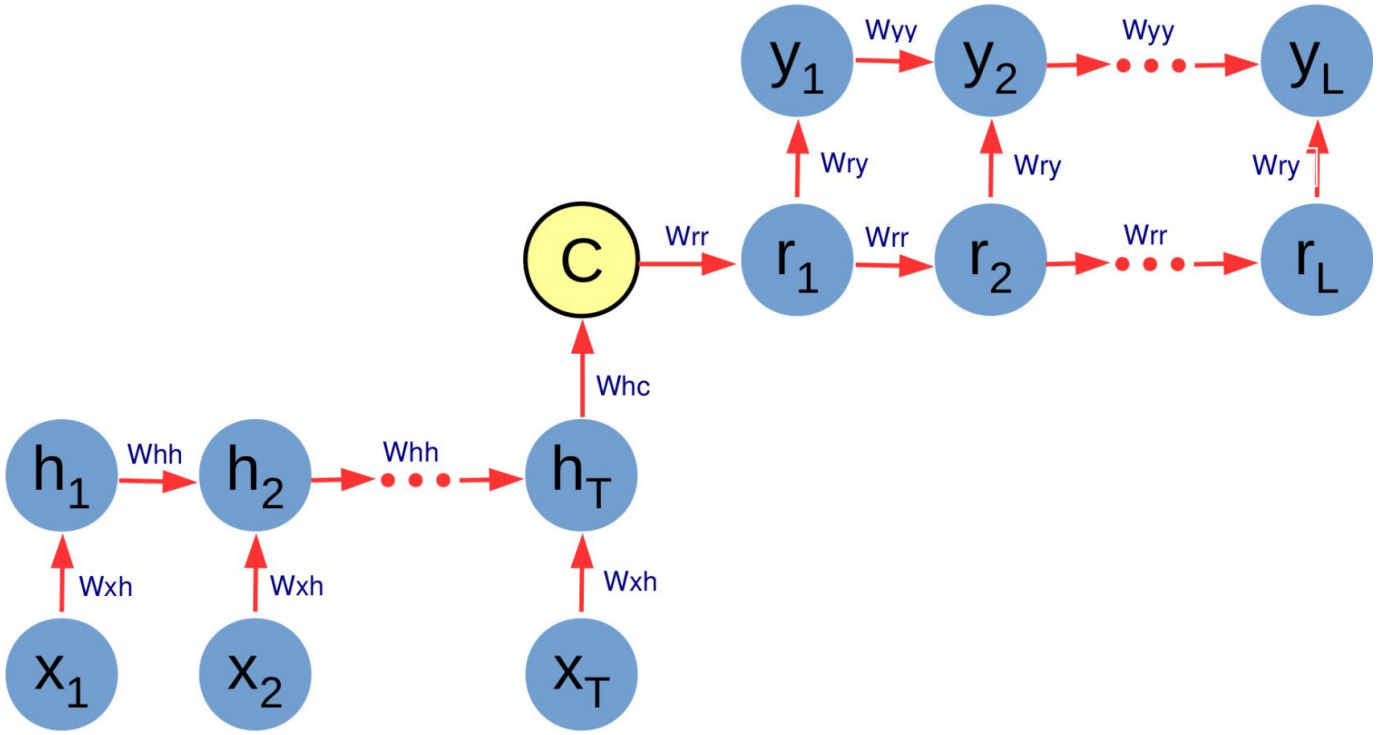
Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ciencia de la Computación

Sistemas Urbanos Inteligentes

Mecanismos de atención

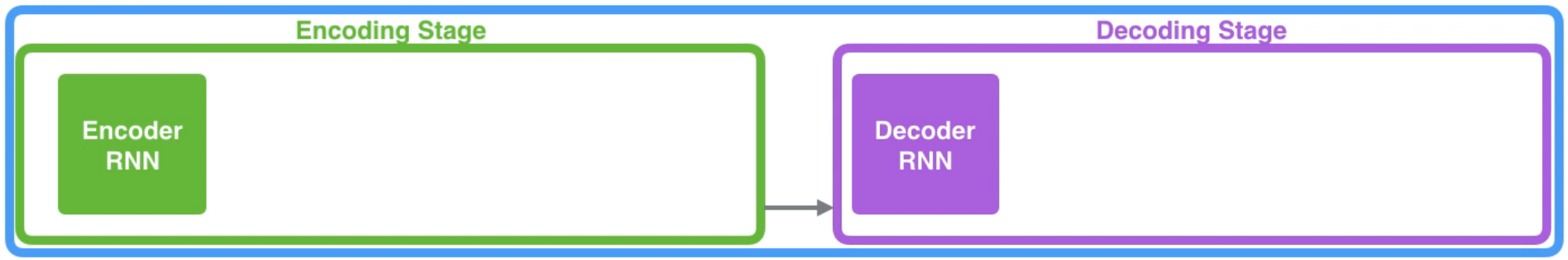
Hans Löbel
Dpto. Ingeniería de Transporte y Logística
Dpto. Ciencia de la Computación

El modelo más simple considera un *encoder* que genera un único vector contextual C



Neural Machine Translation

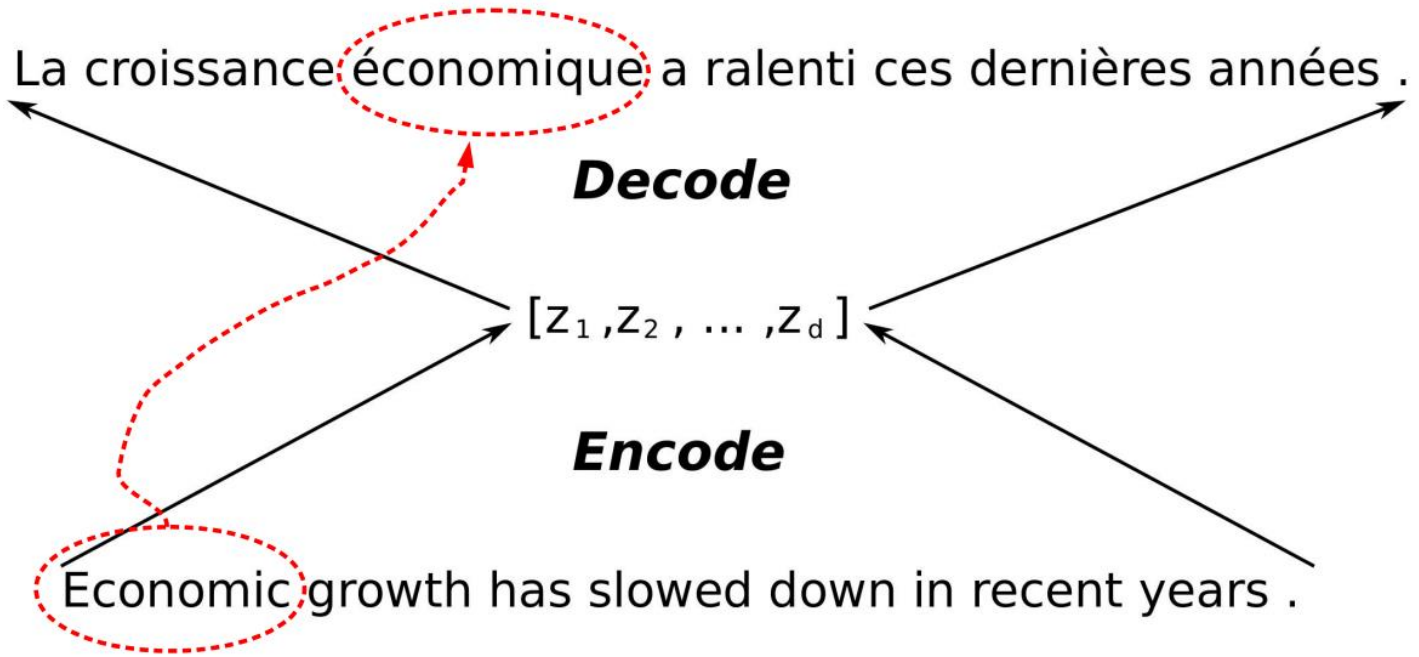
SEQUENCE TO SEQUENCE MODEL



Je suis étudiant

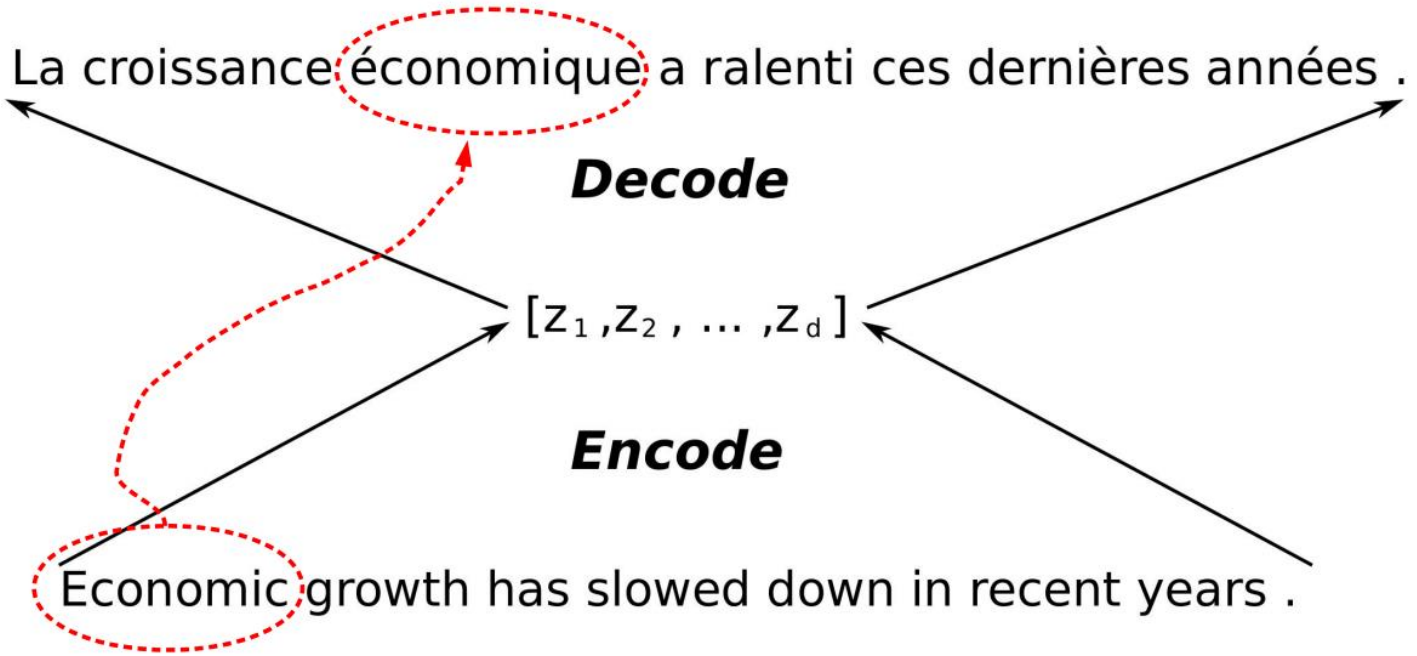
¿Es un solo contexto suficiente?

- Si bien estos modelos son poderosos, dependen completamente de que el contexto esté bien representado en C, ya que es la única conexión entre *encoder* y *decoder*.
- Más aún, asumen que todos los pasos pueden basarse en el mismo contexto inicial. El problema de esto es que un contexto fijo le da una importancia fija a cada parte de la entrada.

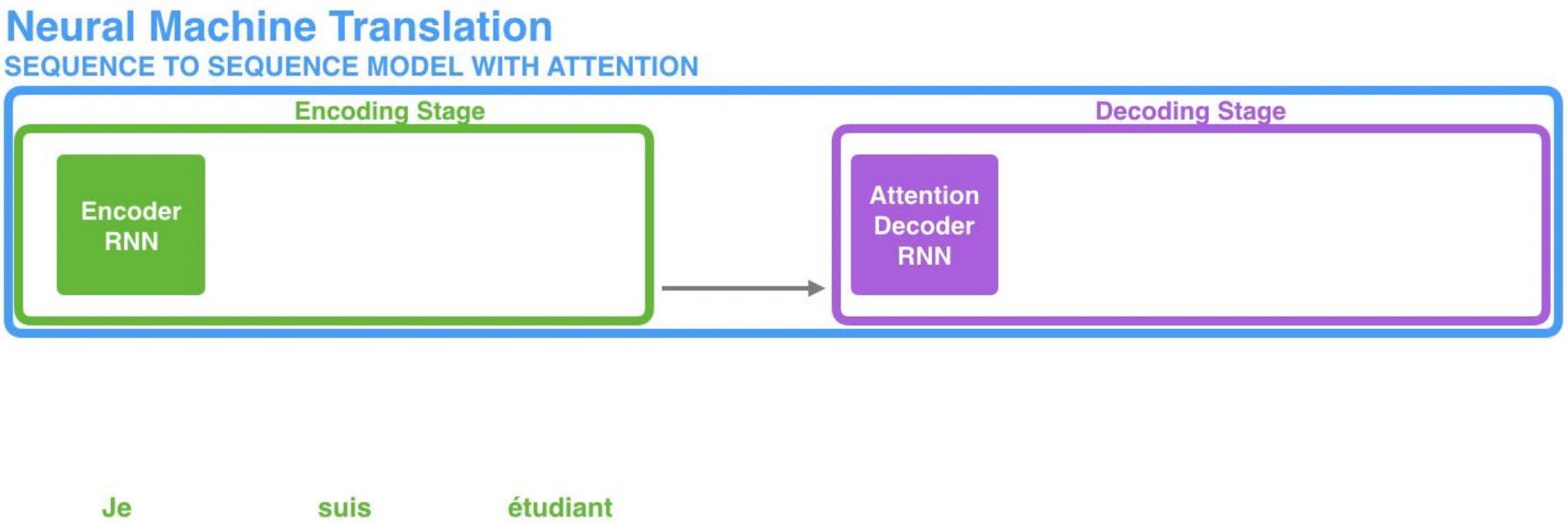


Una posible solución es manejar contextos adaptativos usando atención

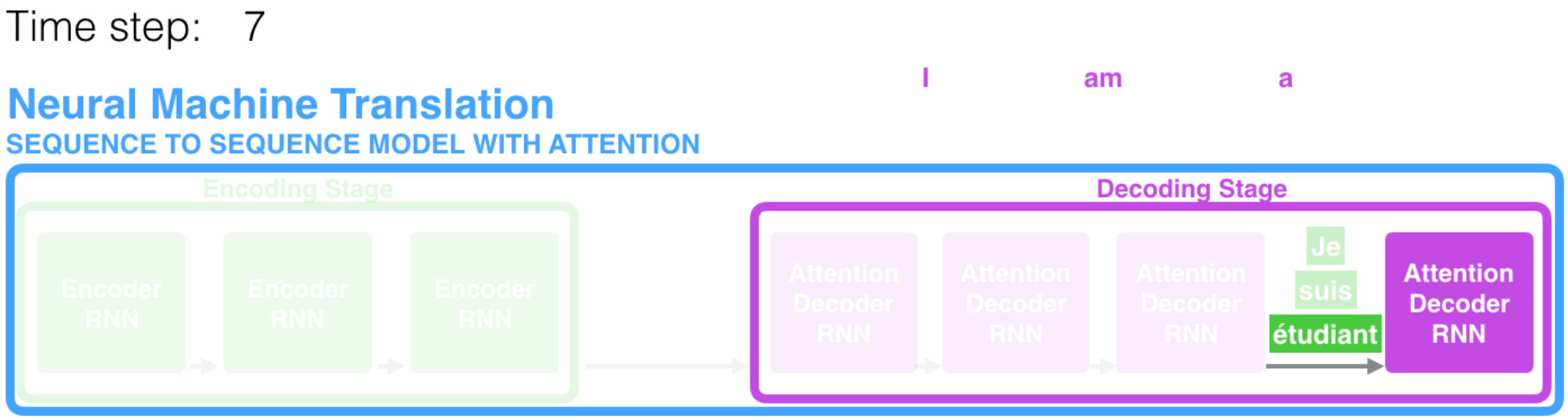
- Un mecanismo de atención es esencialmente una nueva capa que pondera en cada paso la relevancia de cada estado del *encoder*, y por consiguiente, de la secuencia de entrada.
- Esto permite liberar la carga de conocimiento que debe almacenar el vector C y especializarlo de acuerdo al paso particular del proceso.



Veamos un ejemplo simple

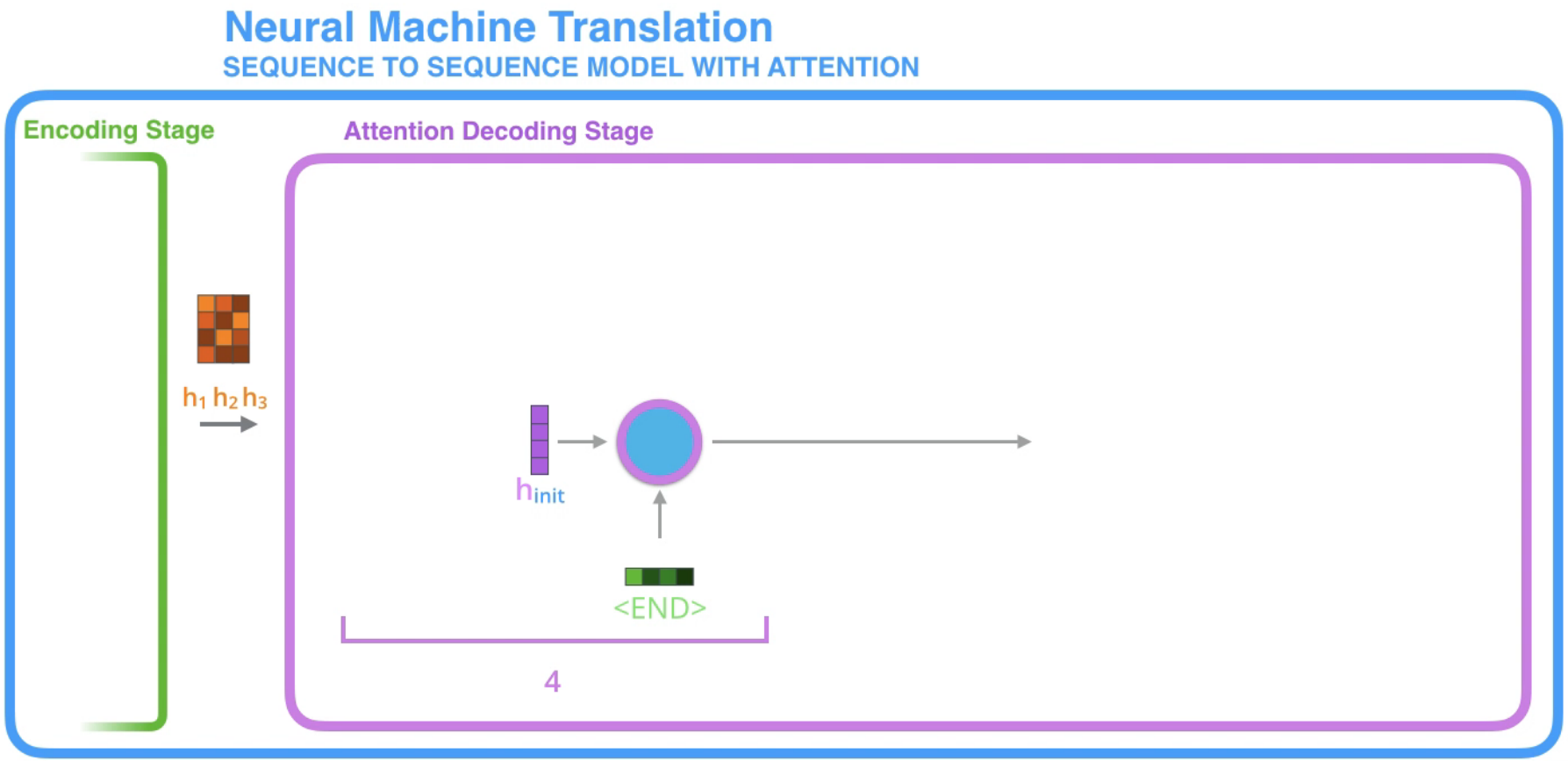


Veamos un ejemplo simple

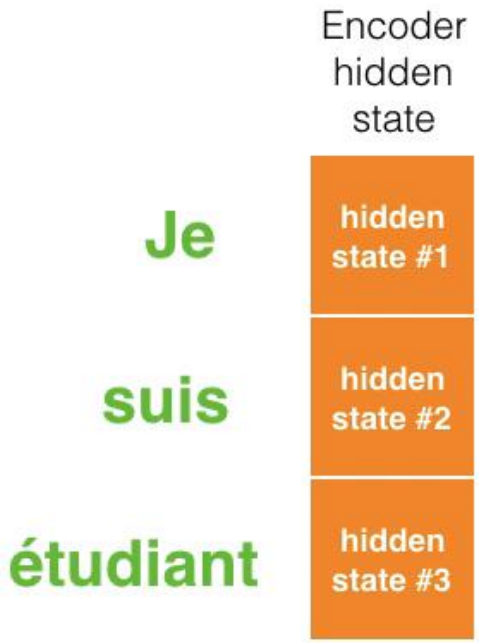


Veamos un ejemplo simple

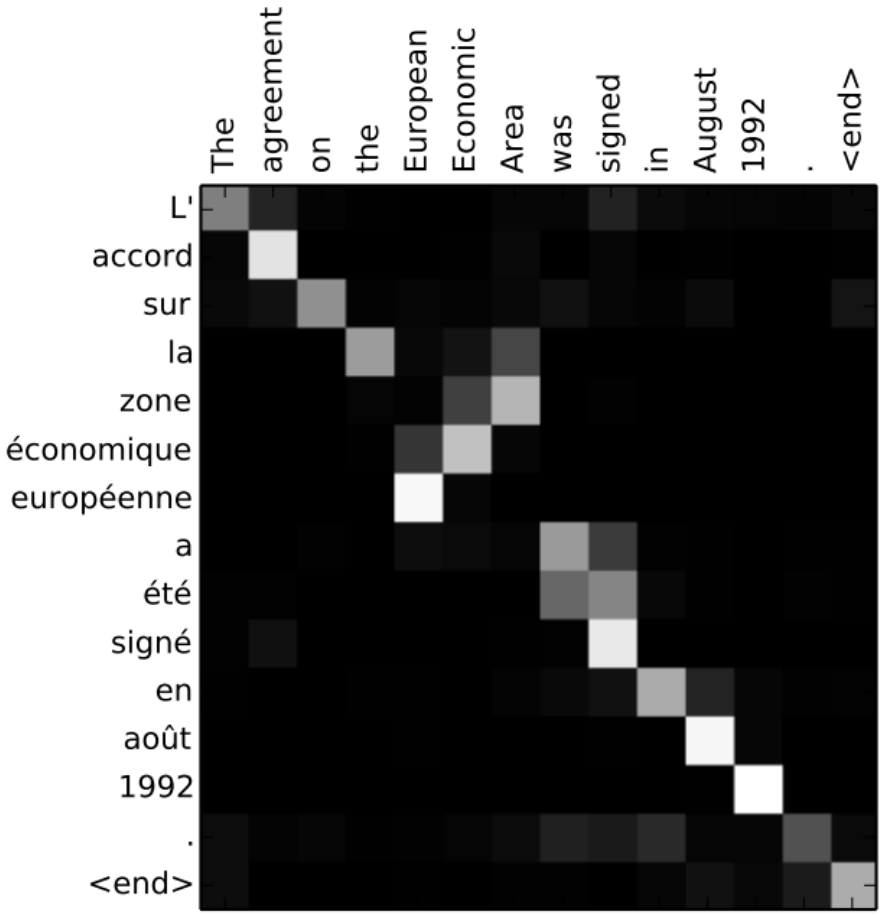
Veamos un ejemplo simple



Veamos un ejemplo simple

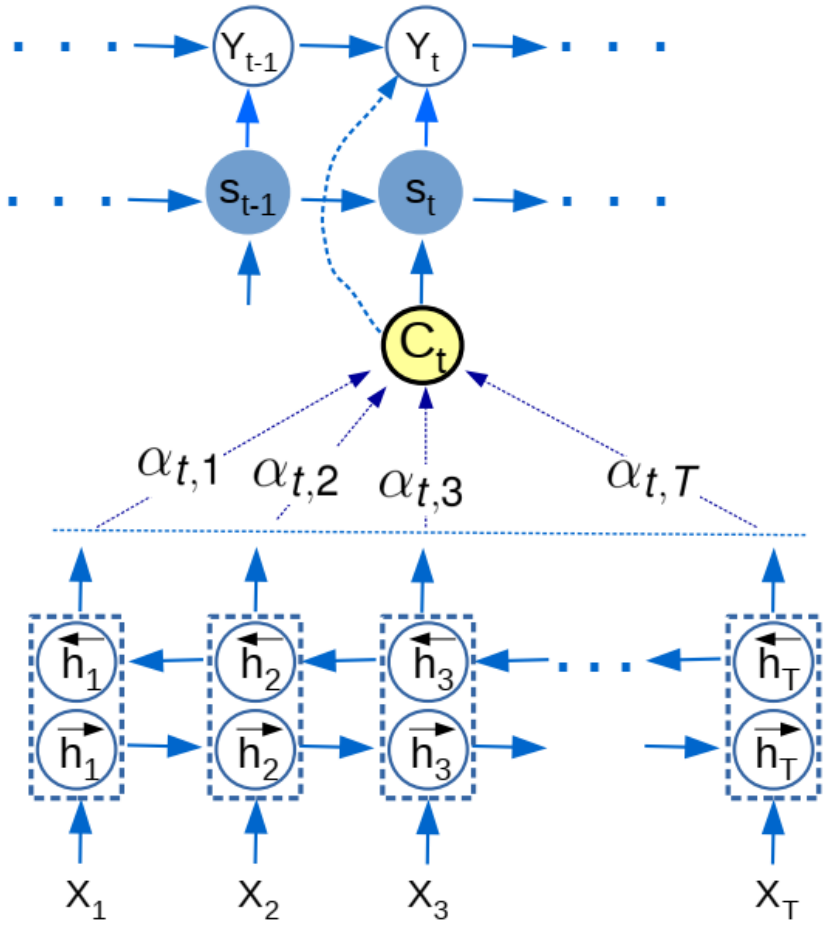


Veamos un ejemplo simple



Y la pregunta del millón es: ¿cómo se obtienen los pesos de la atención?

- Un esquema muy popular fue propuesto en 2015 por Bahdanau et al.



Y la pregunta del millón es: ¿cómo se obtienen los pesos de la atención?

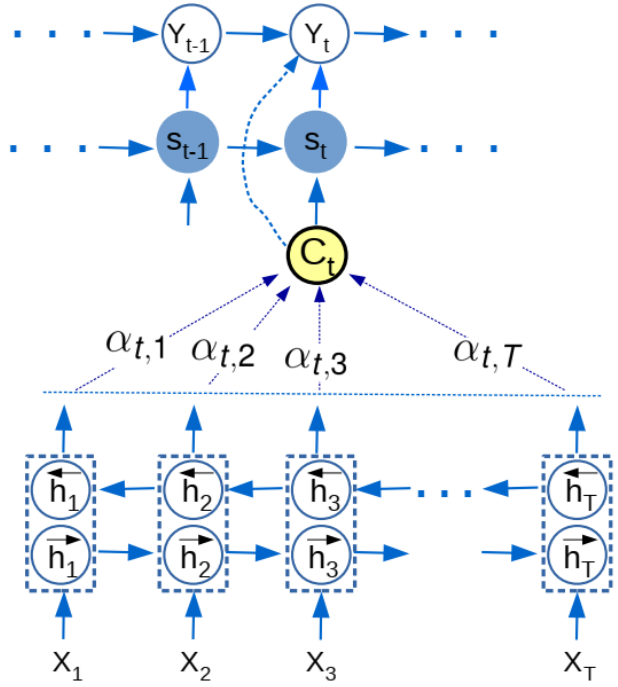
- En este modelo, el contexto adaptativo es utilizado para guiar todo el proceso:

$$y_t = \sigma(W_{yy} y_{t-1} + W_{sy} s_t + W_{cy} C_t)$$
$$s_t = \sigma(W_{ss} s_{t-1} + W_{cs} C_t)$$
$$C_t = \sum_{i=1}^{T} \alpha_{t,i} <\overrightarrow{h}_i, \overleftarrow{h}_i>$$

- La atención es capturada por los pesos $\alpha_{t,i}$, que codifican la relevancia de cada estado oculto del *encoder*:

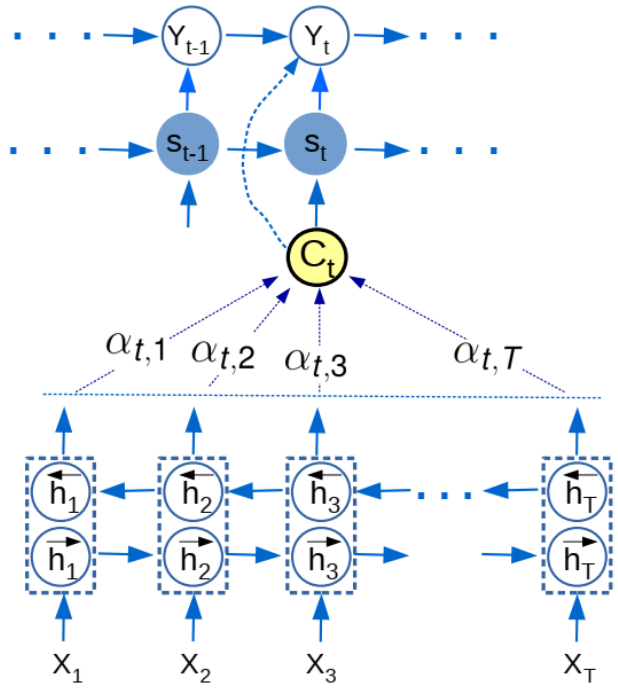
$$\hat{\alpha}_{tj} = V_c^T \sigma(W_c s_{t-1} + U_c h_j)$$

$$\alpha_{t,j} = \frac{\hat{\alpha}_{t,j}}{\sum_k \hat{\alpha}_{tk}}$$

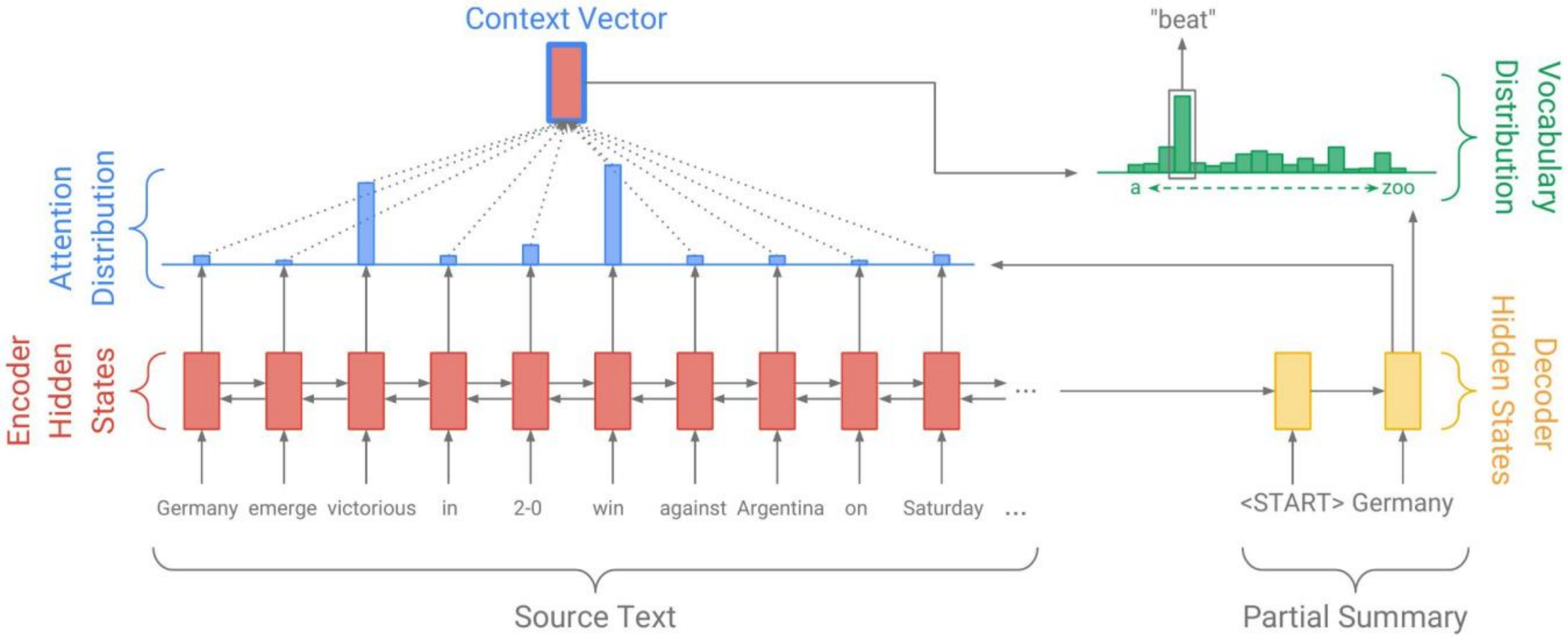


Y la pregunta del millón es: ¿cómo se obtienen los pesos de la atención?

- Al ser calculados mediante operaciones típicas de redes neuronales, la atención es diferenciable, por lo que se puede estimar mediante SGD.
- Este tipo de mecanismo de atención es conocido como *soft-attention* y es el más utilizado en modelos seq2seq.
- ¿Es posible extender este esquema de atención para que utilice, por ejemplo, otras fuentes de información?



Veamos más ejemplos: resumen de documentos (See y Manning, 2017)



Veamos más ejemplos: resumen de documentos (See y Manning, 2017)

Article (truncated): andy murray came close to giving himself some extra preparation time for his wedding next week before ensuring that he still has unfinished tennis business to attend to . the world no 4 is into the semi-finals of the miami open , but not before getting a scare from 21 year-old austrian dominic *thiem* , who pushed him to 4-4 in the second set before going down 3-6 6-4 , 6-1 in an hour and three quarters . murray was awaiting the winner from the last eight match between tomas berdych and argentina 's juan monaco . prior to this tournament *thiem* lost in the second round of a challenger event to soon-to-be new brit *aljaz* bedene . andy murray pumps his first after defeating dominic *thiem* to reach the miami open semi finals . *muray* throws his *sweatband* into the crowd after completing a 3-6 , 6-4 , 6-1 victory in florida . murray shakes hands with *thiem* who he described as a ' strong guy ' after the game . and murray has a fairly simple message for any of his fellow british tennis players who might be agitated about his imminent arrival into the home ranks : do n't complain . instead the british no 1 believes his colleagues should use the assimilation of the world number 83 , originally from slovenia , as motivation to better themselves .

andy murray defeated dominic *thiem* 3-6 6-4 , 6-1 in an hour and three quarters .
murray was awaiting the winner from the last eight match between tomas berdych and argentina 's juan monaco .
prior to this tournament thiem lost in the second round of a challenger event to soon-to-be new brit *aljaz* bedene .

Veamos más ejemplos: *image captioning* revisitado (Xu et al., 2015)

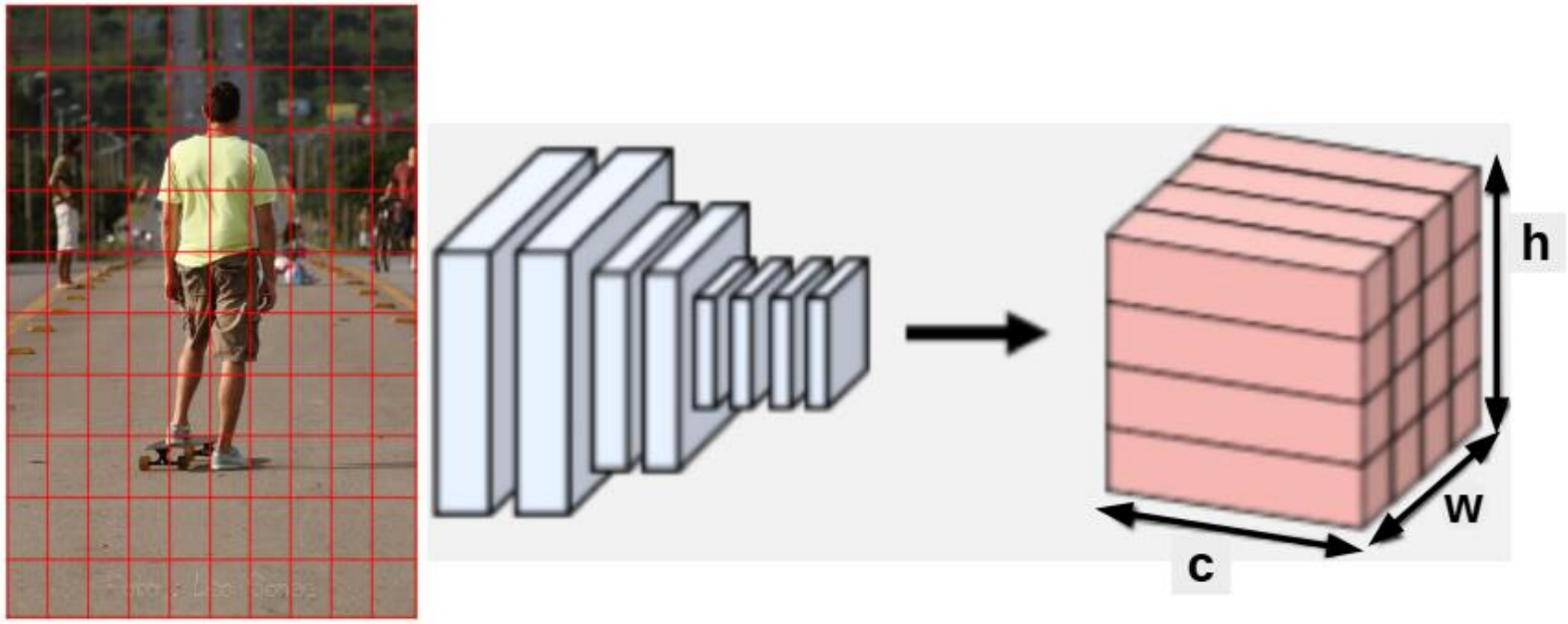
- En el esquema seq2seq, este problema es considerado como una traducción desde una imagen a texto (imágenes consideradas como un lenguaje).
- Si bien puede sonar intuitivo, este esquema tiene dos problemas fundamentales:

    1. La ubicación de la palabras (estructuras visuales relevantes) en la imágenes no está predefinida.
    2. La lista de palabras posibles (vocabulario visual) es desconocida.

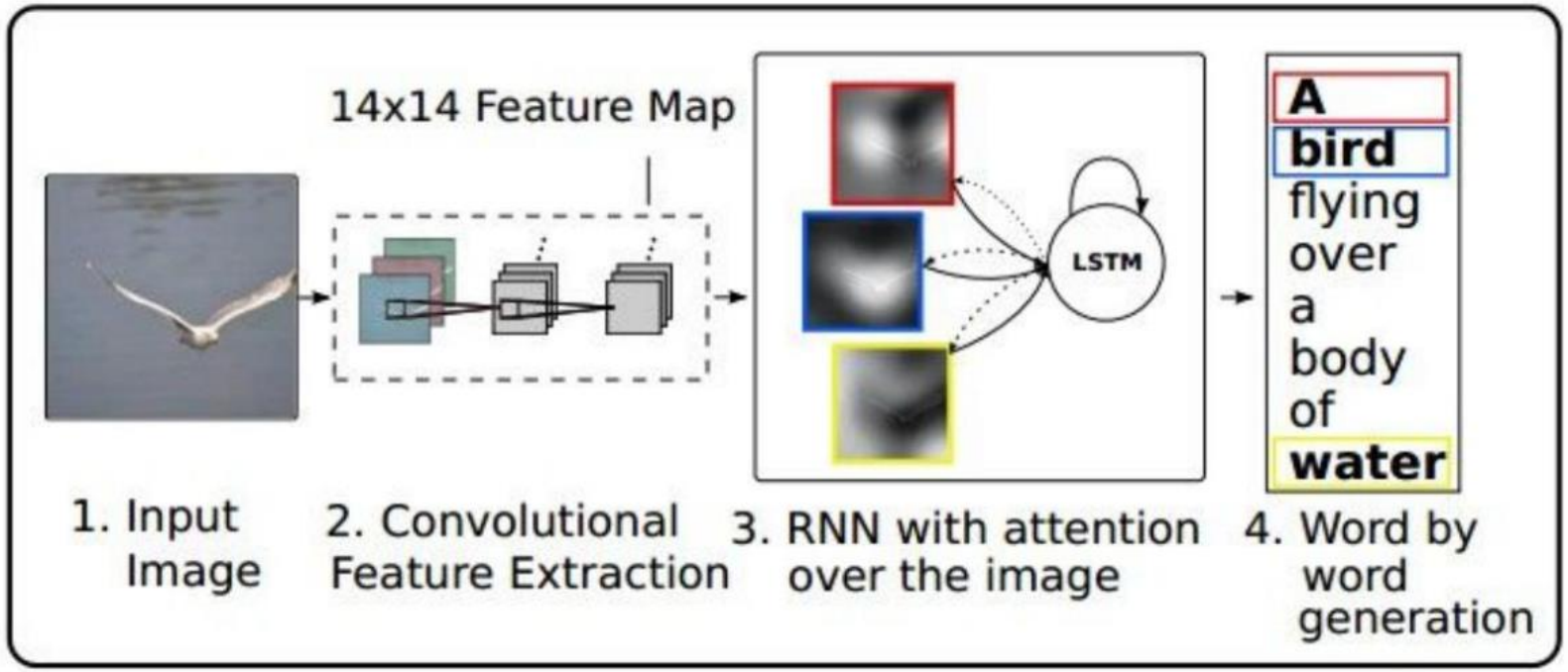
A woman with a little girl in a park, the woman is throwing a fresbee.

Veamos más ejemplos: *image captioning* revisitado (Xu et al., 2015)

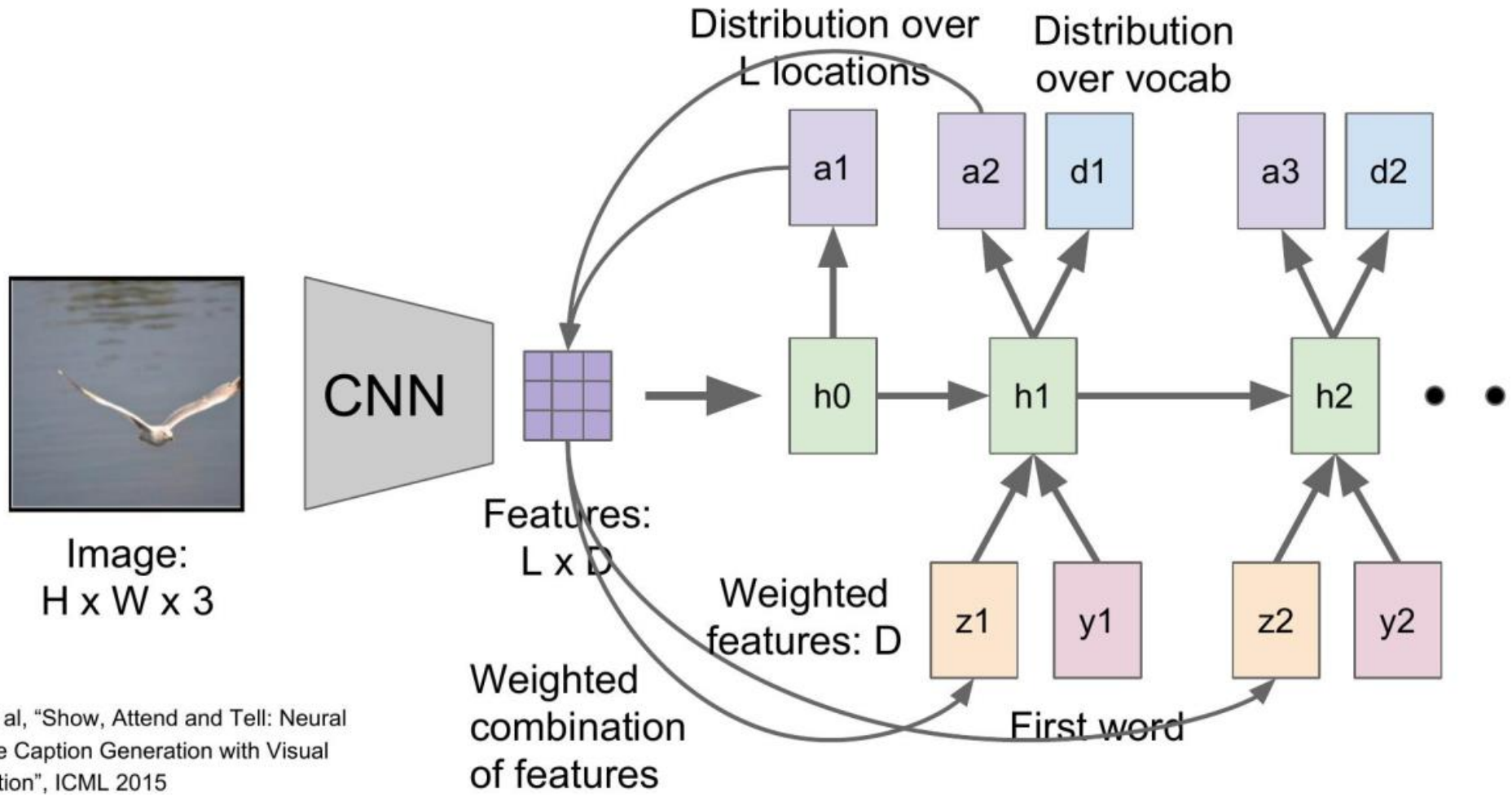
Si bien hay muchas maneras de enfrentar esto, la más directa es asumir como orden una grilla regular y como vocabulario visual las features de la última capa convolucional de una CNN (¿por qué no las generadas por las capas densas?)



Veamos más ejemplos: *image captioning* revisitado (Xu et al., 2015)



Veamos más ejemplos: *image captioning* revisitado (Xu et al., 2015)



Veamos más ejemplos: *image captioning* revisitado (Xu et al., 2015)

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

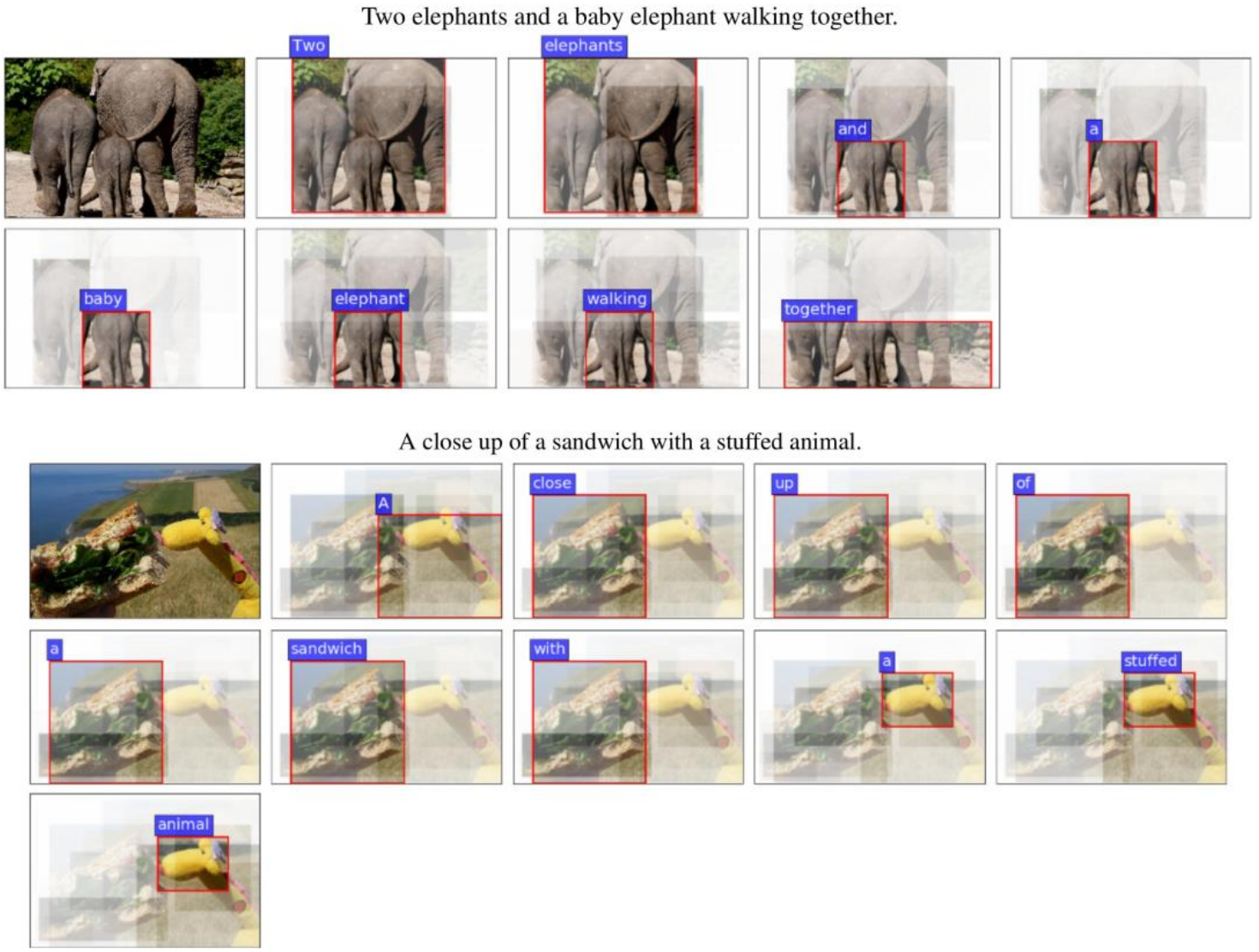
A stop sign is on a road with a mountain in the background.

A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

A giraffe standing in a forest with trees in the background.

Veamos más ejemplos: *image captioning* mejorado con atención sobre objetos (Anderson et al., 2018)



Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ciencia de la Computación

Sistemas Urbanos Inteligentes

Mecanismos de atención

Hans Löbel
Dpto. Ingeniería de Transporte y Logística
Dpto. Ciencia de la Computación