\mathbf{h}^{[i]}})) \end{aligned}$$

- *puntaje* es una función que mide la importancia de 2 estados y $\mathbf{h}^{[t]}$ es el estado en el paso t del decodificador.

Funciones *puntaje* (1)

- Basadas en contenido

$$\text{puntaje}(\mathbf{h}^{[t]}, \hat{\mathbf{h}}^{[i]}) = \mathbf{h}^{[t]\top} \hat{\mathbf{h}}^{[i]} \text{ (producto punto)}$$

$$\text{puntaje}(\mathbf{h}^{[t]}, \hat{\mathbf{h}}^{[i]}) = \frac{\mathbf{h}^{[t]\top} \hat{\mathbf{h}}^{[i]}}{\sqrt{T}} \text{ (producto punto escalado)}$$

$$\text{puntaje}(\mathbf{h}^{[t]}, \hat{\mathbf{h}}^{[i]}) = \mathbf{h}^{[t]\top} \mathbf{W}_a \hat{\mathbf{h}}^{[i]} \text{ (general)}$$

$$\text{puntaje}(\mathbf{h}^{[t]}, \hat{\mathbf{h}}^{[i]}) = \mathbf{W}_a \left[\mathbf{h}^{[t]}; \hat{\mathbf{h}}^{[i]}\right] \text{ (concatenación)}$$

- Basada en ubicación

$$\alpha_{t,i} = \mathbf{W}_a \mathbf{h}^{[t]}$$

Consultas, llaves y valores en funciones *puntaje*

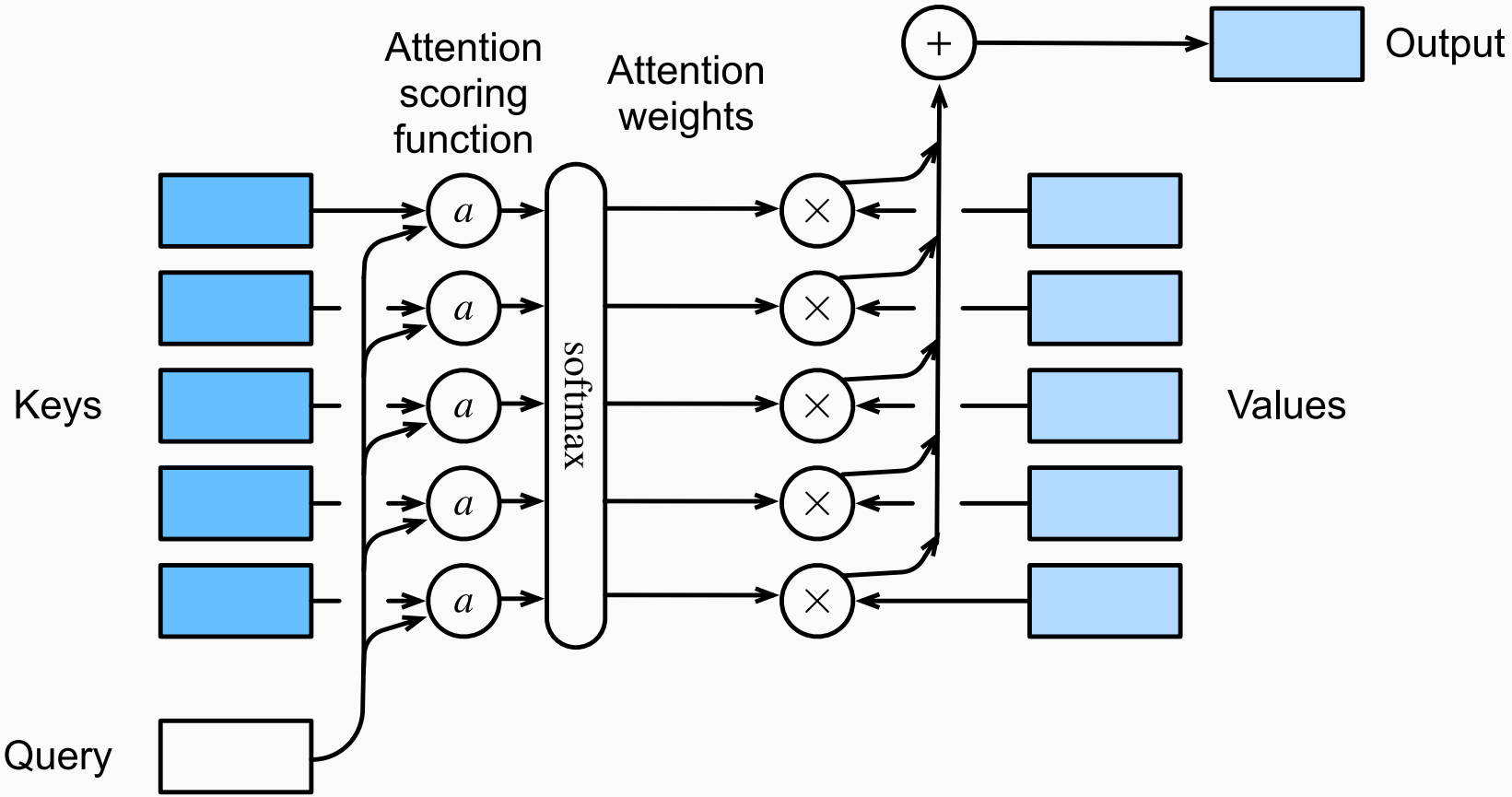

Figura tomada de Zhang et al. Dive into Deep Learning, 2022

Atención de Bahdanau (aditiva)

$$\textit{puntaje}(\mathbf{q}, \mathbf{k}) = \mathbf{w}_v^\top \tanh(\mathbf{W}_q \mathbf{q} + \mathbf{W}_k \mathbf{k}) \in \mathbb{R},$$

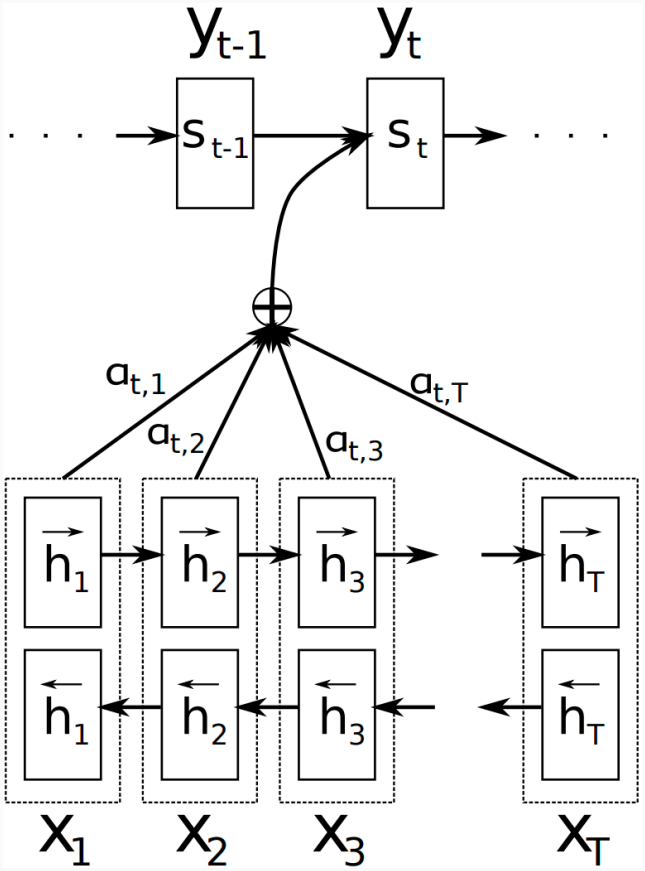

Imagen tomada de Bahdanau et al. Neural machine translation by jointly learning to align and translate, ICLR, 2015

Atención de Luong (global)

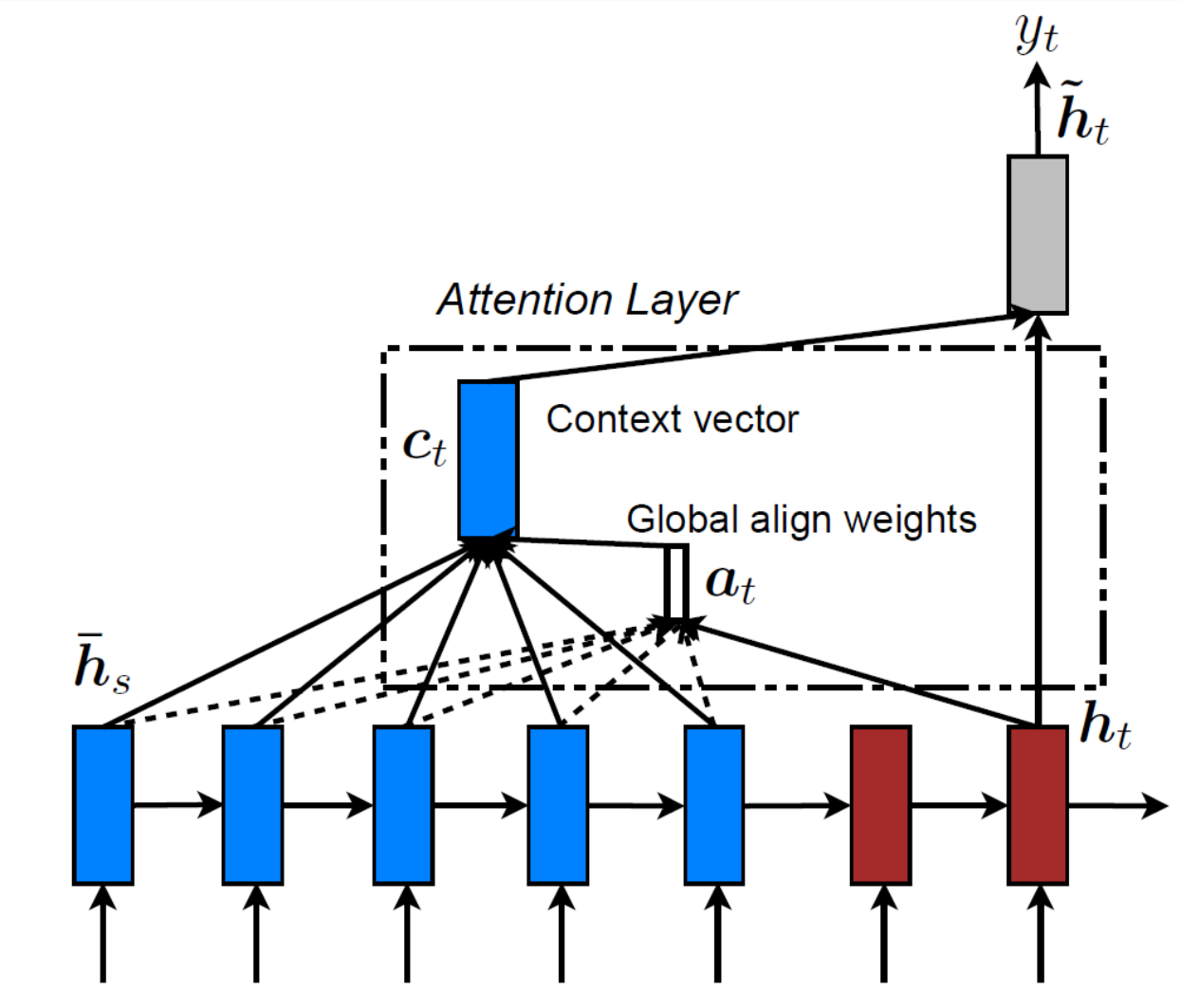

Imagen tomada de Luong et al. Effective approaches to attention-based neural machine translation, EMNLP, 2015

Atención de Luong (local)

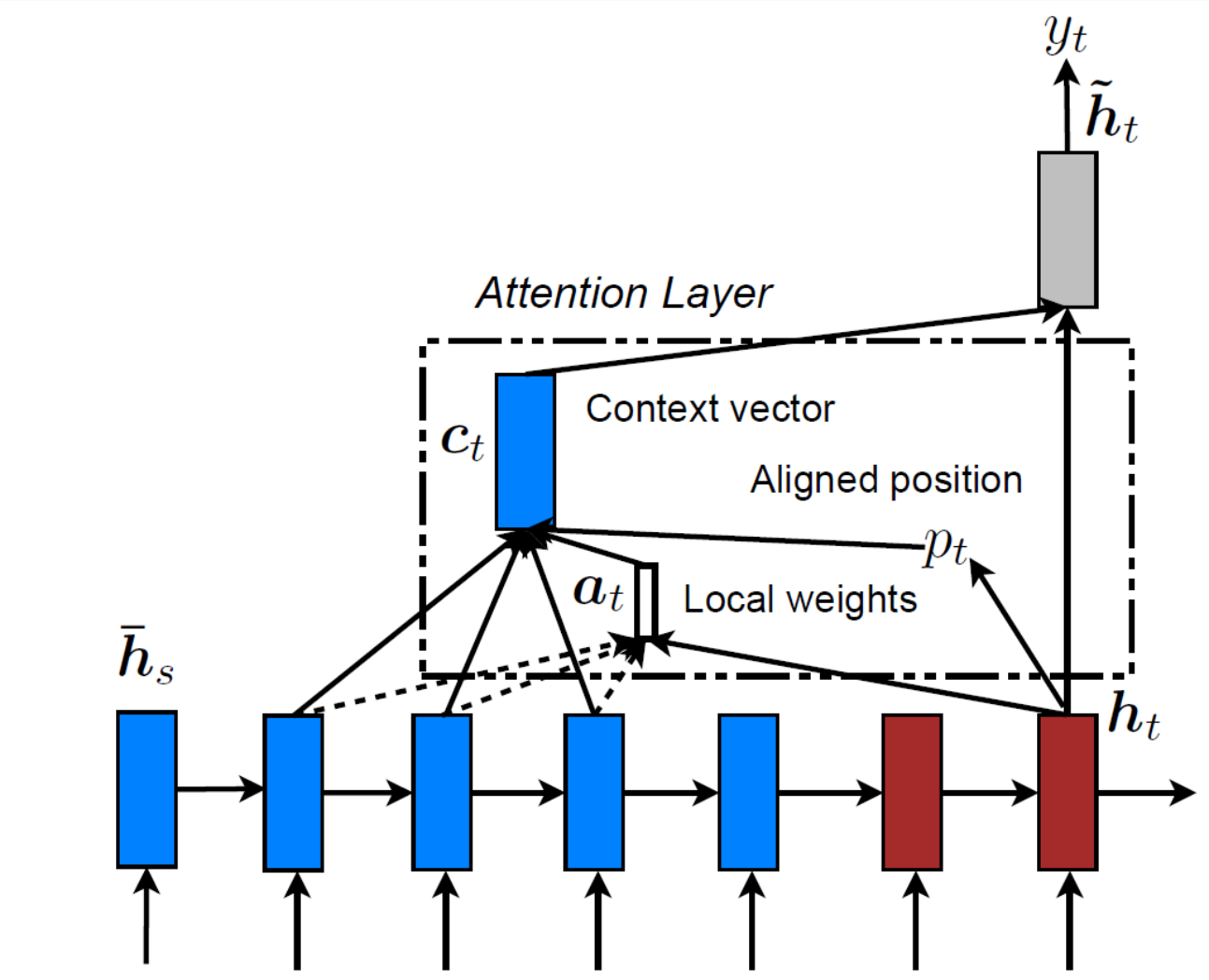

Imagen tomada de Luong et al. Effective approaches to attention-based neural machine translation, EMNLP, 2015

Atención con Luong (alimentación de entradas)

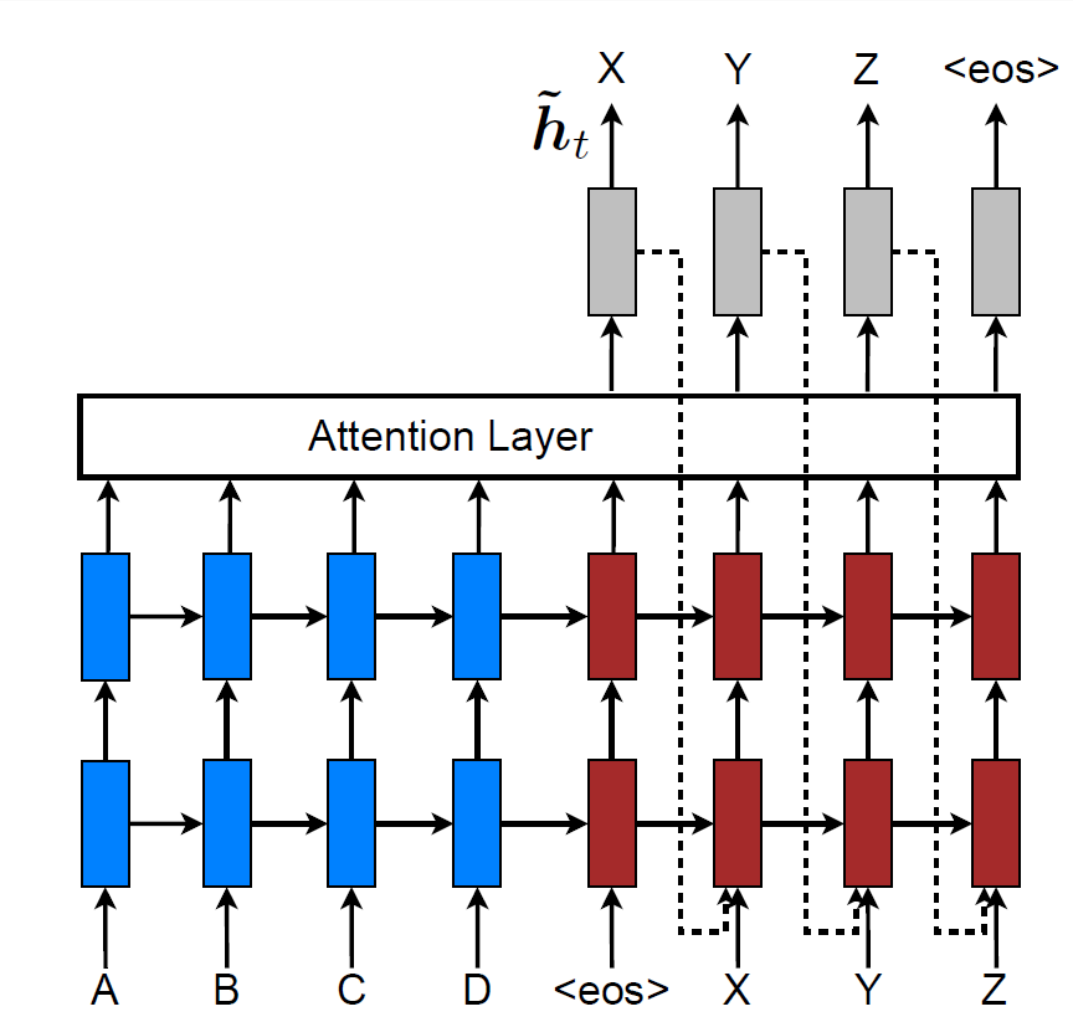

Imagen tomada de Luong et al. Effective approaches to attention-based neural machine translation, EMNLP, 2015

Efecto del tamaño de secuencia en la atención

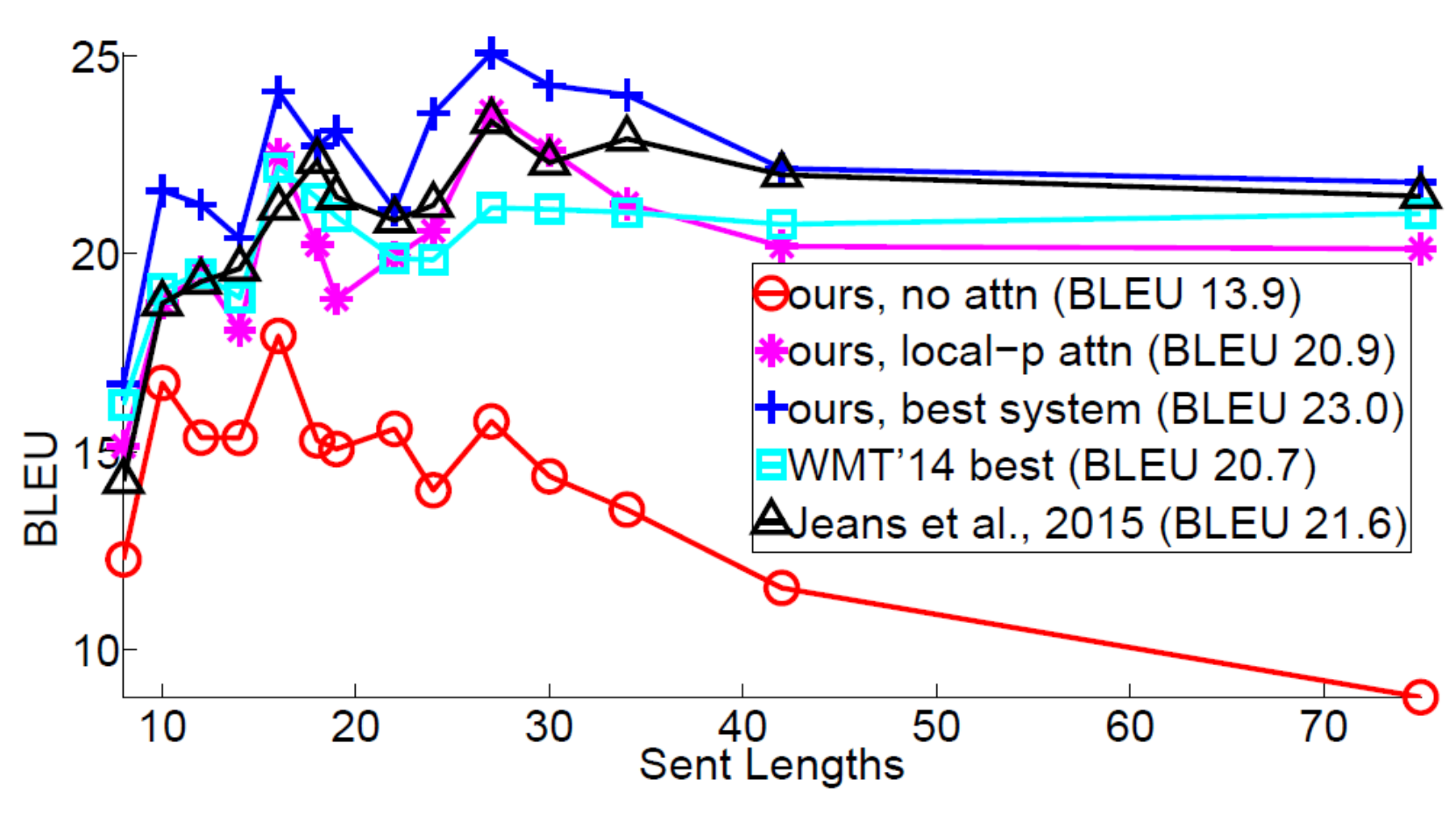

Imagen tomada de Luong et al. Effective approaches to attention-based neural machine translation, EMNLP, 2015

Visualización de puntuaciones de la alineación

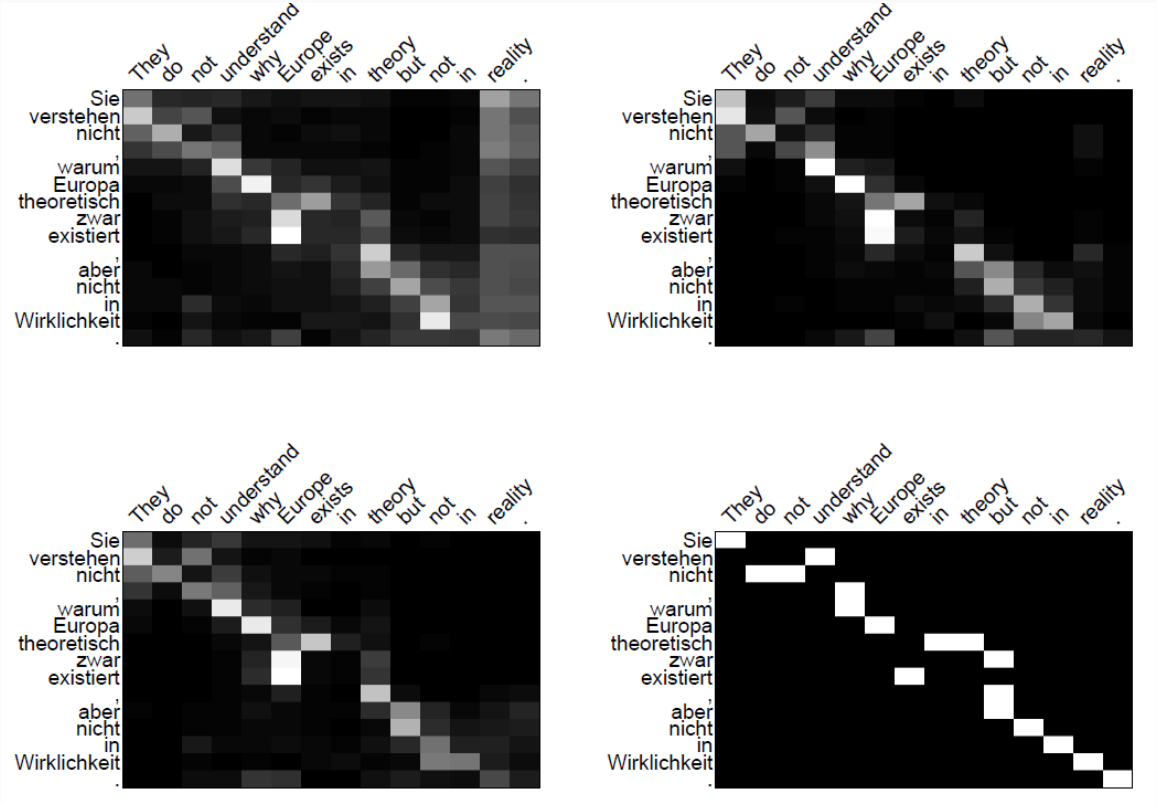

Imagen tomada de Bahdanau et al. Neural machine translation by jointly learning to align and translate, ICLR, 2015

Atención en descripción de imágenes (1)

- Permite enforcarse a sólo ciertas partes de la imagen entrada al producir cada palabra en la salida
- Modelo aprende a qué partes ponerle atención en cada paso

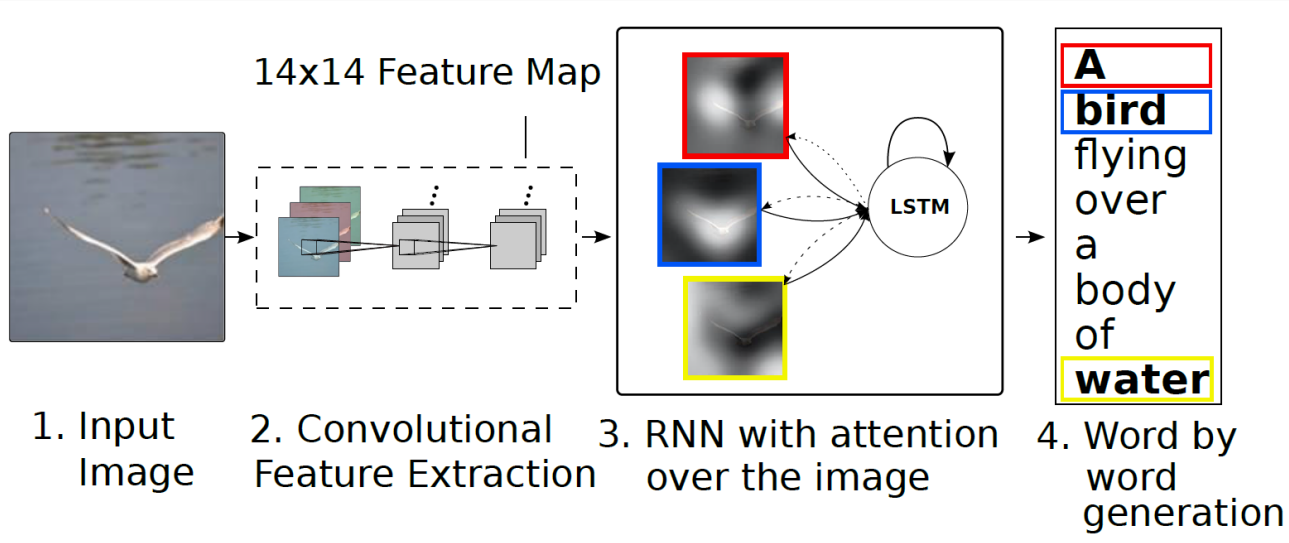

Imagen tomada de Xu et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML, 2015

Atención en descripción de imágenes (2)

- **Dura:** se toma en cuenta una sola region de la imagen
- **Suave:** se toma en cuenta cada región de la imagen en distinta proporción, de forma similar a la atención de Bahdanau

Figure 2. Attention over time. As the model generates each word, its attention changes to reflect the relevant parts of the image. "soft" (top row) vs "hard" (bottom row) attention. (Note that both models generated the same captions in this example.)

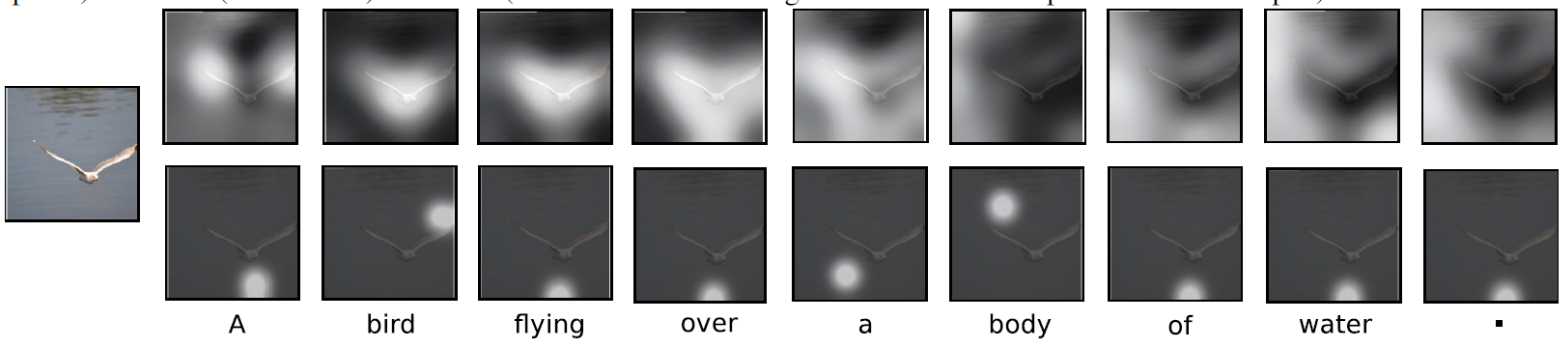

Imagen tomada de Xu et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML, 2015

Atención en descripción de imágenes (3)

- Podemos visualizar a qué partes de la imagen le pone atención el modelo al producir cada palabra de la descripción.

Imagen tomada de Xu et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML, 2015

Atención en descripción de imágenes (4)

- La atención nos permite entender mejor los errores que comente el modelo.

Figure 5. Examples of mistakes where we can use attention to gain intuition into what the model saw.

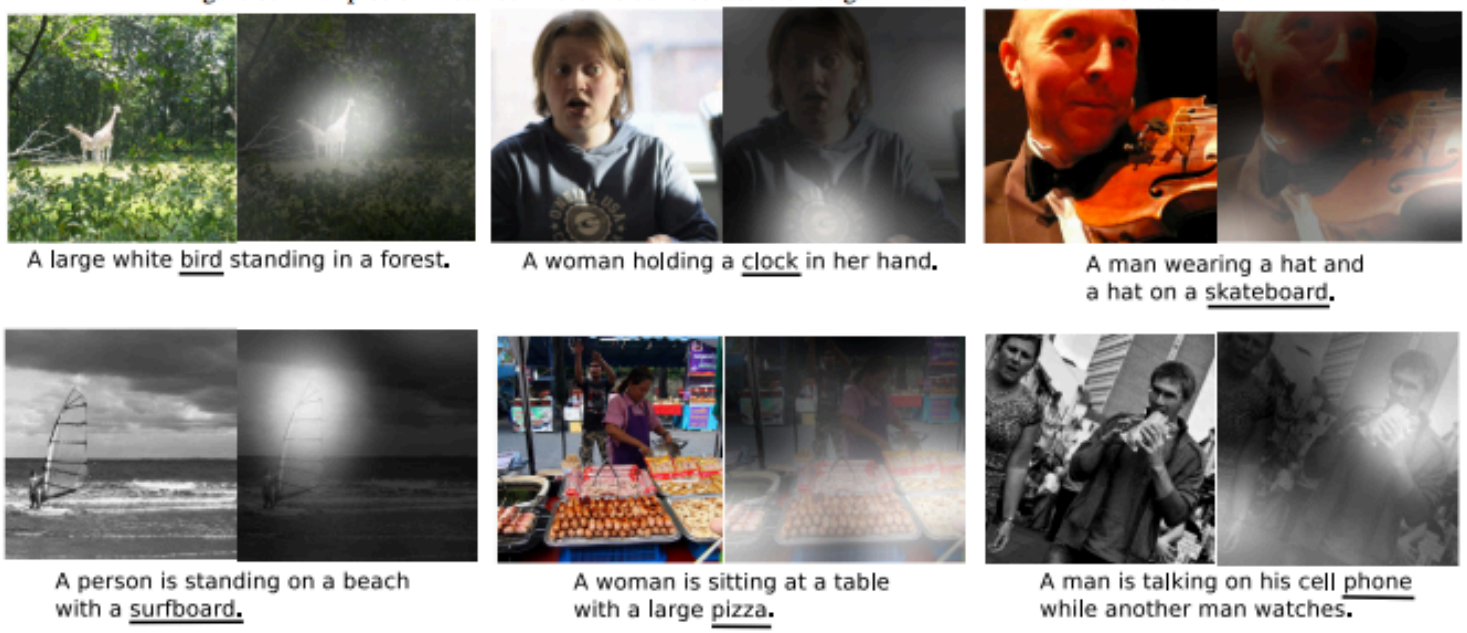

Imagen tomada de Xu et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, ICML, 2015